Abstract

In reinforcement learning (RL) tasks, the agent learns to make good decisions by interacting with an environment. Namely, the agent perceives the state of the environment and it acts in order to maximize the cumulative return, which is based on a real-valued reward function on each step. To learn the optimal strategy/policy, the agent needs to do exploration, i.e. trying different actions to gather information of the environment, which often results in entering unsafe or undesirable states. In this case, the desired policy should not only seek for maximization of the expected cumulative return from the environment, but also take the safety concerns during the process into consideration. In this project, we use the Markov Decision Processes (MDPs) to model the decision-making progress and consider constrained MDPs with safety constraints. We aim to design an RL algorithm, based on the modification of the optimality criterion, to learn the policy that performs well in the original RL task while not violating the safety constraints specified.