北京理工大學

数学与统计学院学术报告

Attack-Resistant Uniform Fairness for Linear and Smooth Contextual Bandits

报告人： Wenjia Wang（National University of Singapore）

时间： 6月12日 10：00

地点： 文萃楼E708

摘要： Modern digital platforms and service systems increasingly use contextual bandits for online decision-making, but their deployment can create unfair exposure across arms, threatening platform sustainability and supplier trust. We study contextual bandits under a uniform (1-\delta)-fairness constraint, which requires preferential treatment to be justified by true rewards across all contexts and time horizons, closing statistical loopholes. We develop algorithms for linear and smooth rewards that achieve nearly minimax-optimal regret while guaranteeing \left(1-\tilde{O}(1/T)\right)-fairness, and characterize the asymptotically small price of fairness. We further show that this merit-based notion is vulnerable to signal manipulation: an adversary with only \tilde{O}(1) budget can trigger fairness-specific failures without noticeably affecting regret. To address this, we design robust, corruption-adaptive algorithms that attain minimax-optimal regret under C-budgeted attacks while preserving \left(1-\tilde{O}(1/T)\right)-fairness. Numerical and real-world results confirm strong fairness and efficiency.

个人简介： Wenjia Wang is an Assistant Professor (Presidential Young Professor) in the Department of Industrial Systems Engineering & Management at National University of Singapore. He obtained his Ph.D. in the School of Industrial & Systems Engineering at Georgia Institute of Technology. Wenjia Wang's research interests include uncertainty quantification, stochastic simulation, machine learning, computer experiments, and nonparametric statistics.

主办单位：北京理工大学数学与统计学院
School of Mathmatics and Statistics, Beijing Institute of Technology