An Intelligent Path Selection Algorithm Based on Deep Reinforcement Learning

Mengchen Sun
College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, 211006, China
1175996497@qq.com

Abstract: Path selection is the most important algorithm in intelligent devices such as robots. At present, the traditional path-planning algorithm has achieved some results, but it lacks the ability of environmental perception and continuous learning. In order to solve the above problems, this paper proposes an intelligent path selection algorithm based on deep reinforcement learning, which uses the learning ability of deep learning and the decision-making ability of reinforcement learning to realize the autonomous path planning of robots and other equipment. Simulation results show that the proposed algorithm has faster convergence, efficiency and accuracy.

1. Introduction

As an emerging technology, mobile robot technology is the product of combining key technologies of multiple disciplines, which plays a very important role in promoting the development of intelligent field. The key technology in autonomous navigation technology is to realize robot autonomous path planning [1]. How to realize autonomous path planning is the key problem. Path planning technology refers to the robot planning an optimal collision free path from the starting to the end in the environment with obstacles.

Autonomous navigation is the symbol of intelligent mobile robot. Typical robot navigation technology includes magnetic navigation technology of logistics robot, simple landmark navigation technology, advanced visual navigation technology and new navigation technology. Magnetic navigation technology is generally the main traction device of automatic guided vehicle (AGV)[2]. This technology applies electromagnetic induction technology to robot tracking and navigation. The disadvantage is that with the accumulation of time, the cumulative error generated in the navigation process will increase infinitely. This method is not suitable for long-term accurate navigation and positioning; Landmark navigation is to set a special landmark at a special position in the path, obtain the landmark information through the specific sensor inside the robot, and calculate the current coordinate position of the mobile robot; Visual navigation realizes navigation by detecting road signs and obstacles through cameras.

At present, some progress has been made in the research of path planning algorithm, but its effect is still unsatisfactory. Global path planning algorithm and local path planning algorithm, such as Dijkstra algorithm, artificial potential field method, and SLAM technology[3]. Due to the problems of low computational efficiency, path deviation and falling into a dead cycle, there is a huge step for improvement, especially the need to realize the autonomous navigation of the robot in a complex environment.

To solve the above problems, this paper proposes a path planning method based on deep...
reinforcement learning [4], which realizes the autonomous navigation of robot by combining deep learning network and reinforcement learning. At present, the research fields based on deep reinforcement learning mainly include intelligent control of manipulator, robot navigation, robot path planning, and autonomous motion control of handling robot, cooperative control of multiple robots and so on.

2. The proposed approach
Deep reinforcement learning, combining the strong expression ability of deep learning with the full autonomous learning ability of reinforcement learning, can automatically and autonomously learn the effective features from the original data, and directly complete the mapping from the original data to the action output. It mainly consists of three parts: Environment (E), Agent (A) and Sum unit (S). The procedure of path selection algorithm based on deep reinforcement learning is presented as follows:

(1) Collecting the data, \( n \) is the total number of counting periods;

(2) In each period, the agent obtains the data from the environment. By executing the agent internal strategy to obtain the corresponding action who will be input to the environment and sum unit. Meanwhile, the environment returns a reward signal for each decision of the agent;

(3) In each period, the sum unit accumulates according to the action.

(4) After all the data processing is completed; sum unit returns the results to the environment. The environment evaluates the returned results and transmits the final evaluation signal to the agent.

In deep reinforcement learning, the parameters in the strategy network are trained by maximizing the expectation of cumulative reward. Convolutional neural network (CNN) is used in deep learning to complete the end-to-end data processing.

This paper uses the algorithm based on strategy gradient to optimize. In general, the goal of optimization is to maximize the cumulative reward. According to the definition of value function, the value function represents the expected reward that can be obtained by executing the corresponding strategy in the current state, while maximizing the cumulative reward, that is, when all possible states are the initial states, it can be replaced by the value function of the initial state, by deriving \( v_{\pi_0}(s) \),

\[
\nabla v_{\pi_0}(s) = \nabla \sum_{a \in A} [v_{\pi}(a|s)q_{\pi}(s, a)] = \sum_{s \in S} \sum_{k=0}^{\infty} P(s \rightarrow x, k, \pi) \sum_{a} [v_{\pi}(a|x)q_{\pi}(x, a)]
\]

(1)

In the above equation, \( S^d(s) = \sum_{k=0}^{\infty} P(s \rightarrow k, \pi) \). It can be seen from the above formula that the gradient of the objective function does not depend on the state distribution. What we need to do is to make the expectation of the sampling result approximate or satisfy the expression through some sampling method. The reinforce algorithm can be obtained by replacing \( q_{\pi}(s, a) \) in the equation (1) with the sample return estimate,

\[
\nabla J(\theta) = \mathbb{E} \left[ \sum_{t=1}^{M} \sum_{i=1}^{T} \ln \left( \frac{\pi(a_t|s_t, \theta)}{\pi(a_t|s_t, \theta)} \right) G_t \right]
\]

(2)

Where \( q_{\pi}(s_t, a_t) = \mathbb{E}_{\pi} \left[ G_t | s_t, a_t \right], G_t = \sum_{k=t}^{T} r_{k+1} \). M is the number rounds. The gradient method can be used to update the parameters. In this paper, we use the optimization method of deep learning to solve the parameters, and use equation (2) to construct the loss function as follows;

\[
L(\theta) = -\frac{1}{M} \sum_{i=1}^{M} \sum_{t=1}^{T} \ln \left( \frac{\pi(a_t|s_t)}{\pi(a_t|s_t)} \right) G_t
\]

(3)

Here, we can use the trick \( \nabla v_{\pi}(a|s) = \nabla \ln \left( \frac{\pi(a|s)}{\pi(a|s)} \right) \) to reconstruct the loss function. With the help of the back-propagation algorithm of deep learning, the gradient descent method is used to update the parameters.
Algorithm 1 The deep reinforcement learning algorithm

1. for round \( i = 1 \ldots M \), do
2. \hspace{1em} Get initialization status \( s_0 \)
3. \hspace{1em} for \( t = 1 \ldots T \), do
4. \hspace{2em} Get action \( a_t \) according to \( \pi(a_t|s_t) \)
5. \hspace{2em} Input \( a_t \) into the sum unit to accumulate
6. \hspace{1em} end for
7. \hspace{1em} Get the step sum of summation unit \( y' \)
8. \hspace{1em} Input \( y \) and \( y' \) into reward function to get the current reward
9. \hspace{1em} for \( t = 1 \ldots T \), do
10. \hspace{2em} \( G_t = \sum_{k=t}^{T} r_k \)
11. \hspace{2em} Calculate the gradient of the parameter according to formula (3)
12. \hspace{2em} Updating network parameters by gradient descent method: \( \theta \leftarrow \theta - \alpha \cdot \nabla L(\theta) \)
13. \hspace{1em} end for
14. end for

3. Experimental analysis
The experiment of deep reinforcement learning task is based on ROS framework. The simulation robot is Turtle BOT 3, and the proposed approach is implemented in PYTHON.

As can be seen from Figure 1, with the increase of the number of training rounds, the reward obtained by the robot in environmental exploration is also gradually increasing. Red represents the reward based on the deep reinforcement learning method, and blue represents the reward obtained by the traditional ddpg algorithm. Through the reward value curve, it can be found that the algorithm proposed in this paper is better, the growth of reward value is relatively peaceful and the training efficiency is higher. In addition, it can be seen from the average score curve that the performance of the red curve is better and more stable than the blue curve, indicating that reinforcement learning has advantages on the basis of the general depth framework.
4. Conclusions
In this paper, an intelligent robot path selection algorithm based on deep reinforcement learning is proposed and verified in the simulation environment. The experimental results show that the algorithm based on deep reinforcement learning can significantly improve the learning efficiency of the system, and it can be seen that the growth trend of reward is more stable.

References
[1] Wang C P, Liu P, Qin H Z. Overview of intelligent path planning algorithms for mobile robots [J]. Sensors and Microsystems, 2018, 8: 5-8.
[2] Huo L, Zhang X Y, Zhang C X. Research on a road sign for visual navigation and its fast recognition method [J]. Modern electronic technology, 2017, 11: 140-144.
[3] Mirowski P, Pascanu R, Viola F, et al. Learning to navigate in complex environments[J]. arXiv preprint arXiv:1611.03673, 2016:1-16.
[4] Arulkumaran K, Deisenroth M P, Brundage M, et al. Deep reinforcement learning: A brief survey [J]. IEEE Signal Processing Magazine, 2017, Vol.34, No.6, pp.26-38.