Business Analytics, Artificial Intelligence and Cherry Blossom Conference
22-23 March 2025, Washington DC

Day 1- Boot Camp: Predict-then-Optimize: Decision-Focused Learning
Invited Talks Schedule

Morning Session: Saturday, March 22, 2025, 10:00-11:40am

1. **Milind Tambe, Harvard University and Google DeepMind**
2. **Kai Wang, Georgia Institute of Technology**

Title: From Prediction to Action: Decision-Focused Learning for Impactful Applications (Joint Talk)

Abstract: Traditional machine learning often focuses on prediction accuracy, but many real-world problems require optimal decision-making. Decision-focused learning (DFL) addresses this gap by directly integrating optimization problems into the learning process. In this talk, we introduce the core concepts of DFL and showcase its transformative potential in public health, conservation and other applications. We will explore key challenges and solutions for DFL, including differentiable optimization layers, end-to-end learning frameworks, and novel methods for learning task-specific loss functions that replace the need for hand-crafted surrogates. We will draw from our recent research and highlight practical applications, including a deployed system for improving maternal health outcomes, and discuss how DFL can lead to more effective interventions in resource allocation, for example for optimized anti-poaching strategies, and other critical domains. We will also discuss the future challenges and opportunities in this exciting and impactful field.

Noon Session: Saturday, March 22, 2025, 12:10-1:00pm

3. **Pascal Van Hentenryck, Georgia Institute of Technology**

Title: Predict and Optimize in Energy Markets

Abstract: This presentation reviews the fundamentals of Predict and Optimize for Energy markets applications. It discusses their benefits on practical applications, computational Issues that arise in practice, and how they can be addressed.

Afternoon Session: Saturday, March 22, 2025, 3:00-4:40pm

4. **Hamsa Bastani, The Wharton School of the University of Pennsylvania**

Title: Improving Access to Essential Medicines in Sierra Leone via Decision-Aware Machine Learning

Abstract: A critical challenge in healthcare systems in Low- and Middle-Income Countries (LMICs) is the efficient and equitable allocation of scarce resources, particularly essential medicines. This problem is complicated by limited high-quality data, which restricts the applicability of traditional data-driven techniques. We propose a novel machine learning framework for essential medicines allocation, which leverages a combination of multi-task learning and decision-aware learning to improve sample efficiency and ensure equitable allocation. In collaboration with the Sierra Leone national government, our framework has been deployed nationwide as a decision support tool to help reduce waste and improve essential medicines allocation. Our evaluation using synthetic difference-in-differences analysis demonstrates a 21% increase in medicine consumption, with no changes to the supply, improving access for approximately 3.7 million women and children under five. Through experimental validation, we demonstrate that our approach also significantly outperforms baseline approaches. Our work demonstrates the tangible impact of machine learning in optimizing high-stakes decisions in resource-constrained settings, improving efficiency while ensuring equity and cost-effectiveness.

5. **Vahid Rostami, Pendulum Systems**

   Title: Toward an Agentic Framework Integrating Foundation Models for Adaptive, Human-Centric Decision-Making

   Abstract: We propose a novel agentic framework designed to enhance adaptive, human-centric decision-making in complex supply chain networks. Existing forecasting and optimization methods typically focus on isolated components, neglecting the interconnected, temporal dynamics critical to comprehensive supply chain management. Additionally, traditional approaches often struggle to incorporate dynamically evolving conditions, vaguely defined human objectives, and the complexity arising from supply chain dynamics that are only accessible through diverse and heterogeneous information sources. To overcome these limitations, our framework integrates advanced agentic systems with cutting-edge foundation models specifically adapted to the domain-specific language and actions of supply chains. This synergy enables a holistic representation and optimization of supply chain dynamics by building an intelligent system that comprehensively understands customer-specific supply chain operations through digesting and integrating diverse structured and unstructured data sources. Consequently, it transforms ambiguous or imprecise user objectives into a tailored set of actionable strategies that align closely with overall business priorities and goals.

Concluding Boot Camp: Saturday, March 22, 2025, 4:50-4:30pm

6. **Panel Discussion**

   Title: Panel Discussion and Wrapping Up the Boot Camp

Business Analytics, Artificial Intelligence and Cherry Blossom Conference
22-23 March 2025, Washington DC

Day 2: Recent Advances
Contributed Talks Schedule

Morning Session: Sunday, March 23, 2025, 8:30-10:00am

1. **Irene Aldridge, Cornell University**

   Title: Conditional Optimization

   Abstract: This paper shows that in any optimization problem, the expected loss always equals the covariance between the optimal decision and the decision variables, such as costs. The decision variables may be known or unknown at the time the optimal decision is made. This paper further shows that any optimization function can be estimated in advance from historical or even simulated data. This allows researchers to significantly reduce computational costs by pre-computing and reusing the optimization functions and avoiding peak computational costs.

2. **Long He, George Washington University**

   Title: Proactive Policing: A Resource Allocation Model for Crime Prevention with Deterrence Effect

   Abstract: This paper addresses police resource allocation across multiple locations, aiming to minimize the overall cost of potential crimes. Unlike previous literature focused on reactive police tasks, we propose a proactive approach that emphasizes crime prevention through deterrence. To account for the deterrence effect of police resources on crime, we employ the multinomial logit model to calibrate the distribution of crime locations. Our model sheds light on two facets of the deterrence effect in proactive policing—crime control diffusion and crime displacement—relevant to modern crime patterns from both criminology and economics perspectives. We also investigate the structural properties of our problem and its relation to mixture-of-logits assortment optimization. Additionally, we provide reformulations for mixed-integer linear/conic programs that can be solved directly using conventional optimization software. Finally, we showcase the efficacy of our model through a data-driven case study on the allocation of surveillance cameras in New York City.

3. **Tito Homem-De-Mello, Adolfo Ibañez University**

   Title: Forecasting Outside the Box: Application-Driven Optimal Pointwise Forecasts for Stochastic Optimization

   Abstract: The exponential growth in data availability in recent years has led to new formulations of data-driven optimization problems. One such formulation is that of stochastic optimization problems with contextual information, where the goal is to optimize the expected value of a certain function given some contextual information (also called features) that accompany the main data of interest. The contextual information then allows for a better estimation of the quantity of interest via machine learning methods, thereby leading to better solutions. Oftentimes, however, machine learning methods yield just a pointwise estimate instead of an entire distribution. In this paper we show that, when the problem to be solved is a class of two-stage stochastic programs (namely, those with fixed recourse matrix and fixed costs), under mild assumptions the problem can be solved with just one scenario. While such a scenario—which does not have to be unique—is usually unknown, we present an integrated learning and optimization procedure that yields the best approximation of that scenario within the modeler's pre-specified set of parameterized forecast functions. Numerical results conducted with inventory problems from the literature (with synthetic data) as well as a bike-sharing problem with real data demonstrate that the proposed approach performs well when compared to benchmark methods from the literature.

4. **Rui Gao, University of Texas at Austin**

   Title: Neural-Network Mixed Logit Choice Model: Statistical and Optimality Guarantees

   Abstract: The mixed logit model, widely used in operations, marketing, and econometrics, represents choice probabilities as a mixture of multinomial logits. We examine its formulation as a single-hidden-layer neural network that approximates the mixture distribution using equally weighted components. Despite its simplicity, this architecture poses intriguing theoretical challenges that merit deeper exploration.

   We establish a universal, dimension-independent bound on the model's root-mean-squared approximation error and demonstrate that overparameterization does not lead to overfitting when appropriate second moment- and entropy-regularization are applied. Building on these statistical insights, we propose a regularized parameter learning problem. We prove that, when the number of neurons approaches infinity, noisy stochastic gradient descent achieves global convergence at a nearly optimal rate, despite the non-convexity of the problem. Empirical studies on both synthetic and real datasets substantiate our theoretical findings. Our results highlight the effectiveness of overparameterized neural network representations, combined with efficient training algorithms, in learning choice models with strong performance guarantees.

5. **Mo Liu, University of North Carolina at Chapel Hill**

   Title: Value of One Data Point: Active Label Acquisition in Assortment Optimization

   Abstract: Predicting customers' preferences based on their features is crucial for personalized assortment optimization. When building this prediction model, using informative data can significantly increase the expected revenue from personalized assortments. This paper studies how to sequentially and actively collect informative data to construct this prediction model. We introduce a novel concept, the 'value of one data point,' which evaluates the marginal contribution of acquiring a specific customer's preference to the expected revenue in personalized assortment optimization, given the existing training set. Notably, this value drops to zero once the optimal assortment for this specific customer is determined. To estimate this value and identify important customers for acquiring their preferences, we derive a feature-dependent upper bound. This bound provides significant insights into the importance of each data point for revenue growth. Based on this upper bound, we develop a personalized incentive policy for effectively collecting survey data from customers to obtain their preferences. We provide non-asymptotic guarantees for both the cumulative incentives and the revenue from the final prediction model. Theoretically, we show that our personalized incentive policy requires smaller cumulative incentives than any fixed incentive policy to achieve the same level of revenue. Furthermore, our numerical experiments validate the effectiveness of our personalized incentive algorithms over fixed strategies.

Afternoon Session: Sunday, March 23, 2025, 1:30-3:00pm

6. Meng Qi, Cornell University

Title: Integrated Conditional Estimation-Optimization

Abstract: Many real-world optimization problems involve uncertain parameters with probability distributions that can be estimated using contextual feature information. In contrast to the standard approach of first estimating the distribution of uncertain parameters and then optimizing the objective based on the estimation, we propose an integrated conditional estimation-optimization (ICEO) framework that estimates the underlying conditional distribution of the random parameter while considering the structure of the optimization problem. We directly model the relationship between the conditional distribution of the random parameter and the contextual features and then estimate the probabilistic model with an objective that aligns with the downstream optimization problem. We show that our ICEO approach is asymptotically consistent under moderate regularity conditions and further provide finite performance guarantees in the form of generalization bounds. Computationally, performing estimation with the ICEO approach is a non-convex and often non-differentiable optimization problem. We propose a general methodology for approximating the potentially non-differentiable mapping from estimated conditional distribution to the optimal decision by a differentiable function, which greatly improves the performance of gradient-based algorithms applied to the non-convex problem. We also provide a polynomial optimization solution approach in the semi-algebraic case. Numerical experiments are also conducted to show the empirical success of our approach in different situations including with limited data samples and model mismatches.

7. Chao Qin, Stanford University

Title: Admissibility of Completely Randomized Trials: A Large-Deviation Approach

Abstract: When an experimenter has the option of running an adaptive trial, is it admissible to ignore this option and run a non-adaptive trial instead? We provide a negative answer to this question in the best-arm identification problem, where the goal of the experimenter is to select and deploy a treatment arm post-experiment with strong bounds on the regret from this selection. We demonstrate that simple adaptive designs, which sequentially eliminate underperforming arms, universally and strictly dominate non-adaptive completely randomized trials when there are at least three treatment arms. This dominance is characterized by a notion called efficient exponent, which quantifies a design's statistical efficiency in large within-experiment populations. Our result resolves an open problem posed in Qin (2022). This is a joint work with Guido Imbens and Stefan Wager.

8. **Neha Sharma, University of Pennsylvania**

Title: Information Design and Coordination in Cloud Kitchens – How Human-Tech Overrides Affect Order Fulfillment Time?

Abstract: Cloud kitchens, or ghost kitchens, have transformed food delivery by centralizing multiple brands in one location. Multi-brand cloud kitchens further enhance convenience, allowing customers to order from various brands in one transaction. While this offers convenience for customers, it also presents operational challenges, particularly in efficient workflow and real-time information sharing for order fulfillment.

Our analysis of data from a multi-brand cloud kitchen platform in India, which operates 69 kitchens with 10-48 restaurant brands each, finds that multi-brand orders represent only 0.01% to 2% of total orders but have fulfillment times about 50% longer (10.6 additional minutes) than single-brand orders. This indicates inefficiencies in processing such orders. Additionally, many brands collaborate by offering dishes from partner brands, complicating order processing. In inter-brand collaborations, customer orders reflect only the brand menu the dish is ordered from, not the brand that prepares it. We classify orders into three categories: single-brand orders where all dishes are from the same brand; explicit multi-brand orders where dishes come from different brands; and implicit multi-brand orders where dishes are ordered from one brand but prepared by multiple brands.

Field visits revealed that the platform uses brand information from customer orders to send dish details to cooking stations. This often leads to incorrect dish assignments in implicit multi-brand orders, requiring chefs to reassign dishes manually. This disrupts the first-come, first-served (FCFS) workflow and deprioritizes other orders. Our findings underscore the need for better technology-driven order sequencing and information design. Explicit multi-brand orders experience longer fulfillment times, while implicit orders benefit from chef-led adjustments. To improve efficiency, it’s essential to assess information-sharing practices and sequencing policies in cloud kitchens.

Research Question: This study investigates: (1) How do platform rules for order sequencing and information sharing affect fulfillment times for different order types? (2) When should the platform adjust sequencing rules or prioritize multi-brand orders to minimize delays? (3) What is the ideal information-sharing policy that ensures fulfillment efficiency while considering temperature-sensitive dishes and overall equity?

Empirical Strategy: Our empirical analysis uses 6.6 million orders from a large Indian cloud kitchen platform over 16 months (June 2021 – September 2022), considering order composition, brand associations, timestamps, kitchen workload, and chef availability. To evaluate the impact of sequencing and information-sharing policies, we model the cloud kitchen as a fork-join queuing network. Orders are divided into dishes, which are then sent to the respective brand’s cooking station for preparation. Once all dishes are prepared, they are synchronized for final packaging and delivery.

Closed-form solutions for fork-join queues with multiple servers and batch arrivals do not exist, so we estimate key system parameters using the Simulated Method of Moments (SMM). This estimated model allows us to simulate different policies that vary in order sequencing and information transparency. Each policy is assessed on its impact on fulfillment times, synchronization for temperature sensitive orders, and overall efficiency.

Contributions: Our work enhances the literature on information design and human-tech interaction by combining large-scale empirical data with a data-driven queuing model. It provides insights into coordination dynamics in multi-brand cloud kitchens, informing platform policies that balance fulfillment times, food quality, and system-wide equity. The findings also apply to other service environments, such as hospital laboratories and pharmaceutical manufacturing.

9. Leann Thayaparan, Johns Hopkins University

Title: UMOTEM: Upper Bounding Method for Optimizing over Tree Ensemble Models

Abstract: Machine learning has become core to forecasting and planning. However, when decision makers are provided with trained and more complex machine learning models, often these models are difficult to optimize over then. When tree-based ensemble models, such as Random Forest or XGBoost, are used in optimization formulations, they require an exponential number of binary decision variables. Optimization problems of this type do not scale well. We propose UMOTEM (Upper Bounding Method for Optimizing over Tree Ensemble Models), an algorithm for solving a constrained optimization problem where the objective function is determined by a tree ensemble model. The algorithm narrows the region of decision variables to an approximate region of optimality by iteratively optimizing using upper bounds as it moves down the trees in the ensemble, at each step only using information available at that depth of the tree. This significantly improves the problem's complexity, with the number of binary variables scaling only linearly, quickly outpacing the exponential growth of the alternative formulations. We show how this method can be used to jointly predict and optimize to save time building sub-optimal branches of the decision trees. We prove an expected optimality gap bound for Random Forest in terms of the forest's in-sample error and leaf separation and show when it is tight. We demonstrate computationally that our algorithm can capture at least 90% of optimality on a variety of datasets. Finally, we show through work with Oracle Retail, for one of their fashion retailer clients, how UMOTEM can increase revenue by 12-13%.

10. Sainan Zhang, George Washington University

Title: Wildfire-Driven Distribution System Operation: New Chance-Constrained Model with Decision-Dependent Probabilities

Abstract: Power line unavailability leads to disruptions of power flow and can be caused by both external factors (e.g., wildfire) and endogenous factors (e.g., power flow levels). This paper introduces a new chance-constrained stochastic programming model with decision-dependent uncertainty in which the line availability probability is affected by the magnitude of wildfire and further distorted by power flow levels. The model determines changes in network topology under the consideration of wildfire ignition risks and minimizes the total system costs associated with active power generation, switching action, and penalties for power imbalance while ensuring operational reliability. To model the stochastic availability of power lines, we represent how wildfires affect the line availability probability (exogenous uncertainty) which is subsequently modified using a distortion function that accounts for how power flow decisions (endogenous uncertainty) impact line availability. We use and calibrate a copula function to model the dependency across power lines and their availability. We derive a deterministic and equivalent reformulation of the chance-constrained model that takes the form of a nonconvex mixed-integer nonlinear programming (MINLP) problem. Estimating a piecewise linear generator function for the copula, we reformulate the MINLP problem into a mixed-integer linear programming problem. The numerical tests showcase the applicability and effectiveness of the proposed reformulation and algorithmic approach and provide practically relevant insights.

Business Analytics, Artificial Intelligence and Cherry Blossom Conference

22-23 March 2025, Washington DC

Day 2: Recent Advances

Student Talks Schedule

Students' Session: Sunday, March 23, 2025, 10:30am-12:00pm

1. **Rakesh Allu, Samuel Curtis Johnson Graduate School of Management, Cornell University**

   Title: Ranking Quality and User Engagement on an Online B2B Platform

   Abstract: Online business-to-business (B2B) platforms are increasingly investing in data science teams to develop machine-learning algorithms to provide personalized rankings-based user recommendations. These algorithms continuously evolve over time due to frequent innovations by data scientists and dynamic learning of weights from users' recent activity. Thus, there is a growing need to develop ways to monitor the performance of these algorithms over time. Two key questions faced by platforms are: (i) how to periodically measure the quality of rankings presented to the user using real-time ranked transactions data and, (ii) how to examine the effect of improving ranking quality on usage (i.e., the number of transactions) and browsing effort of the user. These questions are challenging to address because of simultaneity between usage and effort, and censoring in ranked transactions data. Using detailed transaction-level data from a B2B platform, we propose methods to measure ranking quality, develop a position-level model of a user's decision to transact and scroll, and develop estimation approaches to overcome censoring. Our analysis reveals a position-wise dynamic that results in an increase in platform usage and a decrease in browsing effort with improvement in ranking quality. We find that users in different product categories respond differently to ranking quality improvements. We present a framework for the platform to utilize this heterogeneity in prioritizing its efforts to improve the recommendation system. Our position-level method can be used to evaluate and manage continuous improvements to recommendations systems on any ranked page with a cascade browsing model.

2. **Lin An, Tepper School of Business, Carnegie Mellon University**

   Title: Real-Time Personalization with Simple Transformers

   Abstract: Real-time personalization has advanced significantly in recent years, with platforms utilizing machine learning models to predict user preferences based on rich behavioral data on each individual user. Traditional approaches usually rely on embedding-based machine learning models to capture user preferences, and then reduce the final real-time optimization task to one of nearest-neighbors, which can be performed extremely fast both theoretically and practically. However, these models struggle to capture some complex user behaviors, such as sequence effects, complementarity, or variety effects, which are essential for making accurate recommendations. Transformer-based models, on the other hand, are known for their practical

ability to model sequential behaviors, and hence have been intensively used in personalization recently to overcome these limitations. However, optimizing recommendations under transformer-based models is challenging due to their complicated architectures. In this paper, we address this challenge by considering a specific class of transformers, showing its ability to represent complex user preferences, and developing efficient algorithms for real-time personalization.

We focus on a particular set of transformers, called simple transformers, that contain a single self-attention layer. We show that simple transformers are capable of capturing complex user preferences, such as variety effects, complementarity and substitution effects, and various choice models, which traditional embedding based models cannot capture. We then develop an algorithm that enables fast optimization of real-time personalization tasks based on simple transformers. Our algorithm achieves near-optimal performance and has sub-linear runtime. Finally, we demonstrate the effectiveness of our approach through an empirical study on large datasets from Spotify and Trivago. Our experiment results show that (1) given data on past user behavior, simple transformers can model/predict user preferences substantially more accurately than nontransformer models and nearly as accurately as more-complex transformers, and (2) our algorithm completes simple-transformer-based recommendation tasks quickly and effectively. Comparing against two natural benchmark algorithms, our algorithm on average achieves objective values 4.5% higher than Beam Search and 6.5% higher than k-Nearest Neighbor.

3. **Sandeep Chitla, Leonard N. Stern School of Business, New York University**

Title: Consumers’ Cart-Building Behavior in Online Grocery: A Structural Approach Using Generative AI

Abstract: Problem definition: How do customers build their online grocery shopping carts across multiple product categories, and how do operational factors such as prices, delivery fees, and promotions influence their choices? Existing multi-category choice models in operations, marketing, and economics often overlook the dynamic and sequential nature of cart-building and struggle with scalability in real-world scenarios involving tens of thousands of products. Machine learning models, while scalable, often function as black boxes with limited interpretability. Our approach bridges this gap by combining the scalability of machine learning with the interpretability of structural choice models. Methodology/results: We model customers as agents who sequentially select product categories based on their needs and choose products within categories according to their preferences. Using Generative AI techniques, we estimate the utilities driving these choices, extending the application of generative AI beyond its conventional domains of language, image, audio, and video to build a customer digital twin that replicates customer behavior on online platforms. Our analysis leverages a dataset from a leading quick-commerce platform in India, comprising 8.5 million orders from 350,000 customers. Our model accurately predicts the next product category added to the shopping cart 6.6% of the time from approximately 2,000 categories, with the true category appearing in the top 10 predictions 38.8% of the time. Managerial implications: Our model’s structural insights reveal that customers strongly respond to delivery fee waivers over flat discounts of the same value, highlighting the behavioral characteristic of avoiding additional service fees. Our counterfactual analyses reveal how increasing delivery thresholds impacts cart value and product demand, with discretionary items like chips and longer shelf-life items like onions experiencing an increase in demand as thresholds rise.

4. **Ali Dasouqi, Carey Business School, Johns Hopkins University**

Title: Navigating the Real World: The Limitations of AI in Address Orientation and the Need for Subject Matter Expertise

Abstract: Artificial Intelligence (AI) has made remarkable strides in fields such as natural language processing, predictive analytics, and automation. However, when it comes to navigating real-world spatial data, AI still struggles with fundamental tasks, such as accurately determining the front-facing elevation of a building. This paper presents an analysis of multiple instances where AI failed to correctly interpret the orientation of residential and commercial properties based on address data and mapping technologies. Despite leveraging satellite imagery, street view analysis, and geospatial data, AI systems still misidentified the primary entrance or driveway alignment, leading to inaccurate conclusions about a property's front-facing direction.

The challenges encountered in this study highlight a broader issue: AI, even with access to vast amounts of data, lacks the nuanced judgment and contextual awareness that human experts possess. While mapping tools provide raw data, human knowledge, such as understanding local building conventions, road alignments, and urban planning nuances, remains indispensable for accurate interpretation. This paper argues that AI alone is insufficient for many real-world business analytics applications and that integrating subject matter expertise is essential for practical AI deployment.

To address these challenges, businesses and AI developers must adopt a hybrid approach that leverages AI's efficiency alongside human intuition and expertise. This could involve developing AI models trained on expert-validated geospatial data, improving context-aware AI algorithms, and designing interactive AI-human collaboration frameworks for decision-making. By incorporating domain knowledge into AI-powered solutions, businesses can enhance AI's reliability and make it more practical for applications requiring spatial intelligence.

This paper contributes to the growing discourse on AI's limitations and the importance of human-AI collaboration. It underscores the need for AI innovations that rely not merely on data-driven inference but also integrate subject matter expertise to ensure accuracy, usability, and real-world applicability in business analytics and decision-making.

5. **Shivam Kumar, Applied and Computational Math and Statistics, University of Notre Dame**

Title: Estimation and Inference for Change Points in Functional Regression Time Series

Abstract: In this research, we study the estimation and inference of change points under a functional linear regression model with changes in the slope function. We present a novel Functional Regression Binary Segmentation (FRBS) algorithm which is computationally efficient as well as achieving consistency in multiple change point detection. This algorithm utilizes the predictive power of piece-wise constant functional linear regression models in the reproducing

kernel Hilbert space framework. We further propose a refinement step that improves the localization rate of the initial estimator output by FRBS, and derive asymptotic distributions of the refined estimators for two different regimes determined by the magnitude of a change. To facilitate the construction of confidence intervals for underlying change points based on the limiting distribution, we propose a consistent block-type long-run variance estimator. Our theoretical justifications for the proposed approach accommodate temporal dependence and heavy-tailedness in both the functional covariates and the measurement errors. Empirical effectiveness of our methodology is demonstrated through extensive simulation studies and an application to the Standard and Poor's 500 index dataset.

6. **Imran Pervez, Electrical and Computer Engineering, King Abdullah University of Science and Technology**

Title: Integrated Learning and Optimization for Congestion Management and Cost Minimization in Real-Time Electricity Market

Abstract: Economic dispatch (ED) and DC optimal power flow (DCOPF) are optimization formulations in power systems that determine power dispatch decisions to minimize generator operational costs. ED excludes transmission constraints, while DCOPF incorporates them to limit power flow. These formulations involve unknown parameters that must be estimated before solving the optimization problem. In ED, load (consumer demand) is the unknown parameter, whereas in DCOPF, the load and the power transfer distribution factor (PTDF) matrix are unknowns where the PTDF matrix coefficients represent a linearized approximation of power flows over the transmission lines.

We develop novel Integrated Learning and Optimization (ILO) methodology to train ED and DCOPF unknowns, ensuring their decisions improve economic operation when deployed in real-world power system applications. In ILO training pipeline, predictions of unknown parameters are evaluated using contextual features, passed through an optimization problem to generate decisions, and compared against target (true data)-driven decisions. The resulting discrepancy forms a regret function, whose gradient is backpropagated to update the prediction model.

The solution to ED/DCOPF with unknown parameters being trained using proposed ILO formulations provide decisions to improve the economic operation when deployed in a power system application which in this work is the real-time electricity market. In electricity market, prior to real-time market (RTM) operation, the day-ahead market is cleared which provide forecast for the unknown parameter load, while ISO provide forecast for the unknown parameter PTDF in the ED/DCOPF optimization setting. The ED/DCOPF optimization is then solved to generate power dispatching decisions corresponding to forecast. The ED/DCOPF decision generated using forecast differs from the decision generated using real-time data (true values for the unknowns in the optimization formulations realized in real-time) which create supply-demand imbalance and line congestion. The discrepancy between forecast-driven decision and true decision is corrected in real-time market, where an independent system operator (ISO) interact with real-time market participants willing to ramp-up/-down their generators to achieve supply-demand balance by correcting forecast-driven decisions. Moreover, ISO control line impedance prior to real-time operation to minimize line congestion. The ramping operation in real-time market to correct supply-demand imbalance comes at a higher cost than the actual market cost with ramping-up (↑) operation being more expensive than the ramping-down (↓) operation. Moreover, the line impedances for different lines used to generate PTDF matrix for DCOPF solution require impedance of each line having a specific correlation to other lines in order to generate DCOPF decisions for minimum line congestion which may not be available to ISO.

Inspired by the above market and congestion management procedures, the ILO formulation in this work is designed to train ED and DCOPF unknowns to minimize real-time market ramping costs and finding correlations between line impedances to minimize line congestion rather than conventional training where unknowns are trained to accurately estimate the target data and do not consider market economic operations and line congestion. We designed ILO regret function with the objective to minimize ramping costs emphasizing ↓ operation more than ↑ operation, and minimize line congestion. The gradient of the regret function with the above-mentioned objectives is backpropagated to update load and PTDF matrix model parameters. The load and PTDF after ILO training when deployed in ED/DCOPF power scheduling provide dispatch decisions with minimum ramping costs favoring ↓ more than ↑ in RTM and minimum line congestion which improves generator operational costs while further minimizing RTM ramping costs. Our regret function is designed by thoroughly analyzing the feasible region of ED/DCOPF optimization problems and performing sensitivity analysis for the unknown parameters. We derived insights particularly for DCOPF feasible region to understand PTDF sensitivity analysis. Our analysis show the PTDF matrix despite being an equality constraint in DCOPF formulation and infeasible with respect to true PTDF matrix corresponds to minimum line congestion (DCOPF decision equivalent to ED decision) if trained within a certain sensitive range. The above-mentioned sensitivity analysis for regret function design along with understanding the impact of one unknown on another during training, their impact at different stages of market operations is in general missing in the literature. The proposed methodology with the resulting regret function is compared to sequential learning and optimization (SLO) which train load and PTDF forecasts for accuracy rather than economic operation. Our experimentation proves the superiority of proposed ILO methodology in minimizing RTM ramping costs and line congestion, thereby improving the economic operation with a significant amount compared to SLO.

7. **Zhiyuan Tang, Naveen Jindal School of Management, University of Texas at Dallas**

Title: Match Made with Matrix Completion: Efficient Offline and Online Learning in Matching Markets

Abstract: Online matching markets face increasing needs to accurately learn the matching qualities between demand and supply for effective design of matching policies. However, the growing diversity of participants introduces a high-dimensional challenge in practice, as there are a substantial number of unknown matching rewards and learning all rewards requires a large amount of data. We leverage a natural low-rank matrix structure of the matching rewards in these two-sided markets, and propose to utilize matrix completion (specifically the nuclear norm regularization approach) to accelerate the reward learning process with only a small amount of offline data. A key challenge in our setting is that the matrix entries are observed with matching interference, distinct from the independent sampling assumed in existing matrix completion

literature. We propose a new proof technique and prove a near-optimal average accuracy guarantee with improved dependence on the matrix dimensions. Furthermore, to guide matching decisions, we develop a novel “double-enhancement” procedure that refines the nuclear norm regularized estimates and further provides near-optimal entry-wise estimations. Our paper makes the first investigation into adopting matrix completion techniques for matching problems. We also extend our approach to online learning settings for optimal matching and stable matching by incorporating matrix completion in multi-armed bandit algorithms. We present improved regret bounds in matrix dimensions through reduced costs during the exploration phase. Finally, we demonstrate the practical value of our methods using both synthetic data and real data of labor markets.

8. **Christopher Yeh, Department of Computing and Mathematical Sciences, California Institute of Technology**

Title: End-to-End Conformal Calibration for Optimization Under Uncertainty

Abstract: Machine learning can significantly improve performance for decision-making under uncertainty in a wide range of domains. However, ensuring robustness guarantees requires well-calibrated uncertainty estimates, which can be difficult to achieve with neural networks. Moreover, in high-dimensional settings, there may be many valid uncertainty estimates, each with their own performance profile—i.e., not all uncertainty is equally valuable for downstream decision-making. To address this problem, this paper develops an end-to-end framework to learn uncertainty sets for conditional robust optimization in a way that is informed by the downstream decision-making loss, with robustness and calibration guarantees provided by conformal prediction. In addition, we propose to represent arbitrary convex uncertainty sets with partially input-convex neural networks, which are learned as part of our framework. Our approach consistently improves upon two-stage estimate-then-optimize baselines on concrete applications in energy storage arbitrage and portfolio optimization.