Near-Optimal Learning and Planning in Separated Latent MDPs

Abstract

We study computational and statistical aspects of learning Latent Markov Decision Processes (LMDPs), a natural problem class of partially observable reinforcement learning. In this model, the learner interacts with an MDP drawn at the beginning of each epoch from an unknown mixture of MDPs. To sidestep known impossibility results, we consider several notions of separation of the constituent MDPs.

The main thrust of this paper is in establishing a *phase-transition* phenomena of learnability, in terms of the horizon length. For horizon length below a (nearly sharp) statistical threshold, it is impossible to learn a near-optimal policy in polynomial samples; On the other hand, sample-efficient learning can be achieved when the horizon length slightly exceeds the threshold. On the computational side, we show that under a weaker assumption of separability under the optimal policy, there is a quasi-polynomial algorithm with time complexity scaling in terms of the statistical threshold. We further show a near-matching time complexity lower bound under the exponential time hypothesis.