Chi Jin

Title: Is RLHF More Difficult than Standard RL? A Theoretical Perspective

Abstract: Reinforcement learning from Human Feedback (RLHF) learns from preference signals, while standard Reinforcement Learning (RL) directly learns from reward signals. Preferences arguably contain less information than rewards, which makes preference-based RL seemingly more difficult. In this talk, we will provably show that, for a wide range of preference models, we can solve preference-based RL directly using existing algorithms and techniques for reward-based RL, with small or no extra costs. Specifically, (1) for preferences that are drawn from reward-based probabilistic models, we reduce the problem to robust reward-based RL that can tolerate small errors in rewards; (2) for general arbitrary preferences where the objective is to find the von Neumann winner, we reduce the problem to multiagent reward-based RL which finds Nash equilibria for factored Markov games under a restricted set of policies. The latter case can be further reduced to adversarial MDP when preferences only depend on the final state.

Bio: Chi Jin is an assistant professor at the Electrical and Computer Engineering department of Princeton University. He obtained his PhD degree in Computer Science at University of California, Berkeley, advised by Michael I. Jordan. His research mainly focuses on theoretical machine learning, with special emphasis on nonconvex optimization and Reinforcement Learning (RL). In nonconvex optimization, he provided the first proof showing that first-order algorithm (stochastic gradient descent) is capable of escaping saddle points efficiently. In RL, he provided the first efficient learning guarantees for Q-learning and least-squares value iteration algorithms when exploration is necessary. His works also lay the theoretical foundation for RL with function approximation, multiagent RL and partially observable RL.