VIJAY SUBRAMANIAN, University of Michigan, Ann Arbor

Bayesian Learning of Optimal Policies in Markov Decision Processes with Countably Infinite State-Space

Models of many real-life applications—queuing models of communication networks—have a countably infinite state-space. Algorithmic and learning procedures that have been developed to produce optimal policies mainly focus on finite state settings, and do not apply to these models. To overcome this lacuna, we study the problem of optimal control of a family of discrete-time countable state-space Markov Decision Processes (MDPs) governed by an unknown parameter $\theta \in \Theta$, and defined on a countably-infinite state space $\mathcal{X} = \mathbb{Z}_+^d$, with finite action space $\mathcal{A}$, and an unbounded cost function. The random unknown parameter θ^* is generated via a given fixed prior distribution on Θ. To optimally control the unknown MDP, we propose an algorithm based on Thompson sampling with dynamically-sized episodes: at the beginning of each episode, the posterior distribution formed via Bayes' rule is used to produce a parameter estimate, which then decides the policy applied during the episode. To ensure the stability of the Markov chain obtained by following the policy chosen for each parameter, we impose ergodicity assumptions. From this condition and using the solution of the average cost Bellman equation, we establish an $\tilde{O}(\sqrt{|\mathcal{A}T})$ upper bound on the Bayesian regret of our algorithm, where T is the time-horizon. Finally, to elucidate the applicability of our algorithm, we consider two different queuing models with unknown dynamics, and show that our algorithm can be applied to develop approximately optimal control algorithms.

This is joint work with Saghar Adler at the University of Michigan, Ann Arbor.