Robust and Safe RL

Amirabbas Afzali[1] and Amirreza Velae[2], [1]amirrezavelae@gmail.com, [2]amir8afzali@gmail.com.

Proposal: Reinforcement learning (RL) has emerged as a highly effective framework for solving Markov decision processes (MDPs). As RL is increasingly applied to complex and real-world environments, ensuring the robustness of solutions becomes crucial. However, current RL algorithms often struggle with robustness against uncertainties, disturbances, and structural changes in the environment.

Prior robust RL methods typically only study noise introduced independently across time[1]. These methods mostly categorize the perturbations into four main types: (i) Transition robust designs, which address uncertainties in system dynamics by adjusting state transition probabilities; (ii) Disturbance robust designs, which incorporate external forces to model system behavior uncertainties; (iii) Action robust designs, which mitigate the impact of corrupted agent outputs on system transitions; and (iv) Observation robust designs, which handle distortions in the perceived system state. Each of these approaches modifies a different aspect of the MDP framework. These four categories of robustness are mostly considered from the perspectives of Optimal Control (H_∞ *control* and *Belza form*), Optimization (*bi-level* problem) and Game Theory (*adversarial reinforcement learning*).

We consider the environment (or, as many other papers call it, nature) to be adversarial towards the agent, attempting to minimize the agent's reward. This perspective leads to common worst-case scenario cases where the optimization problem is converted to a minimax problem. Consequently, many authors view this optimization as a zero-sum two-player game and seek to find the Nash equilibria[5]. However, this viewpoint does not generalize for most cases. For example, in Multi-Agent Reinforcement Learning (MARL), the game involves more than two players, making the Nash equilibria NP-hard to find. Additionally, preparing for the worst case is not necessarily beneficial, as the agent may overlook some valuable rewards.

Safe reinforcement learning ensures agents avoid harmful actions during both training and deployment by integrating safety constraints[2]. Reinforcement Learning with Human Feedback (RLHF) [3] can enhance safe RL by using human input to guide the agent's behavior, ensuring it aligns with safety standards and avoids risky actions[4].

In some advances papers, authors consider editing *Hamilton-Jacobi-Bellman equation* (HJB) which is derived from H_∞ control and *Belza Form*. In robust RL the Bellman operator slightly modified under some uncertainty or perturbation sets.

Therefore, research on this topic is ongoing, and authors are attempting to address this problem in various ways. For example, authors in [6], proposed the first policy-based approach, Robust Natural Actor-Critic (RNAK), for robust RL under function approximation. Additionally, they proposed two novel uncertainty sets to improve sim-to-real approaches.

References:

1. https://www.mdpi.com/2504-4990/4/1/13
2. https://arxiv.org/pdf/2403.18539
3. https://arxiv.org/pdf/2310.12773
4. https://arxiv.org/pdf/2106.05091
5. https://arxiv.org/pdf/2307.12062
6. https://arxiv.org/abs/2307.08875