PURDUE UNIVERSITY | Krannert School of Management

Supply Chain & Operations Management Seminar

Dr. Ruihao Zhu

Assistant Professor, Supply Chain & Operations Management
Krannert School of Management, Purdue University

Safe Optimal Design with Applications in Policy Learning

Friday, October 22, 2021 | 10:00—11:10 am

Bio

Ruihao Zhu is an Assistant Professor at Purdue Krannert School of Management. He received his Interdisciplinary Ph.D. in Statistics from the Massachusetts Institute of Technology and B.Eng. degrees in Electrical Engineering and Computer Science from both the Shanghai Jiao Tong University and the University of Michigan in 2015. He spent the summers of 2020 and 2019 as a research scientist intern with Amazon and Google Research. Ruihao's research seeks to address fundamental and practical challenges in recommendation systems, digital experimentation, retailing & pricing, and supply chain analytics by connecting optimization, machine learning, and economics. As part of it, he has been collaborating with companies across different industries, such as consumer packaged goods and manufacturing. His works have been recognized by a Finalist in INFORMS Service Science 2021 Best Cluster Paper Award, an Honorable Mention in INFORMS George E. Nicholson 2019 Student Paper Competition, and a Finalist in POMS-JD.com 2019 Best Data-Driven Research Paper Competition.

Abstract

Motivated by practical needs in online experimentation and off-policy learning, we study the problem of *safe optimal design*, where we develop a data *logging policy* that efficiently explores while achieving competitive rewards with a baseline *production policy*. We first show, perhaps surprisingly, that a common practice of mixing the production policy with uniform exploration, despite being safe, is sub-optimal in maximizing information gain. Then we propose a safe optimal logging policy for the case when no side information about the actions' expected rewards is available. We improve upon this design by considering that side information and also extend both approaches to a large number of actions with a linear reward model. We analyze how our data logging policy impacts errors in off-policy learning. Finally, we empirically validate the benefit of our designs by conducting extensive experiments.