Provably Robust and Well-Behaved Reinforcement Learning

Tengyang Xie

tx10@illinois.edu

Motivations

- RL has been used in many *safety-critical* applications (e.g., medical treatment design, autonomous driving).
- The observation from real-world applications may contain unavoidable uncertainty.

Goal: Robust RL against (adversarial) perturbations on state observations.

- Rigorous: *Provable* guarantees on performance and robustness.
- Practical: The algorithm is implementation-friendly.

Challenges

On Theoretical:

- Define "robustness" from the first principles.
  For RL, what should be the analog of certified radius[1] given the most basic assumptions (e.g., the Lipschitz condition of MDPs)?
- Bound the performance loss (induced by robustness) with explicit dependence on model/function/robustness conditions.
- Sequential error propagation due to the MDP model.

On Empirical:

- The algorithm might need to optimize over the whole perturbation set with complicated optimization procedures (e.g., [Zhang et al., 2020]).

[1] In classification problems, the concept of certified radius can be derived straightforwardly from Lipschitz + margin.

Proposed Approach: RL with Randomized Smoothing

Idea: Leverage randomized smoothing [Cohen et al., 2019] using an extended value iteration (EVI) style approach.

An example FQI/DQN-style approach:

$$(\text{For } t \in [T]) \quad Q_{t+1} \leftarrow \operatorname*{argmin}_{Q \in \mathcal{Q}} \left\| Q(s,a) - r - \max_{a'} \textsc{Smooth}[Q_t](s',a') \right\|_{2,\text{data}}^2$$

Output $\textsc{Smooth}[Q_T]$.

Q: Why EVI rather than use a smoothed Q-function/policy directly?

A: EVI-style approaches could ensure *Bellman consistency*, which is conjectured to have better performance loss. (will make it formal)

Reference

Jeremy Cohen, Elan Rosenfeld, and Zico Kolter. Certified adversarial robustness via randomized smoothing. In *International Conference on Machine Learning*, pages 1310–1320, 2019.

Huan Zhang, Hongge Chen, Chaowei Xiao, Bo Li, Duane Boning, and Cho-Jui Hsieh. Robust deep reinforcement learning against adversarial perturbations on observations. *arXiv preprint arXiv:2003.08938*, 2020.