数学与系统科学研究院

计算数学所网络学术报告

报告人： 王梦迪 副教授

（ 普林斯顿大学 ）

报告题目：

Some theoretical results on model-based reinforcement learning

邀请人： 刘歆 研究员

报告时间：2020 年 11 月 20 日（周五）

上午 9:00-9:35

报告工具：腾讯会议（ID：521 3538 2330）
会议密码：311311

Abstract：

We discuss some recent results on model-based methods for reinforcement learning (RL) in both online and offline problems.

For the online RL problem, we discuss several model-based RL methods that adaptively explore an unknown environment and learn to act with provable regret bounds. In particular, we focus on finite-horizon episodic RL where the unknown transition law belongs to a generic family of models. We propose a model based 'value-targeted regression' RL algorithm that is based on optimism principle: In each episode, the set of models that are `consistent' with the data collected is constructed. The criterion of consistency is based on the total squared error of that the model incurs on the task of predicting values as determined by the last value estimate along the transitions. The next value function is then chosen by solving the optimistic planning problem with the constructed set of models. We derive a bound on the regret, for arbitrary family of transition models, using the notion of the so-called Eluder dimension proposed by Russo & Van Roy (2014).

Next we discuss batch data (offline) reinforcement learning, where the goal is to predict the value of a new policy using data generated by some behavior policy (which may be unknown). We show that the fitted Q-iteration method with linear function approximation is equivalent to a model-based plugin estimator. We establish that this model-based estimator is minimax optimal and its statistical limit is determined by a form of restricted chi-square divergence between the two policies.

<u>Bio</u>：

Mengdi Wang received her Ph.D. in Electrical Engineering and Computer Science from Massachusetts Institute of Technology in 2013. At MIT, Mengdi was affiliated with the Laboratory for Information and Decision Systems and was advised by Dimitri P. Bertsekas. Mengdi joined Princeton University in 2014. She received the Young Researcher Prize in Continuous Optimization of the Mathematical Optimization Society in 2016 (awarded once every three years), the Princeton SEAS Innovation Award in 2016, the NSF Career Award in 2017, the Google Faculty Award in 2017, and the MIT Tech Review 35-Under-35 Innovation Award (China region) in 2018. She is currently serving as an associate editor for Operations Research.

欢迎大家参加！