Big Data 2020

Conference Program

Monday, August 24

10:00–10:15 am Introduction & Welcome by Horng-Tzer Yau

10:15-11:15 am Varda Shalev, Tel Aviv University
Title: "From Big Data to small medical meaningful insights"

Abstract: Medicine is very complexed. The volume of data that is generated daily is huge. There are many dimensions that one must take in account while treating a patient continuous and holistically. The best physician cannot do it without two team members: a smart decision support system (based on AI and Big Data and an empowered cooperative patient. The way to achieve the goal is based mostly on leadership and true collaboration.

11:20-12:00 pm Joseph Dexter, Dartmouth College
Title: "Strategies for Clear Communication About COVID-19"

Abstract: Containment strategies for the COVID-19 pandemic involve non-pharmaceutical interventions, such as social distancing, requiring broad public compliance. Given the widespread proliferation of complex, contradictory, and false information about COVID-19, it is vital that members of the public be able to understand and use recommendations for health protective behavior from trustworthy sources. In this talk, I will discuss a recent cross-sectional study of official written information about COVID-19. Using both standardized readability formulas and stylometric markers of text complexity, the study demonstrates that most public health guidance about the pandemic may be challenging to understand, suggesting an urgent need for the development of more accessible and inclusive approaches to communication.

12:00–1:15 pm Lunch Break

1:15-1:55 pm Nicole Immorlica, Microsoft
Title: "Incentivizing Exploration with Selective Data Disclosure"

Abstract: We study the design of rating systems that incentivize efficient social learning. Agents arrive sequentially and choose actions, each of which yields a reward drawn from an unknown distribution. A policy maps the rewards of previously-chosen actions to messages for arriving agents. The regret of a policy is the difference, over all rounds, between the expected reward of the best action and the reward induced by the policy. Prior work proposes policies that recommend a single action to each agent, obtaining optimal regret under standard rationality assumptions. We instead assume a frequentist behavioral model and, accordingly, restrict attention to disclosure policies that use messages consisting of the actions and rewards from a subsequence of past agents, chosen ex ante. We design a policy with optimal regret in the worst case over reward distributions. Our research suggests three components of effective policies: independent focus groups, group aggregators, and interlaced information structures.
Joint work with Jieming Mao, Aleksandrs Slivkins, and Zhiwei Steven Wu.

2:00-2:40 pm Amin Saberi, Stanford University

Title: "Matching in Dynamic Environments"

Abstract: The theory of matching with its roots in the work of mathematical giants like Euler and Kirchhoff has played a central and catalytic role in combinatorial optimization for decades. More recently, the growth of online marketplaces for allocating advertisements, rides, or other goods and services has led to new interest and progress in this area. I will start the talk by giving examples from various industries and survey a few models, algorithms, and open problems in the context of ride sharing.

Tuesday, August 25

10:15-11:15 am Sanjeev Arora, Princeton University
Title: "Opening the black box: Toward mathematical understanding of deep learning"

Abstract: Deep learning has led to significant progress on old problems of AI and machine learning. But mathematical understanding of this technique is still lacking. The talk will survey the main mathematical questions and the hurdles the confront researchers trying to answer them. It will also highlight the inadequacies of traditional optimization-based language for thinking about deep learning.

11:20-12:00 pm Vira Semenova, University of California, Berkeley
Title: "Machine Learning for Causal Inference"

Abstract: We study the problem of estimating average welfare in a dynamic discrete choice problem. We first show that value function is orthogonal to the conditional choice probability. Second, we give a correction term for the transition density of the state variable. The resulting orthogonal moment is robust to misspecification of the transition density and does not require this nuisance function to be consistently estimated. Third, we generalize this result by considering the weighted expected value. In this case, the orthogonal moment is doubly robust in the transition density and additional second-stage nuisance functions entering the correction term. We complete the asymptotic theory by providing bounds on second-order asymptotic terms. Joint work with Victor Chernozhukov and Whitney Newey.

12:00–1:15 pm Lunch Break

1:15-1:55 pm Juan Camilo Castillo, University of Pennsylvania
Title: "Who Benefits from Surge Pricing?"

Abstract: In the last decade, new technologies have led to a boom in dynamic pricing. I analyze the most salient example, surge pricing in ride hailing. Using data from Uber in Houston, I develop an empirical model of spatial equilibrium to measure the welfare effects of surge pricing. The model is composed of demand, supply, and a matching technology. It allows for temporal and spatial heterogeneity as well as randomness in supply and demand. I find that, relative to a counterfactual with uniform pricing, surge pricing increases total welfare by 1.59\% of gross revenue. The gains mainly go to riders: rider surplus increases by 5.25\% of gross revenue, whereas driver surplus and platform profits decrease by 1.81\% and 1.77\% of gross revenue, respectively. Riders at all income levels benefit, while disparities in driver surplus are magnified.

1:55–2:00 pm Closing Remarks by Scott Kominers