Review-12

Cautious Adaptation For Reinforcement Learning in Safety-Critical Settings

Real-world settings require the Reinforcement Learning (RL) agent to cautious yet optimal behavior. To that end, the work presents a Safety Critical Adaptation (SCA) framework which allows the agent to adapt to scenarios by executing a safe policy. The framework trains the agent in a non-safety critical simulation setting. Following pretraining, the agent is made to adapt safety-critical scenarios where catastrophic states penalize the agent heavily. The work further proposes a solution based on SCA framework in order to yield risk-averse policies for cautious adaptation. Cautious Adaptation in Reinforcement Learning (CARL) makes use of model-based RL to capture uncertainity and estimate risk followed by planning in sfety-critical settings to avoid catastrophic states. CARL empirically demonstrates risk-averse behavior with fewer failures in comparison to RL baselines.