Dr. Zhengling Qi

Assistant Professor

School of Business

The George Washington University

Title: **Efficient Batch Policy Learning in Markov Decision Processes**

Abstract: In this talk, I will discuss the batch (off-line) reinforcement learning problem in infinite horizon Markov Decision Processes. Motivated by mobile health applications, we focus on learning a policy that maximizes the long-term average reward. Given limited pre-collected data, we propose a doubly robust estimator for the average reward and show that it achieves statistical efficiency bound. We then develop an optimization algorithm to compute the optimal policy in a parametrized stochastic policy class. The performance of the estimated policy is measured by the difference between the optimal average reward in the policy class and the average reward of the estimated policy. Under some technical conditions, we establish a strong finite-sample regret guarantee in terms of total decision points, demonstrating that our proposed method can efficiently break the curse of horizon. Finally, the performance of the proposed method is illustrated by simulation studies.