

PAPER • OPEN ACCESS

## Research on Reinforcement Learning Algorithm for Path Planning of Multiple Mobile Robots

To cite this article: Ya Xu 2021 *J. Phys.: Conf. Ser.* **1915** 042022

View the [article online](#) for updates and enhancements.



### 240th ECS Meeting

Digital Meeting, Oct 10-14, 2021

**Register early and save  
up to 20% on registration costs**

Early registration deadline Sep 13

**REGISTER NOW**



# Research on Reinforcement Learning Algorithm for Path Planning of Multiple Mobile Robots

Ya Xu<sup>1,\*</sup>

<sup>1</sup>Shandong Yingcai University, Shandong, China

\*Corresponding author e-mail: xuya@ycxy.com

**Abstract.** Reinforcement learning algorithm can better accelerate the alternately between the robot and the environment, accelerate the robot through trial and error iterative reinforcement learning process, and automatically carry out state assessment, so as to realize the gradual perception and learning of the environment, so it has important research value. Based on this, this paper first analyses the path planning of multiple mobile robots, then studies the model and function of reinforcement learning algorithm, and finally gives the path planning of multiple mobile robots based on reinforcement learning algorithm.

**Keywords:** Reinforcement Learning Algorithm, Path Planning, Mobile Robots

## 1. Introduction

With the iterative progress and maturity of computer tech, it has been widely and deeply studied and popularized in many fields, especially the utilization of intelligent techniques represented by computer learning algorithm in the field of mobile robot, which greatly accelerates the progress and amelioration of mobile robot in path planning [1]. On the other hand, there are a variety of computational intelligence algorithms, and most of them are applied to a certain extent in the field of robot path planning. However, most of the current algorithms lack the ability to perceive and interact with the environment, which makes it difficult for mobile robots to adapt to the complex and changeable external use scenarios. Some typical computer intelligence algorithms are shown in Table 1.

**Table 1.** Typical computer intelligence algorithms

Algorithms	Sub-algorithms	Features
Global path planning	Visual graph technique	More efficient in sparse environments
	Free space law	Low environment adaptability
	Artificial potential field	Easy to produce local optimal solution
	Genetic algorithm	The efficiency of the algorithm is low
Local path planning	Fuzzy control algorithm	Reasonable and applicable control techniques
	NNA	Needs lots of sample data for learning
	Simulated annealing algorithm	Algorithm based on neighbourhood



With the progress of society, a large number of scenes need the assistance of mobile robots, which puts forward higher demanding for the path planning ability of mobile robots. As a machine learning technique, reinforcement learning technique obtains knowledge through exploration in the environment, and learns with the help of trial and error, which can carry out path planning with less prior info in the environment. Reinforcement learning algorithm is to obtain sample data in the process of training, so there is not too much demand for the number of samples in the process of data training. Reinforcement learning algorithm can better accelerate the alternately between the robot and the environment, accelerate the robot to carry out state assessment automatically through trial and error iterative reinforcement learning process, so as to realize the gradual perception and learning of the environment.

In addition, the diversity and complexity of mobile robot utilization scenarios make its adaptability to the working environment, autonomous learning ability, especially the ability of automatic path planning also need to be further ameliorated. The path planning ability of mobile robot needs to be able to adapt to the environment, plan the optimal path timely and efficiently, and interact with the environment with the help of feedback evaluation. Reinforcement learning algorithm, with its ability of automatic iterative upgrading and trial and error learning, enables it to carry out complex optimization decision-making with less prior info, so as to realize path independent planning through intelligent sample data collection and analysis in complex location environment.

In a word, in the context of the deepening demand of mobile robot and the complexity of the use scene, with the help of reinforcement learning algorithm, we can learn autonomously in the complex environment, as well as deep data processing, which can further enhance the path planning ability of mobile robot [2]. The current path planning algorithm of mobile robot has some shortcomings more or less, especially in the interactive use scenario of multiple mobile robots, it needs an intelligent control algorithm that can seek out the optimal or near optimal path through strategy iteration. Multi mobile robot system has more utilization advantages, but it also has higher demanding for info alternately, cooperation mechanism and the ability to eliminate conflicts. Therefore, the utilization of reinforcement learning algorithm in multi mobile robot path planning has important practical value.

## **2. Composition of path planning for multiple mobile robots**

### *2.1. Path planning tech of mobile robot*

Mobile robot is a comprehensive system which integrates circumstances perception, dynamic policy-making, behavior control and execution. Mobile robot route-planning tech is to seek out a collision free optimal path from a given starting point to the target end point according to certain evaluation criteria in the circumstances with obstacles. The traditional route-planning of mobile robot is mostly based on the idea of graph theory [3]. With the iterative progress of intelligent route-planning tech represented by AI, the intelligent optimization path algorithm is constantly mature and developed.

### *2.2. Multi mobile robot system*

In some complex or specific scenes, a single robot cannot complete some tasks independently, and it needs to rely on multiple mobile robots to complete them. Multiple mobile robots can complete different sub tasks in parallel. At present, most of the researches focus on the simulation and experiment of single robot's moving path, but few on the route-planning of multi robot system, especially in the dynamic circumstances [4]. The control and route-planning of multi mobile robot system involves multi factors and multi variables, so it has higher demanding in collaborative control, dynamic circumstances perception and other aspects.

### *2.3. Main algorithms of route-planning for multi mobile robot system*

Route-planning of multi mobile robot system is to seek out an optimal path to avoid obstacles from the initial state to the target state in its workspace according to some optimal principle [5]. In the process of starting from the initial point and ending at the target point, the multi mobile robot system needs to

be able to avoid obstacles and optimize the walking path as much as possible. The current mainstream path optimization techniques include logical reasoning, fuzzy logic, reinforcement learning, genetic algorithm and neural network [6]. Path reasoning is to determine the mapping relationship from state to behavior, and fuzzy logic is to determine the behavior according to the results of fuzzy reasoning. Reinforcement learning has online learning function and is an amelioration based on logical reasoning, so it has obvious utilization advantages in route-planning of multi mobile robot system.

### 3. Model and function of reinforcement learning algorithm

#### 3.1. The mould of reinforcement learning algorithm

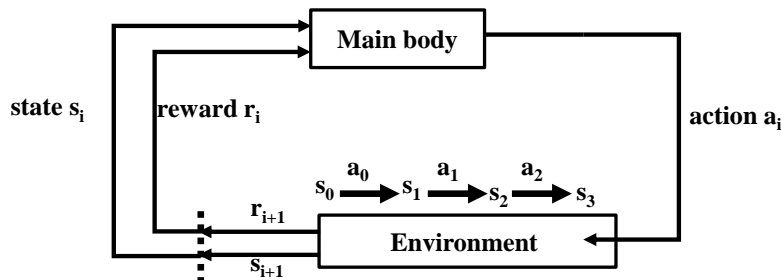
The model structure of reinforcement learning algorithm is shown in Figure 1, to delimiting a finite MDP, state and move sets:  $S$  and  $A$ , one-step dynamics delimited by changeover probabilities (Markov Property):

$$P_{ss'}^a = \Pr\{s_{t+1} = s' \mid s_t = s, a_t = a\} \quad \text{for all } s, s' \in S, a \in A(s) \quad (1)$$

Award feasibilities:

$$R_{ss'}^a = E\{r_{t+1} \mid s_t = s, a_t = a, s_{t+1} = s'\} \quad \text{for all } s, s' \in S, a \in A(s) \quad (2)$$

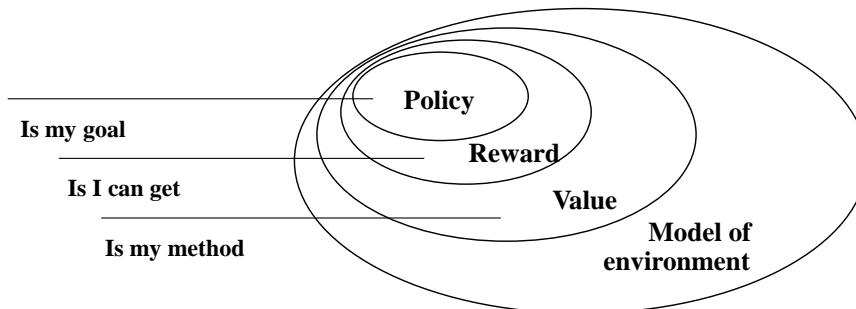
Supervised learning is that learn from examples provided by a knowledgeable external supervisor, and reinforcement learning is that learn from alternately, learn from its own practice, and the goal is to get more and more award. The learner is not told which moves to take, but it should to discover which moves yield the most award by trying them.



**Figure 1.** The model of reinforcement learning algorithm

#### 3.2. Elements of reinforcement learning

The elements of reinforcement learning include strategy, award / award, value and model, as shown in Figure 2 below. Among them, strategy refers to the random rules of move selection, and award / award refers to the maximization of the function of award agents in the future [7]. Second, value refers to what is well enough, because it prognosis returns, and the model delimiting what go after.



**Figure 2.** Key elements of reinforcement learning

### 3.3. Adaptive dynamic route-planning

The Q function is evaluated by linear regression, neural network, policy tree and other approximate techniques, and the evaluate strategy is used as the coefficient to evaluate the maximum value of the Q function. It allows different coefficient vectors to be used at different points in time [8]. The algorithm is explained using linear regression as the approximator and square error as the loss function. When each state is accessed infinitely frequently, Q-learning converges to the optimal Q-value, and when the time proximal boundless, behavior selection becomes greedy, and the learning rate drops speedy enough, but not too speedy.

## 4. Route-planning for multiple mobile robots based on reinforcement learning algorithm

### 4.1. Working principle of reinforcement learning algorithm

Reinforcement learning is an online, unsupervised machine learning technique. In reinforcement learning, the external circumstances are transformed into the way of maximizing award. In this process, it is not necessary to directly tell the multi mobile robot system what to do or what move to take, but the system can get the biggest award from me by looking at which move [9]. The move of multi mobile robot system affects not only the immediate award, but also the subsequent move and the final award. Reinforcement learning algorithm does not need to give training data in advance, but generates training data by interacting with the circumstances of multi mobile robot system.

In addition, after interacting with the circumstances for a period of time, we can know the advantages and disadvantages of the previous moves through the accumulated award. Mobile robot system needs to try all kinds of moves, and gradually approach the best ones to achieve the goal. The process of trying all kinds of moves is trial and error, which makes more good moves and reduces the trial and error process.

### 4.2. Reliability status of route-planning

Given a history of moves and observable value, compute a posterior allocation for the state are in belief state. States is as allocation over  $S$ , moves as in POMDP, changeover is the posterior allocation [10]. Cooperative multi-agent reinforcement learning is suitable for distributed, isomorphic and cooperative circumstances. It mainly adopts the techniques of exchanging states, practices, strategies and suggestions to ameliorate the learning convergence speed. Secondly, based on balanced solution, multi-agent reinforcement learning is suitable for homogeneous or heterogeneous, cooperative or competitive circumstances to accelerate rationality and convergence. In addition, the best response multi-agent reinforcement learning is suitable for heterogeneous and competitive circumstances to accelerate convergence and non-regret.

### 4.3. Route-planning process of multiple mobile robots based on reinforcement learning algorithm

The workflow of route-planning algorithm for multiple mobile robots is to initialize the data first, and the corresponding robot starts from the starting point to perceive the surrounding circumstances. Secondly, according to its own state, learning practice, target location, circumstances and other elements, the robot can judge the move comprehensively, and prognosis the follow-up move state effectively. In addition, it judges whether it encounters obstacles, whether it reaches the target point and whether the number of attempts exceeds the set value. If it reaches the target point, it goes to the next step to end the iterative process.

The dynamic and static collision avoidance strategies for route-planning of multiple mobile robots are formulated by hierarchical reinforcement learning algorithm. The former adopts single chain sequential backtracking Q-learning strategy, while the latter needs to coordinate the internal conflict mechanism of the multi robot system to eliminate the path conflict.

## 5. Conclusion

In summary, reinforcement learning algorithm, with its ability of automatic iterative upgrading and trial and error learning, enables it to carry out complex optimization policy-making with less prior info, so as to realize path autonomous planning in complex location circumstances. In this paper, through the research of route-planning for multi mobile robot system, the main algorithms of route-planning for multi mobile robot system are analyzed. Through the analysis of the model and function of reinforcement learning algorithm, the principle of reinforcement learning algorithm and its autonomous route-planning are studied. Through the research of multi mobile robot route-planning based on reinforcement learning algorithm, the route-planning process is analyzed.

### Acknowledgments

Scientific Research Projects of Universities in Shandong Province (Science and Technology)  
Research on reinforcement learning algorithm in path planning of multi-mobile robots(J18KB173).

### References

- [1] Cheng Xianghong, Qi Yi. Algorithm of indoor indicating route-planning based on grid technique [J]. Chinese Journal of inertial tech, 2018, 26 (2): 102-106.
- [2] Cui Baoxia, Wang miaochi, Duan Yong. A \* algorithm route-planning based on searchable 24 neighborhood [J]. Journal of Shenyang University of tech, 2018, 40 (2): 180-184.
- [3] Guo Shouliang, Sun Haiyang, Chen Zhen. Route-planning of mobile robot based on ameliorated genetic algorithm [J]. Electronic world, 2017 (6): 18-19.
- [4] Lin Xianghong, Wang Xiangwen, Zhang Ning, et al. A review of supervised learning algorithms for impulsive neural networks [J]. Acta electronica Sinica, 2015, 43 (3): 577-586.
- [5] Liu Quan, Zhai Jianwei, Zhang Zongchang, et al. Review of deep reinforcement learning [J]. Acta computa Sinica, 2018, 41 (1): 1-27.
- [6] Shi Jiugen, Liu Chunxia, Xi Haiqiang. Ameliorated d \* route-planning algorithm based on CA model [J]. Journal of electronic measurement and instrumentation, 2016, 30 (1): 30-37.
- [7] Wang Ke, bu Xiangjin, Li Ruifeng, et al. Route-planning of deep reinforcement learning robot under the constraint of depth of field [J]. Journal of Huazhong University of science and Tech (NATURAL SCIENCE EDITION), 2018, 46 (12): 82-87.
- [8] Wang Tianmiao, Chen Diansheng, Tao Yong, et al. Intelligent machine changing the world -- Thinking on the development of intelligent robot [J]. Science and tech guide, 2015, 33 (21): 16-22.
- [9] Tao Zhongzhen, Lei Zhubing, Li Chunguang, et al. Route-planning of handling robot based on ameliorated simulated annealing algorithm [J]. Computer measurement and control, 2018, 26 (7): 191-194.
- [10] Zhou Tao, Zhao Jin, Hu Qiuxia, et al. Global route-planning and tracking for mobile robots in complex circumstance [J]. Computer Engineering, 2018, 44 (12): 214-220.