

Research Article

Averaged Soft Actor-Critic for Deep Reinforcement Learning

Feng Ding, Guanfeng Ma, Zhikui Chen, Jing Gao, and Peng Li

School of Software Technology, Dalian University of Technology, Dalian, China

Correspondence should be addressed to Jing Gao; gaojinghit@gmail.com

DW W [h W # (@ a h W T V d \$ " \$ - D W [e W # 8 W d j S c k \$ " \$ # - 3 U W d f W # * ? S d L Z \$ " \$ # - B g T [e Z W # 3 b d [^ \$ " \$ #

Academic Editor: Ning Cai

Copyright © 2021 Feng Ding et al. is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the advent of the era of artificial intelligence, deep reinforcement learning (DRL) has achieved unprecedented success in high-dimensional and large-scale artificial intelligence tasks. However, the insecurity and instability of the DRL algorithm have an important impact on its performance. The Soft Actor-Critic (SAC) algorithm uses advanced functions to update the policy and value network to alleviate some of these problems. However, SAC still has some problems. In order to reduce the error caused by the overestimation of SAC, we propose a new SAC algorithm called Averaged-SAC. By averaging the previously learned action-state estimates, it reduces the overestimation problem of soft Q-learning, thereby contributing to a more stable training process and improving performance. We evaluate the performance of Averaged-SAC through some games in the MuJoCo environment. The experimental results show that the Averaged-SAC algorithm effectively improves the performance of the SAC algorithm and the stability of the training process.

1. Introduction

To generate fully autonomous agents which can learn to automate behaviors by interacting with the experimental environment via trials and errors is one of the most important tasks in the current artificial intelligence field. In the current artificial intelligence field, the long-term challenge is to create an intelligence system responding to environments in a timely manner. Such intelligence systems include robots that can interact with the surrounding environments and software-based agents that can interact with multimedia devices. Currently, deep reinforcement learning (DRL) whose mathematical framework is the experience-driven autonomous learning is the most important algorithm to address these challenges [1]. For example, the Google AlphaGo defeated the world champion in the Go game. To complete the breakthrough in this field, there are still a lot of work to do. Among them, ensuring the safety of decision making is one of the most important challenges. Because the agent is more inclined to explore unfamiliar states, deep reinforcement learning will be susceptible to the so-called security exploration problem, causing the agent to be in an unsafe state (for example, a mobile robot drove the car into the ditch). The safe intelligence system should ensure the

safety of the controlled objects and reduce the probability of dangerous operations. Since 2018, a series of unmanned vehicle safety accidents have occurred in the United States, highlighting the importance of safety in automatic control tasks. However, many current artificial intelligence methods do not fully control risks. Furthermore, some methods deliberately add exploratory learning with random nature to the solution process. The exploratory learning without security restrictions is likely to bring risks. If the agent directly applies the reinforcement learning method for "trial and error" exploration and learning in real-world tasks, the decision made by the agent may put the system into a dangerous state. The security of deep reinforcement learning has attracted more and more attention. Therefore, to improve the safety and address the noncontrol problems in practical applications, we should find ways to reduce the propagation error in the neural network. At the same time, the stability of the DRL algorithm is a big challenge, which limits the further development of the algorithm. Although the performance of the computer has been greatly improved, the stability of DRL cannot be guaranteed. Therefore, how to improve the security and stability of DRL algorithms for a large number of artificial intelligence learning is one of the most challenging problems in DRL today.

