

Nr: 35

**Reinforcement Learning in Switching Non-Stationary Markov Decision Processes:
Algorithms and Convergence Analysis**

Mohsen Amiri, Department of Computer and Systems Sciences

Posterpresentation

Abstract

Reinforcement learning in non-stationary environments is challenging due to abrupt and unpredictable changes in dynamics, often causing traditional algorithms to fail to converge. However, in many real-world cases, non-stationarity has some structure that can be exploited to develop algorithms and facilitate theoretical analysis. We introduce one such structure, Switching Non-Stationary Markov Decision Processes (SNS-MDP), where environments switch over time based on an underlying Markov chain. Under a fixed policy, the value function of an SNS-MDP admits a closed-form solution determined by the Markov chain's statistical properties, and despite the inherent non-stationarity, Temporal Difference (TD) learning methods still converge to the correct value function. Furthermore, policy improvement can be performed, and it is shown that policy iteration converges to the optimal policy. Moreover, since Q-learning converges to the optimal Q-function, it likewise yields the corresponding optimal policy. To illustrate the practical advantages of SNS-MDPs, we present an example in communication networks where channel noise follows a Markovian pattern, demonstrating how this framework can effectively guide decision-making in complex, time-varying contexts.