

Actor Prioritized Experience Replay (Abstract Reprint)

Baturay Saglam¹, Furkan Mutlu², Dogan Cicek², Suleyman Kozat²

¹Department of Electrical Engineering, Yale University, New Haven, CT, USA

²Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey

Abstract Reprint. This is an abstract reprint of a journal article by Saglam, Mutlu, Cicek, and Kozat (2023).

Abstract

A widely-studied deep reinforcement learning (RL) technique known as Prioritized Experience Replay (PER) allows agents to learn from transitions sampled with non-uniform probability proportional to their temporal-difference (TD) error. Although it has been shown that PER is one of the most crucial components for the overall performance of deep RL methods in discrete action domains, many empirical studies indicate that it considerably underperforms off-policy actor-critic algorithms. We theoretically show that actor networks cannot be effectively trained with transitions that have large TD errors. As a result, the approximate policy gradient computed under the Q-network diverges from the actual gradient computed under the optimal Q-function. Motivated by this, we introduce a novel experience replay sampling framework for actor-critic methods, which also regards issues with stability and recent findings behind the poor empirical performance of PER. The introduced algorithm suggests a new branch of improvements to PER and schedules effective and efficient training for both actor and critic networks. An extensive set of experiments verifies our theoretical findings, showing that our method outperforms competing approaches and achieves state-of-the-art results over the standard off-policy actor-critic algorithms.

References

Saglam, B.; Mutlu, F.; Cicek, D.; and Kozat, S. 2023. Actor Prioritized Experience Replay. *Journal of Artificial Intelligence Research*, 78: 639–672.