

Provably Efficient Offline Reinforcement Learning in Regular Decision Processes

Monday, June 17, 2024 2:30 PM (30 minutes)

We study the problem of offline (or batch) Reinforcement Learning (RL) in episodic Regular Decision Processes (RDPs). RDPs are the subclass of Non-Markov Decision Processes where the dependency on the history of past events can be captured by a finite state automaton. We consider a setting where the automaton that underlies the RDP is unknown, and a learner strives to learn a near-optimal policy using pre-collected data, in the form of non-Markov sequences of observations, without further exploration. We present RegORL, an algorithm that suitably combines automata learning techniques and state-of-the-art algorithms for offline RL in MDPs. RegORL has a modular design allowing one to use any off-the-shelf offline RL algorithm in MDPs. We report a non-asymptotic high-probability sample complexity bound for RegORL to yield an ε -optimal policy, which makes appear a notion of concentrability relevant for RDPs. Furthermore, we present a sample complexity lower bound for offline RL in RDPs. To our best knowledge, this is the first work presenting a provably efficient algorithm for offline learning in RDPs.

Primary author: Dr JONSSON, Anders (University Pompeu Fabra)

Presenter: Dr JONSSON, Anders (University Pompeu Fabra)

Session Classification: Parallel session: Challenges and progress in statistical reinforcement learning