Smooth Contextual Bandits: Bridging the Parametric and Non-differentiable Regimes

Abstract:
Contextual bandit problems model the inherent tradeoff between exploration and exploitation in personalized decision making in marketing, healthcare, revenue management, and beyond. Specifically, the tradeoff is characterized by the optimal growth rate of the regret in cumulative rewards compared to the optimal policy we are trying to learn, and naturally the optimal rate depends on how complex the underlying learning problem is, namely how much can observing reward in one context tell us about mean rewards in another. Curiously, this obvious-seeming relationship is obscured in current theory that separately studies the easy super-extrapolatable and hard super-local problems. So, to characterize the relationship more precisely I will present a nonparametric contextual bandit problem where the expected reward functions belong to a Hölder class with smoothness parameter $\beta$ (approximately meaning they are $\beta$-times differentiable). I will show how this interpolates between two extremes that were previously studied in isolation: non-differentiable bandits ($\beta \leq 1$), where rate-optimal regret is achieved by running separate non-contextual bandits in different context regions, and parametric-response bandits ($\beta = \infty$), where rate-optimal regret can often be achieved with minimal or no exploration due to infinite extrapolatability from one context to another. We develop a novel algorithm that works for any given smoothness setting in between and that operates neither fully locally nor fully globally. We prove its regret is rate-optimal by establishing matching upper and lower bounds, recovering the existing results at the two extremes. In this sense, our work bridges the regret gap between the non-differentiable and parametric bandit problems and between bandit algorithms that exclusively use global or local information, shedding light on the crucial interplay of complexity and regret in dynamic decision making. Time permitting, I will also discuss how to construct valid confidence intervals from data collected by contextual bandits.

Papers:
https://arxiv.org/abs/2106.00418

Bio:
Nathan Kallus is an Assistant Professor in the School of Operations Research and Information Engineering and Cornell Tech at Cornell University. Nathan's research interests include optimization, especially under uncertainty; causal inference; sequential decision making; and algorithmic fairness. He holds a PhD in Operations Research from MIT as well as a BA in Mathematics and a BS in Computer Science from UC Berkeley. Before coming to Cornell, Nathan was a Visiting Scholar at USC's Department of Data Sciences and Operations and a Postdoctoral Associate at MIT's Operations Research and Statistics group.