2021 International Conference on Applied Mathematics, Modeling and Computer Simulation

CS163

中国南方电网 CHINA SOUTHERN POWER GRID

Robot Path Planning Based On Improved Reinforcement

Jian Fang, Kuang Yin, Hongbin Wang, Wenxiong Mo and Fan Yang

Email: fj1261262021@126.com

Guangzhou Power Supply Bureau Co. , Ltd. Guangzhou, China

INTRODUCTION

Mobile robot path planning means that in an environment with obstacles, the mobile robot finds an smooth and collision-free path from the start point to the end point according to a given task or condition. The main goal of path planning is to find the best route based on performance indicators such as distance, time and energy when the robot is in an obstacle environment. Based on the understanding of the environment, path planning is divided into two research directions: global path planning based on environmental a priori complete information and local path planning based on unknown environment. In practical applications, mobile robots need to have the ability to adapt to unknown environments, therefore, solving the problem of robot path planning in unknown environments has great significance to the application and popularization of robot technology, which is the premise and foundation of various application researches of mobile robots.

MAIN BODY

1、Principles of reinforcement learning algorithm

Q-learning is a reinforcement learning algorithm based on value iteration to learn action strategies. The algorithm uses the Q function to find the optimal action-selection strategy, and its core is to continuously update a table which composed by state, action and reward. The basic calculation formula of the algorithm is as follows:

$$Q^*(s,a) = R(s,a) + \gamma \sum T(s,a,s_t) \max Q^*(s_t,a_t)$$

2、Reinforcement Learning Path Planning Based on ε-decreasing

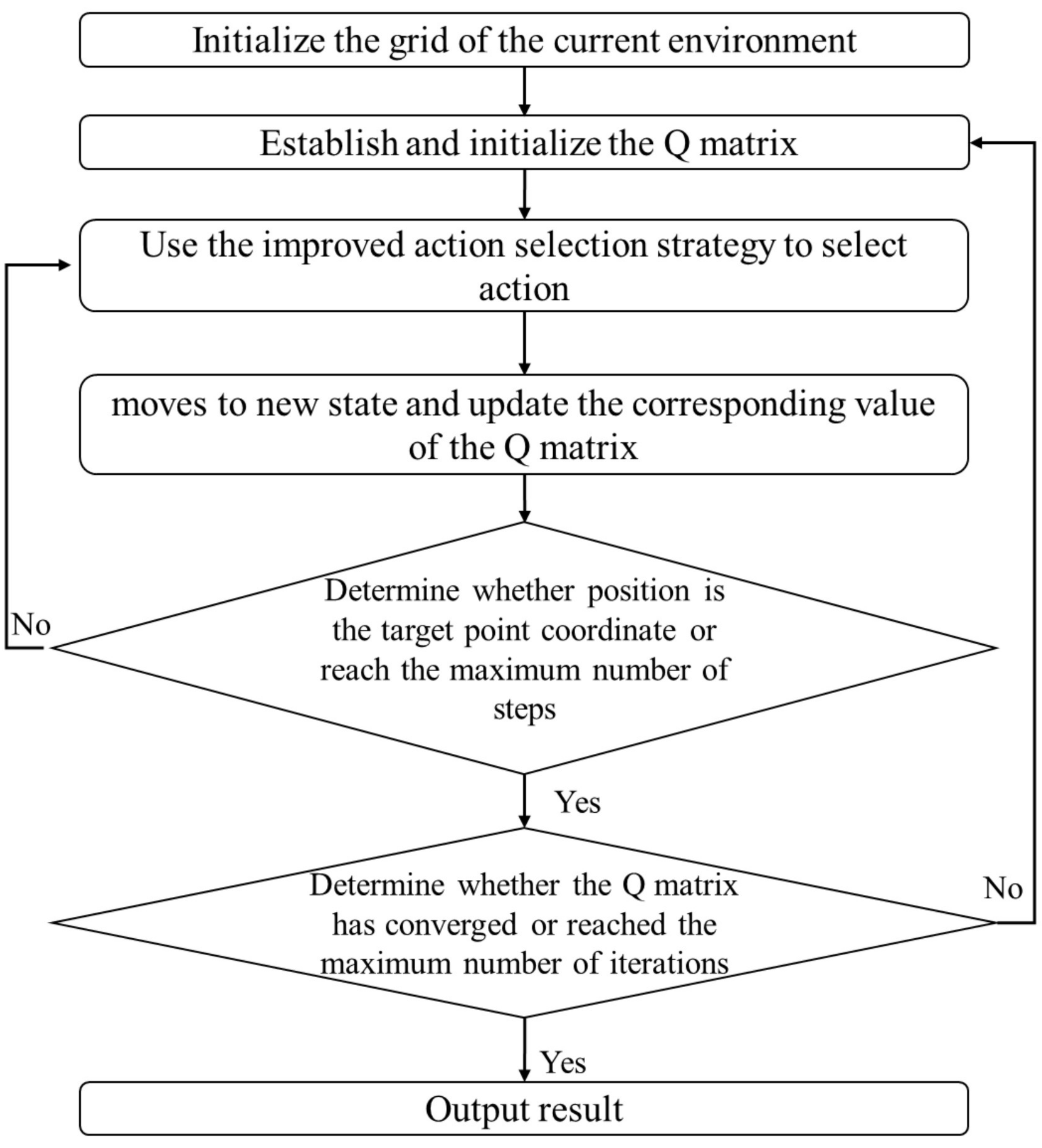
At the beginning, since the agent is facing an unfamiliar environment, it is necessary to explore the environment through random actions as much as possible. As the degree of mastery of environmental information continues to increase, the agent gradually reduces the exploration of the environment. Instead, the agent choose the most beneficial action to bring greater returns. Therefore, it is hoped that ε will maintain a downward trend along with the understanding of the environment throughout the operation of the algorithm, which is called ε-decreasing:

$$\varepsilon = e^{-\left|\frac{Q(S_t,a_b)-Q(S_t,a_t)}{n}\right|}$$

the learning rate determines how much information the agent learns from the environment after each action. However, during the operation of the algorithm, the learning rate of every action must not be the same. When exploring the environment, it is necessary to learn more about the environment, so set a larger learning rate, while when using the environment, set a larger learning rate to prevent falling into the local optimum, so the is set as follows:

$$\alpha = \begin{cases} 0.7 & \text{if } \varepsilon_0 < \varepsilon \\ 0.1 & others \end{cases}$$

3、Calculation Steps



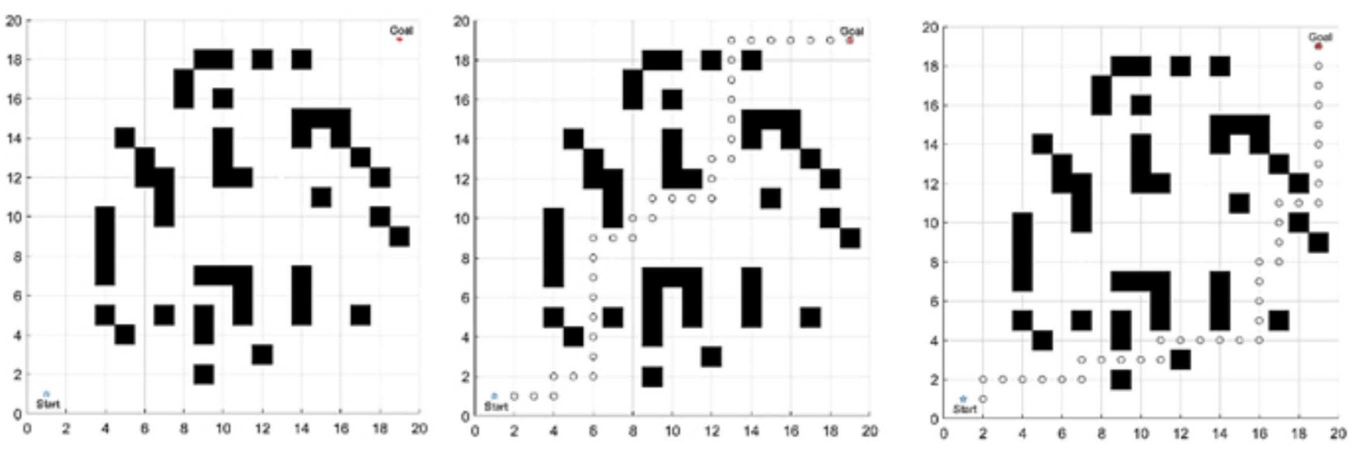
4、Calculation parameters and results

In order to verify the feasibility of the path planning method based on reinforcement learning proposed in this paper, MATLAB is used to simulate it. Table 1 shows the parameters needed for the basic Q-learning algorithm.

Parameter	Value
Map size	20*20
Reward value	100
Penalty value	-100
Learning rate	0.7
Exploration factor	1
Discount factor	0.9
Maximum step size	2000
The maximum number of iterations	1000

Tab.1 Basic Q-learning algorithm parameters

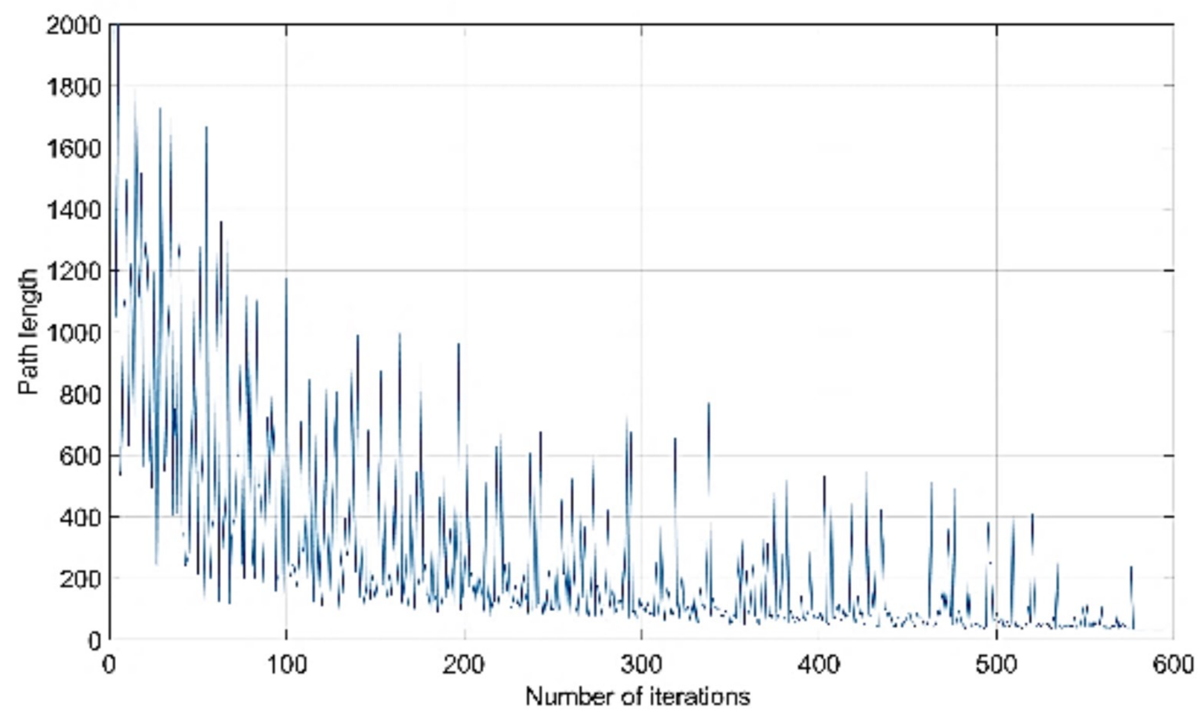
The results as follows, as we can see from Figure 1(b) and (c), both methods can successfully plan robot path in an unknown environment. Although the planned routes are different, the path length is 36, which shows that both methods can find the optimal path in this environment.



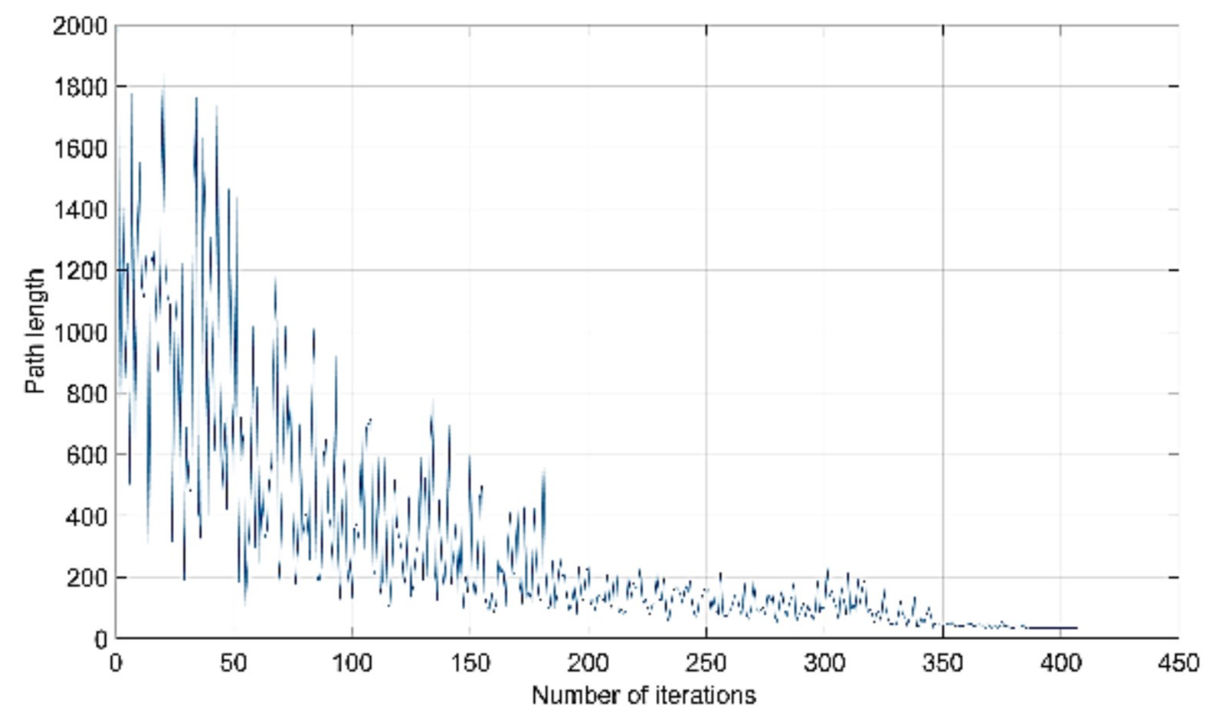
(a) (b) (c)

Fig.1 (a)simulation environment; (b) basic Q-learning algorithm path planning; (c) improved Q-learning algorithm path planning.

In order to further compare the performance of the two methods, Figure 2(a) and (b) show the step size convergence of the two methods. It can be seen that the path length jitter of the two reinforcement learning methods is large at the beginning of the iteration. As the number of iterations increases, the path length shows a downward trend, but the basic Q-learning algorithm still has more frequent and severe jitter in the later stage of convergence, while the improved Q-learning algorithm converges gently.



(a)



(b)

Fig.2 (a) basic Q-learning algorithm path planning; (b) improved Q-learning algorithm path planning.

CONCLUSIONS

Aiming at the problem of mobile robot path planning in an unknown environment, this paper proposes a dynamic action selection strategy based on the traditional Q-learning algorithm. This strategy is guided by the degree of mastery of environmental information, adaptively choose to explore the environment or use the environment, and dynamically set the learning rate according to the state of the agent, effectively improve the situation where the Q-learning algorithm is easy to fall into the local optimum. Compared with the traditional Q-learning algorithm, the method proposed in this paper has a faster convergence speed and shows better path planning performance under the premise of ensuring the optimal path.