DISSERTATION PROPOSAL

Lin An

"Algorithms for Online Decision-Making: Nonstationary Newsvendor, Resource Allocation, and Transformer-Based Recommendations"

Friday, October 4, 2024
3:00pm
Tepper 5219

This dissertation investigates three key problems in online decision-making, focusing on the development of algorithms tailored to distinct challenges in inventory management, resource allocation, and recommendation systems. Each chapter addresses a different facet of decision-making, with a shared emphasis on practical, algorithmic solutions in dynamic environments.

In the first chapter, we tackle the Nonstationary Newsvendor problem, where sequential inventory decisions are made under unknown and nonstationary demand. We propose a policy that achieves order-optimal regret without assuming knowledge of the nonstationarity level. Additionally, we introduce a framework that incorporates generic predictions of arbitrary accuracy, demonstrating that our approach performs optimally even in the absence of prediction accuracy information. Extensive validation on real-world datasets shows that this method significantly reduces the gap between traditional models and those relying on perfect predictions.

The second chapter focuses on the Online Resource Allocation problem in the presence of uncertain arrival models. We introduce an algorithm that leverages predictions in the form of shadow prices while maintaining robustness to unknown prediction accuracy. Our solution achieves optimal performance across both stochastic and adversarial settings. We validate this approach with large-scale experiments on real-world data from H&M, demonstrating its effectiveness in resource allocation decisions when prediction accuracy is variable.

In the third chapter, we explore the application of transformer models to recommendation systems, collaborating with Glance, an AI-based software company. Our research shows that transformer models, by capturing the sequential nature of user behavior, significantly improve recommendation accuracy compared to traditional methods. Furthermore, we develop optimization techniques to ensure real-time performance, balancing speed and accuracy to meet the stringent latency requirements of large-scale, real-world applications like Glance.

Together, these chapters advance the field of online decision-making by providing robust, scalable algorithms that address critical challenges in inventory management, resource allocation, and recommendation systems.

Proposed Committee:
Andrew Li and Ben Moseley (Co-Chairs), R. Ravi, and Evelyn Gong

Proposal Documents:
Link to the first chapter: https://arxiv.org/abs/2305.07993.
Link to the second chapter: https://arxiv.org/abs/2402.13530.