School of Data Science
香港城市大學
City University of Hong Kong

Hong Kong Institute for Data Science
香港城市大學
City University of Hong Kong

ONLINE SEMINAR

Variational Policy Gradient Method for Reinforcement Learning with General Utilities

Date: 7 October 2020 (Wednesday)

Time: 10am – 11am

Seminar link: https://cityu.zoom.us/j/99639389659

ABSTRACT

In recent years, reinforcement learning (RL) systems with general goals beyond a cumulative sum of rewards have gained traction, such as in constrained problems, exploration, and acting upon prior experiences. In this seminar, we consider policy optimization in Markov Decision Problems, where the objective is a general concave utility function of the state-action occupancy measure, which subsumes several of the aforementioned examples as special cases. Such generality invalidates the Bellman equation. As this means that dynamic programming no longer works, we focus on direct policy search. Analogously to the Policy Gradient Theorem \cite{sutton2000policy} available for RL with cumulative rewards, we derive a new Variational Policy Gradient Theorem for RL with general utilities, which establishes that the parametrized policy gradient may be obtained as the solution of a stochastic saddle point problem involving the Fenchel dual of the utility function. We develop a variational Monte Carlo gradient estimation algorithm to compute the policy gradient based on sample paths. We prove that the variational policy gradient scheme converges globally to the optimal policy for the general objective, though the optimization problem is nonconvex. We also establish its rate of convergence of the order O(1/t) by exploiting the hidden convexity of the problem, and proves that it converges exponentially when the problem admits hidden strong convexity. Our analysis applies to the standard RL problem with cumulative rewards as a special case, in which case our result improves the available convergence rate.

Dr Mengdi WANG

GUEST SPEAKER'S PROFILE

Dr Mengdi WANG is an associate professor at the Department of Electrical Engineering and Center for Statistics and Machine Learning at Princeton University. She is also affiliated with the Department of Operations Research and Financial Engineering and Department of Computer Science. Her research focuses on data-driven stochastic optimization and applications in machine and reinforcement learning. She received her PhD 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 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 serves as an associate editor for Operations Research and Mathematics of Operations Research, as area chair for ICML, NeurIPS, AISTATS, and is on the editorial board of Journal of Machine Learning Research. Research supported by NSF, NIH, AFOSR, Google, Microsoft C3.ai DTI, FinUP.

Enquiries: hkids@cityu.edu.hk

All are welcome