Autonomous Learning and Navigation of Mobile Robots Based on Deep Reinforcement Learning

Zhiqiang Lai\textsuperscript{1*}, Zhiwei Jia\textsuperscript{1}, Man Chen\textsuperscript{1}

\textsuperscript{1} School of Electrical & Information Engineering, Changsha University of Science & Technology, Changsha, Hunan 410114, China

\textsuperscript{*}email: laizq@stu.csust.edu.cn

Abstract. Aiming at the problems of convergence difficulties faced by deep reinforcement learning algorithms in dynamic pedestrian environments, and insufficient reward and feedback mechanisms, a data-driven and model-driven navigation algorithm which named GRRL has been proposed. In order to enrich and perfect the reward feedback mechanism, we designed a dynamic reward function. The reward function fully considers the relationship between the robot and the pedestrian and the target position. It mainly includes three parts. The experimental results show that the autonomous learning efficiency and the average navigation success rate of the mobile robot driven by the GRRL algorithm are improved, the average navigation time is shorter. The dynamic reward function we designed has a certain improvement effect on robot navigation.

1. Introduction

Nowadays, mobile robots have gradually been applied in the field of public service. Aiming at the navigation task in the mobile pedestrian environment, this paper proposes a data-driven and model-driven navigation algorithm GRRL, which can improve the efficiency of autonomous learning and navigation success rate, and shorten the average navigation time.

The mainstream mobile robot navigation methods can be divided into data-driven and model-driven according to the driving type. Among them, the model-driven method mainly relies on the traditional model (such as artificial potential field method [1]) to model the relationship between the robot and the surrounding environment, and relies on the path planning algorithm to realize the movement of the robot toward the target position. In recent years, deep learning and reinforcement learning technologies have gradually been applied to robot navigation tasks and have shown certain advantages. The data-driven method enables the robot to learn in the interaction of the surrounding environment, and finally obtain better navigation capabilities.

Robot navigation technology based on deep reinforcement learning [2] can combine the advantages of deep learning and reinforcement learning. Yu Fan Chen et al. [3] used RGB-D images as input information to directly output the robot action decision results, realizing end-to-end control from visual sensors to actions. Man Chen et al. [4] proposed a robot navigation technology based on deep reinforcement learning and artificial potential field method, which combines the attention mechanism and artificial potential field method to achieve rapid convergence and obtain good navigation effects.
2. GRRL

2.1. State
In order to facilitate the analysis of the effect of the algorithm, both robots and pedestrians are abstracted in the process of setting up the state space. The specific status representation is similar to the literature [4]. \( p = [p_x, p_y] \) and \( v = [v_x, v_y] \) are used to represent the position and speed of the robot. \( p' = [p'_x, p'_y] \), \( v' = [v'_x, v'_y] \) are used to represent the position and speed of the \( i \)-th pedestrian. \( r \) is the radius of robot. \( v_{pre} \) is the initial velocity of the robot. \( d_r \) is the distance between the robot and the target position. \( d_i \) is the distance between the robot and the \( i \)-th pedestrian. We assume that the total number of pedestrians is \( n \). According to the above description, the state can be described as two parts: the state \( \text{Rob}_i \) of the robot itself and the state \( \text{Hum}_i \) of the pedestrian. We integrate the above two parts of the state to get the final state representation.

\[
s_i = \left[ \text{Rob}_1, \text{Hum}_1^1, \text{Hum}_1^2, ..., \text{Hum}_1^n \right] \tag{1}
\]

2.2. State value function
Deep reinforcement learning uses deep networks as function approximators. The GRRL algorithm uses the social attention pool module [5] as a state value function. The social attention pool module is mainly composed of three parts, namely the interaction module, the pooling module and the planning module. The interaction module mainly encodes the information between the robot's moving pedestrians, and promotes the state interaction between the robot and the surrounding environment. The pooling module introduces a self-attention mechanism to aggregate interactive information. The planning module is used to calculate the final status information. The main structure of the state value network is shown in Figure 1.

![State value network](Fig. 1)

2.3. Reward function
The above-mentioned state value network fully considers the position of the pedestrian of the robot and realizes the interaction. The original reward function design is too simple and does not fully consider the position information of the robot's pedestrians. In order to enrich and perfect the reward feedback mechanism, we designed a dynamic reward function. The reward function fully considers the relationship between the robot and the pedestrian and the target position.

2.3.1. Target reward function. The ultimate goal of robot navigation is to reach the target position, so the actual distance between the target position and the robot should be inversely proportional to the reward value. That is, the shorter the distance, the greater the reward.

\[
\begin{align*}
   r_i(d_i) &= \alpha \frac{1}{d_i} \\
\end{align*}
\]

\tag{2}
where $\alpha$ is the target reward function factor, and its purpose is to keep the reward function at a reasonable value. and $d_i$ is the actual distance from the mobile robot to the target position, which is updated as the position changes during the navigation process.

2.3.2. obstacle avoidance reward function In addition to the target reward function, this article also sets the obstacle avoidance reward function. The reward function can make the robot avoid moving pedestrians during navigation. In the design process, the pedestrian closest to the position of the mobile robot is mainly considered. According to the idea of human obstacle avoidance, the obstacle avoidance reward function between robots and pedestrians should be proportional to the distance between them. That is, the closer the distance between the pedestrian and the robot, the smaller the value of the obstacle avoidance reward function. The reward function is specifically set as:

$$ r_2(d_2) = \beta d_2 $$

where $d_2$ is the distance between the mobile robot and the pedestrian closest to the robot. $\beta$ is the obstacle avoidance reward function factor, which aims to keep the robot's obstacle avoidance reward function at a reasonable value.

2.3.3. Direction reward function In addition to the objective reward function and obstacle avoidance reward function, we designed the direction reward function to enable the robot to realize the reward feedback in the direction of the navigation process. The design of the reward function is similar to the literature [6], which mainly relies on the attraction and repulsion between the charges. First, we define that mobile robots, pedestrians, and target locations are all charges in the navigation space. Among them, robots and pedestrians have the same kind of electric charge, and the two repel each other. There are different kinds of electric charges between the robot and the target, and the two attract each other. Therefore, the robot is subject to the attraction of the target location and the repulsion of pedestrians in the navigation environment. Determine the direction of movement of the robot based on attractive and repulsive forces $\gamma_1$.

On the basis of the above angle, the direction reward function is determined by combining the actual angle of the robot:

$$ r_3(\gamma_3) = \chi |\gamma_1 - \gamma_2| $$

where $\chi$ is the directional reward function factor, which aims to keep the directional reward function at a reasonable value. $\gamma_2$ is the actual direction of the current robot. The incentive function obtains the specific reward function value by calculating the angle difference between the actual direction of the mobile robot and the ideal direction $\gamma_1$. During the navigation of the mobile robot, the excitation function changes dynamically with each action decision.

2.3.4. Dynamic reward function The final dynamic reward function consists of the above three parts, which are target reward, obstacle avoidance reward and direction reward. Among them, the target reward function is determined according to the target position, and changes with the change of the robot position during the robot navigation. The other two parts of the reward function also change during the navigation of the mobile robot. Our dynamic reward function can enrich the feedback
mechanism in robot navigation, so as to achieve a better navigation effect. The final reward function is expressed as:

\[ r = r_1 + r_2 + r_3 \]  

(5)

3. Experiment

3.1. Experimental details
In order to verify the news of our GRRL algorithm, we use the python language to build the algorithm. Among them, deep reinforcement learning uses the pytorch architecture. We set up 5 pedestrians in a moving state, and designed target positions in the navigation environment. The specific experimental details are similar to the literature [5].

3.2. Experiment
During the experiment, we designed two sets of comparative experiments. These two sets of experiments used SARL and OCRA to drive the mobile robot. The experimental results are shown in Table 1. In addition, we also designed an ablation experiment to verify the role of each part in the dynamic reward function. The specific experimental results are shown in Table 2. We also designed three sets of experimental evaluation indicators to verify the navigation effect. They are navigation success rate, failure rate and navigation time.

| Method   | Success rate | Failure rate | Time |
|----------|--------------|--------------|------|
| GRRL     | 1.00         | 0.00         | 10.52|
| SARL[5]  | 0.99         | 0.01         | 10.58|
| ORCA[6]  | 0.99         | 0.00         | 12.29|

| Reward   | Success rate | Failure rate | Time |
|----------|--------------|--------------|------|
| \( r_1 \) | 0.99         | 0.01         | 10.57|
| \( r_1 + r_2 \) | 0.99      | 0.01         | 10.53|
| \( r_1 + r_2 + r_3 \) | 1.00      | 0.00         | 10.52|

It can be seen from Table 1 that compared to ORCA and SARL algorithms, our algorithm has improved the navigation success rate. The navigation time of the GRRL algorithm is also shorter. Therefore, we have reason to believe that the design of the dynamic loss function makes the navigation effect better. The specific navigation effect is not obvious in the navigation success rate, but there is a certain improvement in the navigation time. Using our algorithm-driven robot can reach the target position more quickly. In addition, ORCA is a non-learning algorithm, and its navigation time is longer than that of reinforcement learning algorithms.

It can be seen from the ablation experiment in Table 2 that the three-part reward function we designed is useful for navigation effects. The direction reward function can improve the success rate of navigation, although this improvement is small. The three-part reward function improves the navigation time to a certain extent. Therefore, the dynamic reward function we designed has a certain improvement effect on robot navigation.
4. Conclusion
This paper proposes the GRRL algorithm for the robot navigation task of the mobile pedestrian scene in the public service field. The algorithm designs a dynamic reward function to enable mobile robots to obtain a rich and reasonable reward mechanism. The dynamic reward function is mainly divided into three parts: target reward, obstacle avoidance reward and direction reward. The results of comparative experiments show that the algorithm-driven mobile robot has a higher navigation success rate and a shorter navigation time.

References
[1] Xue F, Jin S J 2018 Research on Path Planning of Improved Artificial Potential Field for Mobile Robot Based on Laser Sensor Measurement & Control Technology
[2] Xiang J , Li Q , Dong X, et al. 2019 Continuous Control with Deep Reinforcement Learning for Mobile Robot Navigation Chinese Automation Congress (CAC) IEEE
[3] Chen Y F, Liu M, Everett M, et al. 2017 Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning IEEE International Conference on Robotics and Automation (ICRA) IEEE Singapore pp 285-292
[4] Chen M, Li M J, Li Y W, Lai Z Q 2021 Mobile robot navigation based on deep reinforcement learning and artificial potential field method Journal of Yunnan University: Natural Sciences Edition 43(6) pp 1125-1133 doi: 10.7540/j.ynu.20210091
[5] Chen C, Liu Y, Kreiss S, et al. 2019 Crowd-Robot Interaction: Crowd-Aware Robot Navigation With Attention-Based Deep Reinforcement Learning 2019 International Conference on Robotics and Automation (ICRA)
[6] Berg J, Guy S J, Ming L, et al. 2011 Reciprocal n-Body Collision Avoidance Robotics Research