COMPUTER SCIENCE & ENGINEERING
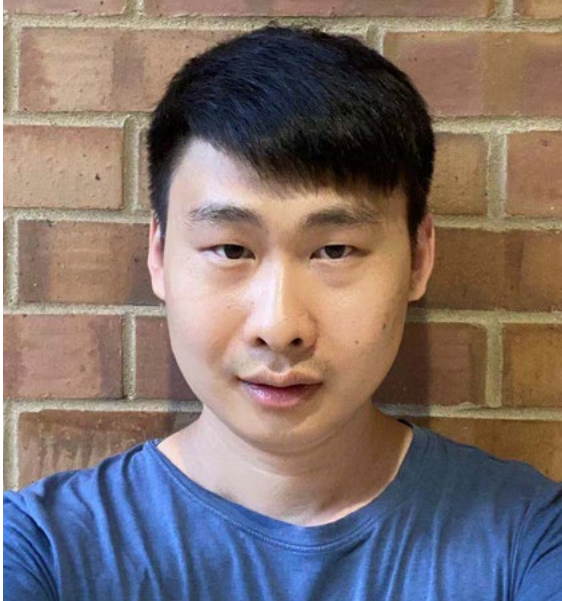## UNIVERSITY OF MICHIGAN

DISSERTATION DEFENSE

SHUANG QIU

Provably Efficient Reinforcement Learning: From Single-Agent MDPs to Markov Games

Thursday, August 5, 2021
10:00am - 12:00pm
[Virtual](#) (Passcode: 128671)

ABSTRACT: Reinforcement Learning (RL) has achieved tremendous empirical successes in real-world decision-making problems. Along with its great empirical achievements, the question of how to design efficient RL algorithms with provable theoretical guarantees has attracted increasing attention. To answer the above question, this dissertation proposes and analyzes novel RL algorithms for single-agent Markov Decision Processes (MDPs) and Markov games, where the Markov game is the multi-agent extension of single-agent MDPs. This thesis covers two paradigms in RL: reward-based online RL and reward-free RL.

Concretely, this dissertation focuses on providing a theoretical analysis of three fundamental and challenging problems in RL: (1) online learning for constrained MDPs, (2) policy optimization for Markov games, (3) reward-free RL with nonlinear function approximations. The first two problems are studied in the scope of the reward-based online RL and the third one is under the reward-free RL setting. For the first problem, this dissertation proposes a provably efficient upper confidence primal-dual algorithm for the scenario that the constraints are stochastically time-varying, the transition model is unknown, and the reward function is adversarial. It further analyzes the upper bounds of the regret and the constraint violation for learning constrained MDPs. For the second problem, this dissertation proposes new optimistic policy optimization algorithms for two-player zero-sum Markov games with structured but unknown transitions and analyzes both players' regret bounds, which generalizes the recent studies on policy optimization for single-agent MDPs in a stationary environment. For the third problem, this dissertation proposes and analyzes new reward-free exploration and planning algorithms for both single-agent MDPs and two-player zero-sum Markov games with powerful kernel and neural function approximators.

CO-CHAIRS: Prof. Jieping Ye