

IBM Research AI



Safe Reinforcement Learning in Constrained Markov Decision Processes

Akifumi Wachi

IBM Research AI

Yanan Sui

Tsinghua University

ICML | 2020

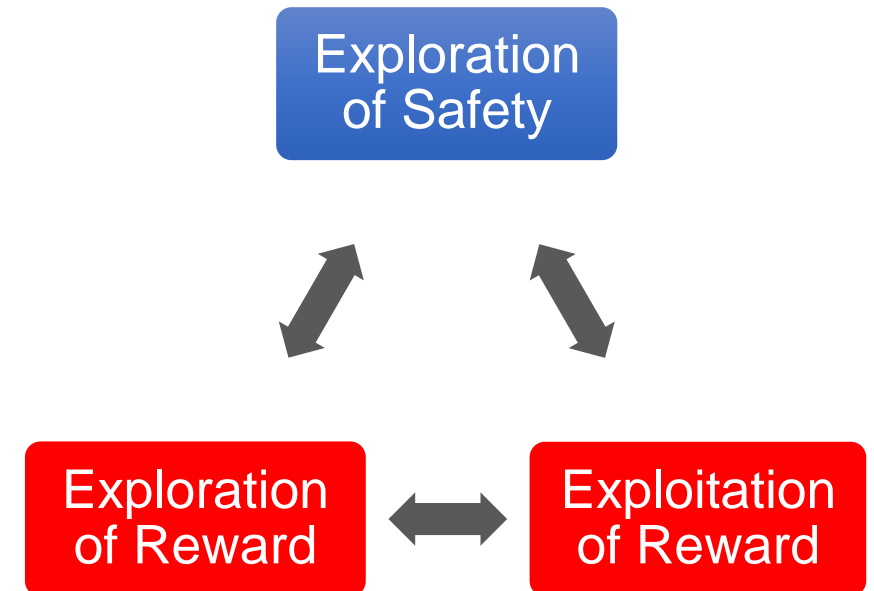
Thirty-seventh International
Conference on Machine Learning

Safe Reinforcement Learning in Constrained Markov Decision Processes

- Consider a safety-constrained Markov Decision Processes (MDPs).
 - Both reward and safety are *unknown a priori*.
 - Our objective is to **maximize the cumulative reward** while **guarantee safety**.

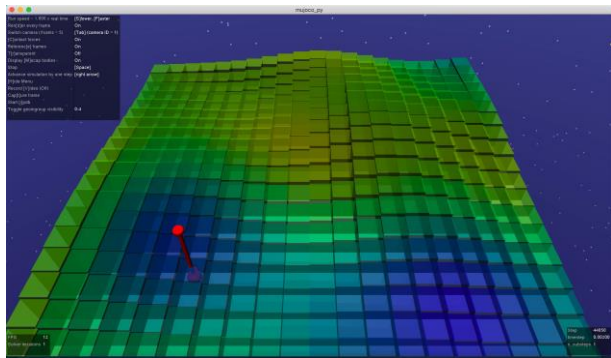
Safe Reinforcement Learning in Constrained Markov Decision Processes

- Consider a safety-constrained Markov Decision Processes (MDPs).
 - Both reward and safety are *unknown a priori*.
 - Our objective is to **maximize the cumulative reward** while **guarantee safety**.
- To solve this problem, we need to balance the three-way tradeoff.
- This work takes a **step-wise approach**.
 1. **Exploration of safety**.
 2. **Optimization of the cumulative reward in the certified safe region**.
- Our algorithm provides theoretical guarantees in terms of both **near-optimality** and **safety**.



Safe Reinforcement Learning in Constrained Markov Decision Processes

- Developed a new simulation environment, called **GP-Safety-Gym**, which is based on Open AI SafetyGym (Ray et al., 2019).
- **Achieved better empirical performance** than other baselines.
 - SafeMDP (Turchetta et al., 2016)
 - SafeExpOpt-MDP (Wachi et al., 2018)
- Also proposed **Early-Stopping of Exploration of Safety (ES²)** algorithm for faster convergence.



Reward (high: **yellow**, low: **blue**)
Safety: height

