FSDM3503

FSDM 2020
November 13-16, 2020
Online Conference

Reward-free Reinforcement Learning Algorithm Using Prediction Network

Zhen Yu[a,b,1], Yimin Feng[a] and Lijun Liu[a,b]
[a] School of Aerospace Engineering, Xiamen University, Xiamen 361005, China
[b] Shenzhen Research Institute of Xiamen University, Shenzhen 518000, China

Introduction

In general reinforcement learning (RL) tasks, the formulation of reward functions is a very important. The reward function is not easy to formulate in a large number of systems. The network training effect is sensitive to the reward function, and different reward value functions will get different results. For a class of systems that meet specific conditions, the traditional RL method is improved. A state quantity function is designed to replace the reward function, which is more efficient than the traditional reward function. Finally, the algorithm was successfully applied in the environment of FrozenLake, and achieved good performance. The experiment proves the effectiveness of the algorithm and realizes rewardless RL in a class of systems.

Methods

Aiming at a class of reinforcement learning problems defined in Chapter 3, this paper designs a reinforcement learning algorithm RFPG with no reward value. This method uses state quantities to provide learning directions for the network. The state quantity can distinguish the target state, failure state and general state of the agent. Experiments in various environments, the experimental results show that in an environment that meets certain conditions, the reinforcement learning method without reward value can still achieve control of the system. In the future, the application scenarios can be further expanded in the research. The RFPG algorithm is not only used for solving fixed target problems, but also for solving optimal state problems

Graphics / Images

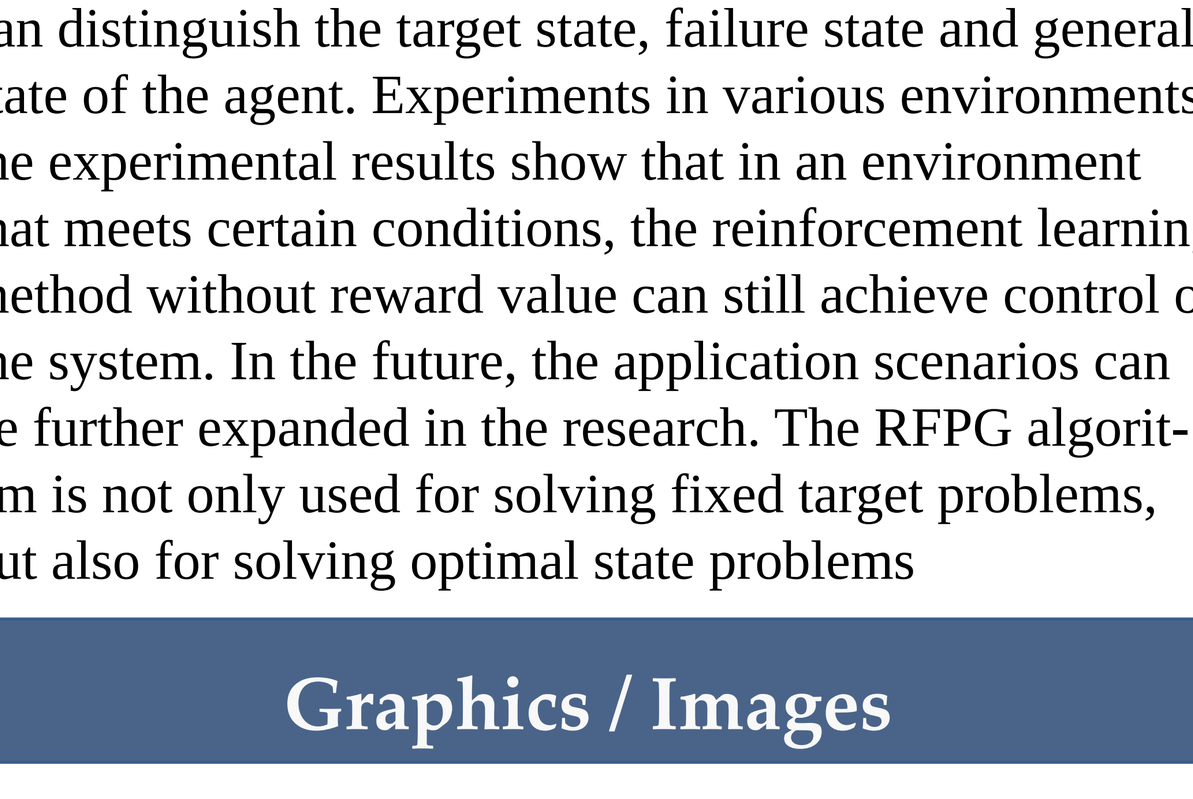

Figure 1 RFPG basic structure

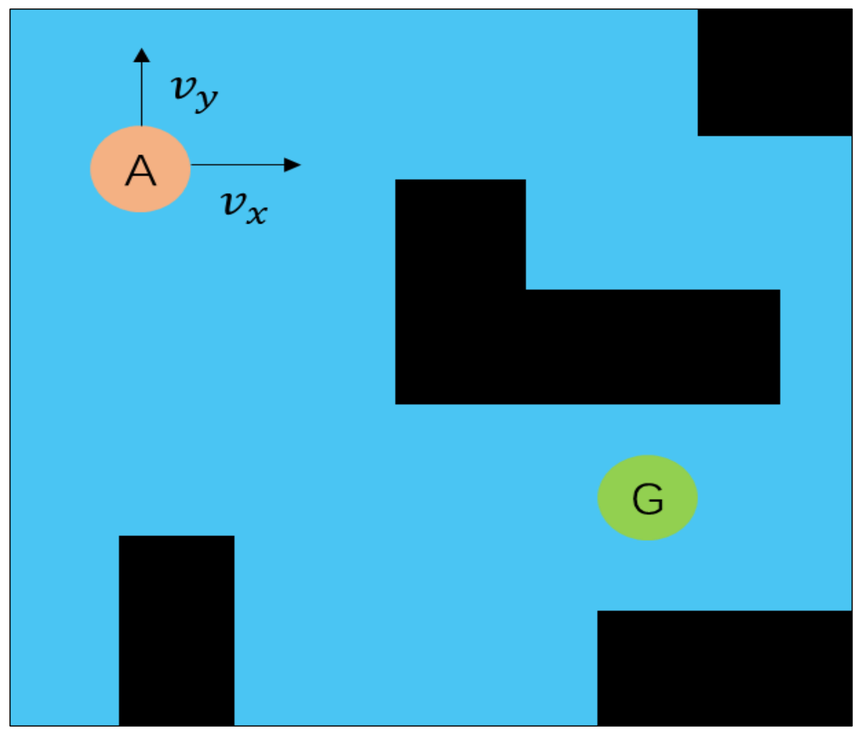

Figure 2 FrozenLake environment

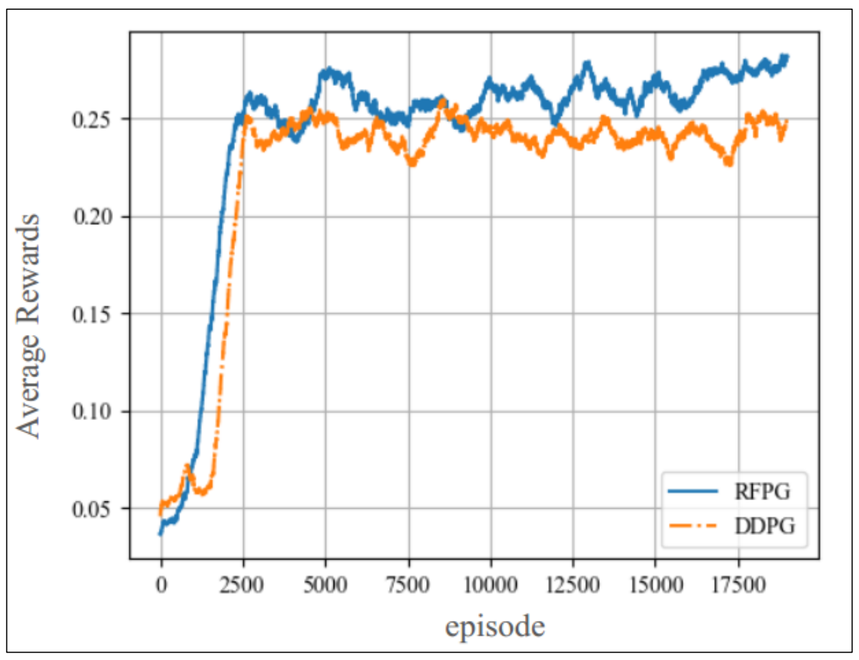

Figure 3 Frozen Lake training process

Conclusions

This paper designs a reinforcement learning algorithm with no reward value called RFPG (Reward Free Policy Gradient) algorithm. The RFPG algorithm includes three networks, as shown in Figure 1,prediction network P, action network A, and value network G

As shown in Figure 2, the environment is a continued state version of FrozenLake in Gym. The environment is a square ice surface with side length n, Agent glides on the ice surface. The starting point of the agent is marked A, and the target point is marked G. Dangerous areas exist on the ice, which are represented as black areas in the figure. Entering the black area indicates that the operation failed.

Figure 3 shows the trend of average returns in the Frozen Lake environment. The performance of the RFPG algorithm in the Frozen Lake environment is superior to the DDPG algorithm. It shows that the RFPG algorithm can still perform quite well without return value