Path planning using deep reinforcement learning based on potential field in complex environment

Qingxuan Jia¹, Maonan Yang¹*, Yu Miao¹ and Xulong Li¹

¹ School of Automation, Beijing University of Posts and Telecommunications, Beijing, 100876, China
*Corresponding author’s e-mail: ymn@bupt.edu.cn

Abstract. This paper introduces a deep reinforcement learning path planning method based on potential field for complex environment. Based on the potential field model in the artificial potential field method, we define states, actions, rewards in reinforcement learning, and use Deep Deterministic Policy Gradient (DDPG) reinforcement learning algorithm for optimization. By training robots in the environment, our method can effectively plan the path in a complex environment with massive obstacles, and avoid trapping in the local minimum region of the potential field.

1. Introduction
The purpose of path planning is to derive a path from the starting position to the target position that does not collide with obstacles in the environment. As a classic path planning method, artificial potential field has been widely used in robot navigation[1]. However, the traditional artificial potential field itself has the shortcomings that it is easy to fall into the local minimum of the potential field and cannot work under the environment of a large number of obstacles[2].

In order to optimize the local minimum of the artificial potential field, Cheng Z [2] introduced the direction vector of the robot to adjust the calculation of repulsion, Ding J [3] improved the gravitational potential field and used the method of pre-planning within the feasible solution area, Ren Y [4] proposed a comparison threshold and established a virtual traction point to solve the local minimum problem. In fact, the above work generally handles the potential field manually, so that the robot escapes from the local minimum.

However, with the development of reinforcement learning, in addition to traditional reinforcement learning such as Q-Learning and Sarsa, the powerful nonlinear fitting ability of deep neural networks provides opportunities for reinforcement learning to solve more complex problems[5], such as Deep Q Network [6], DDPG [7] etc. Deep reinforcement learning has made major breakthroughs. For example, it has achieved superhuman level in Atari games. AlphaGo defeated the world champion of humanity by using hybrid deep reinforcement learning system[8].

In this paper, we consider introducing the powerful nonlinear fitting of neural networks in deep reinforcement learning into the artificial potential field, and propose a DDPG path planning method based on the potential field. Firstly, we model the artificial potential field in a robot working environment with a large number of obstacles. Secondly, for the properties of the potential field, define a continuous action, a continuous state space and a reward function. Finally, we use DDPG reinforcement learning algorithm to train the agent to obtain excellent decision-making ability. The experimental results show that the proposed method can be applied not only in a complex environment...
with a large number of obstacles, but also to solve the local minimum problem of the artificial potential field method.

2. Method
For the convenience of mathematical definition, we assume that the robot is in the square working space of $[-1, 1]$, there are 15 random obstacles with a diameter of 0.03 in the space, and the target position is $(0.5, 0.5)$. We first model the potential field in the environment, and then define the reinforcement learning framework based on the potential field.

2.1. Modeling the potential field environment
Based on the traditional artificial potential field method, we model the gravitational potential field and the repulsive potential field in the workspace of the robot. The gravitational potential field at the target position is as in equation (1),

$$U_{gr}(p_r) = 0.5 \alpha d(p_r, p_g)$$

(1)

where $\alpha$ is the gain gravitational coefficient, $p_r$ is the position of the robot, $p_g$ is the target position, and $d(p_r, p_g)$ is the distance from the robot to the target position.

Model the repulsive potential field at all $m$ obstacle positions, where the repulsive potential field of the $m$-th obstacle is as in equation (2),

$$U_{re}^{(m)}(p_r) = \begin{cases} \frac{1}{2} \beta \left( \frac{1}{d(p_r, p_o^{(m)})} - \frac{1}{d_0} \right), & \text{if } d(p_r, p_o^{(m)}) \leq d_0 \\ 0, & \text{if } d(p_r, p_o^{(m)}) > d_0 \end{cases}$$

(2)

where $\beta$ is the gain coefficient of the repulsive potential field, $d(p_r, p_o^{(i)})$ is the distance from the robot to the $i$-th obstacle, and $d_0$ is the threshold of the repulsive potential field.

Accumulating the gravitational potential field and all repulsive potential fields in the environment, we get the total potential field, as in equation (3),

$$U(p_r) = U_{gr}(p_r) + \sum_{m} U_{re}^{(m)}(p_r)$$

(3)

Figure 1. Visualization of the potential field values of gravitational potential field, repulsive potential field and total potential field.

We set $\alpha = 0.1$, $\beta = 10$ and $d_0 = 0.07$ in the potential field, calculate the gravitational potential field and repulsive potential field separately, and add them up. The visualization of gravitational potential field, repulsive potential field and total potential field is shown in Figure 1. In the working
environment of the robot, the potential field value at the target position is the smallest, and the potential field value farther away from the target position is larger. In the potential field near the obstacle, the closer to the obstacle, the larger the potential field value.

2.2. Definition of reinforcement learning based on potential field

According to the potential field modeled above, we define action space, state space and reward function in reinforcement learning. In the problem of path planning, we expect the robot to be as close as possible to the target position and as far away from obstacles as possible at each step. Therefore, we define the reward value as two parts, including the reward given by the obstacle and the reward given by the target position, such as equation (4),

$$ r = r_{gr} + r_{re} $$  

(4)

where $ r_{gr} = -d(p_r, p_o) $, indicates that the farther the robot is from the target location, the less rewards it will receive. $ r_{re} $ is defined as equation (5),

$$ r_{re} = \sum_{m} d_{re}(p_r, p_{o}^{(m)}) $$  

(5)

where $ d_{re}(p_r, p_{o}^{(m)}) $ is defined as equation (6),

$$ d_{re}(p_r, p_{o}^{(m)}) = \begin{cases} d(p_r, p_{o}^{(m)}), & \text{if } d(p_r, p_{o}^{(m)}) < d_0 \\ 0, & \text{else} \end{cases} $$  

(6)

indicates that when the robot is in the potential field area of the obstacle, the closer the distance to the obstacle, the less the reward value the robot gets.

In order to fully consider the current robot's motion state, gravitational potential field, and all repulsive potential fields, in addition to all potential fields, we introduce the robot's velocity $ \vec{v} $ in the state $ \vec{s} $, as equation (7),

$$ \vec{s}(p_r) = [\vec{F}_{gr}(p_r), \sum_{m} \vec{F}_{re}^{(m)}(p_r), \vec{v}] $$  

(7)

where $ \vec{F}_{gr}(p_r) = \nabla U_{gr}(p_r) $, $ \vec{F}_{re}^{(m)}(p_r) = \nabla U_{re}^{(m)}(p_r) $.

We define the action in reinforcement learning as the distance difference between the next position of the robot and the current position. The action is a 2-dimensional continuous vector, where the two dimensions represent the displacement of the robot along the x-axis and y-axis respectively, as equation (8),

$$ a = [\Delta x, \Delta y] $$  

(8)

Considering that the above action space and state space are high-dimensional continuous spaces, the deep deterministic policy gradient is a method with outstanding performance for high-dimensional action space and state space in reinforcement learning. Therefore, this paper uses DDPG as a reinforcement learning algorithm to optimize the policy.

In the optimization process of DDPG, it contains a main network and a target network as shown in Figure 2. The Actor network in the main network is responsible for the robot's decision-making, where the input is a state vector and the output is an action vector. The Critic network in the main network is responsible for evaluating the value function of the current state and action. The Actor network and Critic network in the target network are responsible for calculating the target value of the value function. By calculating the mean squared difference between the value function output by the Critic network and the target value of the value function as the loss function, the back propagation of the Critic network is performed to update the parameters. Actors of the main network are updated using
the strategy gradient method. To ensure the stability of the target value, the weight of the target network uses the soft update method. The details of the experiment and results are in Section 3.

![Figure 2. Relation and update process of neural network in DDPG.](image)

3. Experiments and results

3.1. Parameter setting and training process
Details of experimental parameters:
- Gain coefficient of potential field $\alpha=0.1$, $\beta=10$
- The Actor network and Critic network in DDPG are 3-layer neural networks
- The discount factor during the update process is 0.95
- The training learning rate is set to 0.01, batch size is 1024
- When the distance between the robot and the target position is less than 0.01, it is regarded as reaching the target position
- In order to prevent the distance of each step of the robot from being too large, set the maximum distance of each step to 0.05

![Figure 3. This is the curve of the mean episode reward during training.](image)
We train the robot 10,000 episodes in the virtual environment. In each episode, the robot's starting position, target position, and the position of all obstacles are random. The condition for the termination of the episode is that the robot reaches the target position or the number of steps of the episode reaches 150. The curve of the average reward value per episode during the training process is shown in Figure 3. It can be seen that when the training reaches about 2,000 episodes, the robot obtains exceptional decision-making ability, so that the average reward value tends to be stable.

3.2. **Obstacles from few to massive**

As shown in the experimental results in Figure 4, when there are few obstacles in the environment, the algorithm can smoothly implement path planning. When there are many obstacles, although the path planned by the algorithm is tortuous, it can still derive a path that does not collide with obstacles. During the experiments, there is no local minimum defect of the traditional potential field method.

![Figure 4. Experiments in the environment of 15, 50, 100, and 150 randomly distributed obstacles.](image)

3.3. **Discussion of local minimum**

In order to prove that compared with the traditional artificial potential field method, the deep reinforcement learning path planning method based on the potential field can avoid falling into the local minimum region of the potential field. As shown in Figure 5, we do an experiment under the environment of a large number of obstacles that are extremely prone to local minimum problems.
Figure 5. Compared with the traditional artificial potential field method, due to the use of a trained deep neural network for decision-making, our method comprehensively considers the speed of the robot and the current position of the potential field state, cannot be trapped in the local minimum area, and can be in a large number of obstacles working in a natural environment.

4. Conclusion

We use the Deep Deterministic Policy Gradient to optimize the strategy of the robot's path planning in the potential field environment, propose a deep reinforcement learning path planning method based on the potential field, which can face the complex work environment where the traditional artificial potential field method is powerless. Our algorithm uses the nonlinear ability of the neural network to reasonably avoid the local minimum region of the potential field, and can even plan a path that does not collide with obstacles in complex environments.

References

[1] Zhu D Q, Yan M Z. Survey on technology of mobile robot path planning. (2010) Control and Decision, (7): 961-967.
[2] Cheng Z, Zhang Z A, Li J Z, Jiang T. (2019) Mobile robots path planning based on improved artificial potential field. Computer Engineering and Applications, 55: 29–34.
[3] Ding J R, Du C P, Zhao Y, Yin D Y. (2016) Path planning algorithm for unmanned aerial vehicles based on improved artificial potential field. Journal of Computer Applications, 2016, 36(1): 287-290.
[4] Ren Y, Zhao H B. (2020) Improved robotic path planning based on artificial potential field method. Computer Simulation, 37(02): 360-364.
[5] Arulkumaran K, Deisenroth M P, Brundage M, et al. (2017) A Brief Survey of Deep Reinforcement Learning. arXiv preprint arXiv:1708.05866.
[6] Mnih V, Kavukcuoglu K, Silver D, et al. (2015) Human-level control through deep reinforcement learning[J]. nature, 518(7540): 529-533.
[7] Lillicrap T P, Hunt J J, Pritzel A, et al. (2015) Continuous control with deep reinforcement learning[J]. arXiv preprint arXiv:1509.02971.
[8] Silver D, Huang A, Maddison C J, et al. (2016) Mastering the game of Go with deep neural networks and tree search. nature, 529(7587): 484-489.