ATLANTIS PRESS

Advances in Intelligent Systems Research, volume 147

International Conference on Network, Communication, Computer Engineering (NCCE 2018)

Planning of Opposite Q Learning Based on Virtual Sub-Target in Unknown Environment

Shengmin Wang [a)], Wei Lin

[1] *Faculty of Computer, Guangdong University of Technology, Guangzhou,510006 China.*

[a)] Corresponding author: 1546003584@qq.com

Abstract. Aiming at the problem of Q value update is slow and easy to produce dimension disaster for Q learning algorithm in complex unknown environment, A path planning algorithm for opposite Q learning robot based on virtual sub targets in unknown environment is proposed .The algorithm is according to the state trajectory of the mobile robot, two state linker are established to record the current sate-action pair and current state-reverse action pairs, from the value of the tail of a single chain, the current state, is traced back to the Q value at the end of a single linker head until the target is reached. Meanwhile, searching for optimal virtual sub target in local detection domain to solve the problem of Q-learning prone to dimension in a large-scale environment. The experiments show that the algorithm can effectively speed up the convergence of learning algorithm and improve the learning efficiency in complex unknown environment and achieve the robot navigation task with the better path.

Key words: Mobile robot; Virtual sub target; Opposite Q-learing; Unknown environment.

INTRODUCTION

With the rapid development of service and logistics industry and the upgrading problems of related industries. Autonomous mobile robots have become a hot spot in the research of security companies and logistics companies [1]. Path planning is one of the key technologies for mobile robots to achieve autonomous navigation. Path planning refers to finding an optimal path from the initial pose to the target pose without collision in an environment with obstacles according to certain evaluation criteria. Currently, most mobile robots perform a predetermined sequence of actions in a well-structured, known map, but in the event of an unstructured or new environment, the mobile robot does not have active learning and adaptive capabilities for the actual environment. In the process of constantly interacting with the environment, adaptive path planning can plan a collision-free path from the starting point to the target point and meet certain optimization criteria [2].

The path planning algorithm in the known static environment has been mature, mainly including A*, RRT, artificial potential field method, visual graph method, genetic algorithm [3] and so on. In the unknown environment, because the mobile robot lacks the knowledge of the environment and cannot identify all obstacles information and location information in the environment, it can only perceive the local information of the robot centered by local sensors. In order to make the robot in the premise of not prejudging obstacle specific information, through continuous interaction with the environment to find a feasible path, some scholars use reinforcement learning approach to path planning and the algorithm through comparison and analysis of Q programming learning algorithm has achieved some results. At the same time, Q learning has a great increase of relative to A* in energy saving. Q learning is an online and unsupervised machine learning algorithm, which has become a research hotspot in the path planning of mobile robots in unknown environment.

HJ Hwang [4] proposes a $Q(\lambda)$ learning algorithm, which accelerates the convergence speed. However, in the process of generating single chain, all the loops in the original state path will be removed, which results in a feasible solution, but it may not be the optimal solution. S Wen [5] applied the EKF-SLAM and A* path planning to a

humanoid robot and tested it in the indoor environment. Mathijs Pieters[6] proposed a learning algorithm based on Q playback experience, through the experience of the function E (s, a) to make up for the previous lack of environment model of cognitive defects, to a certain extent, accelerate the convergence speed of the algorithm, but the algorithm in addition to update the Q value function and update E (s a), increase the time complexity of the algorithm. In addition, the problem of "dimension disaster" of Q value is ignored when the scale of environment is large.

In view of the above research status and shortcomings, this paper proposes the path planning of the antithesis Q learning robot based on the virtual sub target. By establishing bidirectional state chain, the action decision of current state can affect the preceding state action pair quickly, so as to improve the lag of traditional Q learning data transfer and increase the convergence speed. At the same time, the method of finding the optimal virtual sub target in the local detection domain can solve the problem of dimension disaster which is easily produced by Q learning in large-scale environment.

AN OPPOSITE Q LEARNING ALGORITHM BASED ON VIRTUAL SUB-TARGET

Laser Radar and Laser Data Coordinate Conversion

In this paper, the sick laser radar, as shown in Figure 1, has a range of 180 degrees after the interception. With the characteristics of fast speed, high accuracy, strong anti-interference ability and wide range, the robot can acquire the surrounding environmental information quickly and effectively. From the data of the robot sensing information from laser radar rotating 180 degrees obtained by measuring the distance, rotation angle and laser end point, calculate the distance from obstacles around the robot, in the two-dimensional coordinate system, the coordinates of the end point of the laser expressed in the form of $s(s_x, s_y)$, if the laser radar in the global coordinate system for the position $\phi = (\phi_x, \phi_y, \phi_\theta)$, through the ϕ transformation matrix of the T_ϕ, the s conversion to the global coordinate system, As shown in Formula (1).

FIGURE 1. Laser radar and mobile platform

$$T_\phi s = \begin{pmatrix} \cos\phi_\theta & -\sin\phi_\theta \\ \sin\phi_\theta & \cos\phi_\theta \end{pmatrix} s + \begin{pmatrix} \phi_x \\ \phi_y \end{pmatrix} \tag{1}$$

Laser Radar and Laser Data Coordinate Conversion

According to the detection range of sensors for mobile robots, robot center made a local environment as a window, the window in the environment of grid, each grid coordinates calculation according to the current position of the mobile robot, according to the position of the target as a guide, if through the obstacle information and select a virtual the target as a local window target location. As shown in Figure 3, the dashed line circle represents the maximum

scanning range of the laser radar, and the points that converge to the barrier expansion boundary in the R1 scanning region are P1, P2, respectively. These two points will serve as a candidate for the virtual subtarget. Because of P1, the distance from P2 to the target is $d_{L1\to P1}+d_{P1\to R_{goal}}>d_{L1\to P2}+d_{P2\to R_{goal}}$, so P2 is regarded as the best subtarget. In the same way, in position R2, the optimal virtual subtarget point of the robot is P3.

From the above analysis, we can get that (2) is the strategy of the optimal virtual sub-target in the current detection domain:

$$\min(d_{R\to pi}+d_{pi\to R_{goal}}), i\in n \tag{2}$$

Where P_i represents the virtual sub-target candidate node in the current area, R_1, R_2 represents the robot's position, and R_{goal} represents the position of the target point.

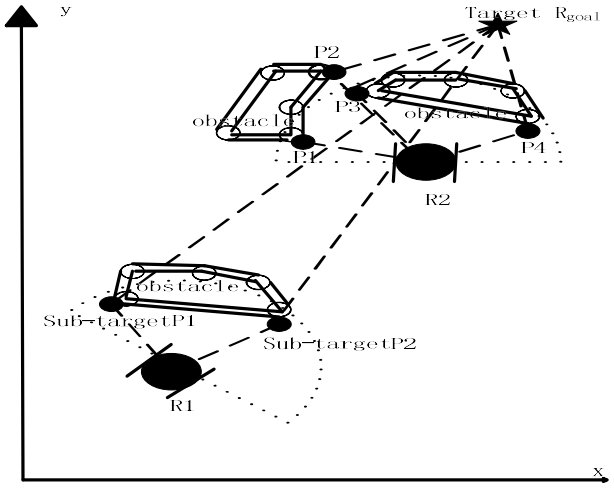


FIGURE 2. Selection of virtual sub-targets

The basic steps of a virtual sub-object generation algorithm are as follows:
Step 1: Obtain n virtual sub-target candidates based on the lidar's scan range and obstacle information.
Step 2: Combine the adjacent visible points with an interval smaller than a fixed value d into one obstacle group.
Step 3: Calculate the optimal virtual sub-target in the current detection area according to formula (2).

Opponent Q Learning Algorithm Steps Based on Virtual Sub-goal

In order to improve the speed of global planning, the density function of obstacles is given:

$$P_{obs}=n_{obs}/n_{total} \tag{3}$$

In which n_{obs} represents the number of obstacles in the window, and n_{total} represents the total number of raster in the window.

The distance L in a window is defined as:

$$L=\begin{cases}1\ \text{Grid length}, & p_{ob}>n2\\ 2\ \text{Grid length}, & n1<p_{ob}<n2\\ 1/3\ \text{Local path length}, & p_{ob}<n1\end{cases} \tag{4}$$

Where $0<n_1<n_2<1$. When there are few obstacles in the window, the robot will take 1/3 of the local path length and then re-plan the new path; when there are more obstacles, it will be planned once every 2 grids; when the obstacles

are particularly large, each time A grid is planned once, and this relatively reduces the number of local path plans. The value of the local path length needs a lot of experiments to determine roughly. The principle of the value is that the global path obtained after the local path is superposed is optimal or nearly optimal. The same choice of window size will also affect the global optimal result. If the value is too small, it is easy to fall into a local optimum. If the value is too large, the convergence speed of the algorithm will be reduced. Therefore, the value of L and the value of the window size all require a large number of experimental settings. The window size set in this paper is determined by setting the detection range of the sensor.

The flow chart is shown as shown in the diagram:

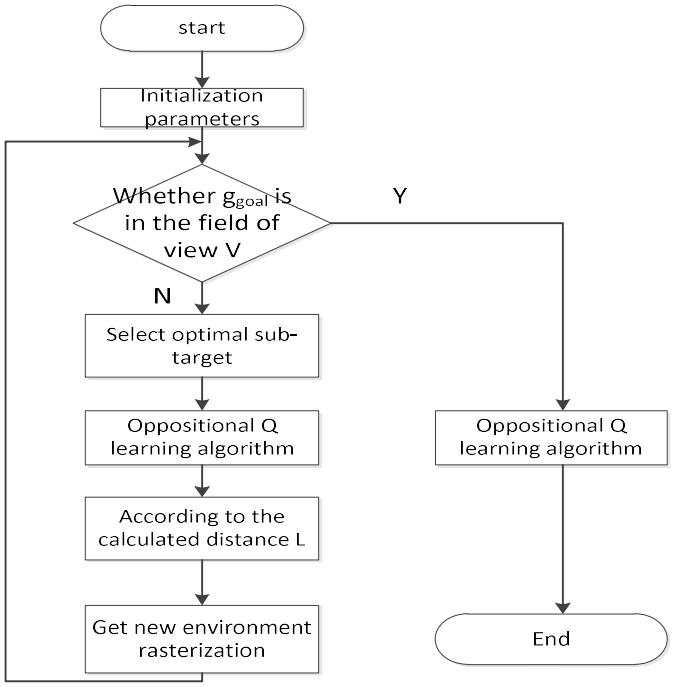


FIGURE 3. The flow charts

Step 1: Initialize the visual radius R, thresholds n_1 and n_2, start point g_{start} and end point g_{goal} of the lidar of the mobile robot.

Step 2: If g_{goal} is within the current field of view V of the mobile robot, the opposite Q learning algorithm is used to plan an optimized path from the current position of the robot to the position of the target point, and the algorithm ends.

Step 3: Select the optimal virtual sub-objective g_{sub_goal} according to formula (2). The opponent Q learning algorithm plans an optimized path from the current position of the robot to the target point and records the length of the path.

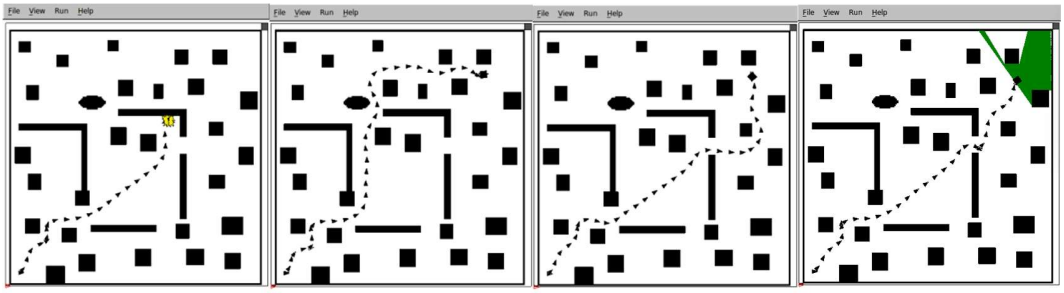
Step 4: Calculate G according to formula (3), calculate P_{obs} according to formula (4), and move L along the local planned path.

Step 5: Update the current environment V according to the current position and view, rasterize the current environment, and then return to the second step.

EXPERIMENT AND ANALYSIS

In order to verify the effect of Q learning algorithm based on virtual sub goal in large-scale environment, we built an unknown environment on the ROS robot simulation platform, including surrounding walls, boxes and baffles. A lot of experiments have been done in 50 * 50 unknown environment. The relevant parameters are set as follows: learning factor λ =0.4, discount factor γ =0.95, and Max iterated algebra count=120. The following four algorithms

are performed on the same condition, which guides the mobile robot to complete the final navigation according to the convergent Q, as shown in Figure 4.



(a)Traditional Q learning (b) $Q(\lambda)$ (c)Opposite Q learning (d)Algorithm in this paper

FIGURE 4. Robot navigation path in complex environment

In order to verify the real-time performance of the proposed algorithm, 20 groups of experiments were carried out in the random environment of the 32 obstacles, and the planning time was shown as shown in Figure 5. The average planning time is 0.152s, which can basically meet the real-time requirements of dynamic path planning for mobile robots.

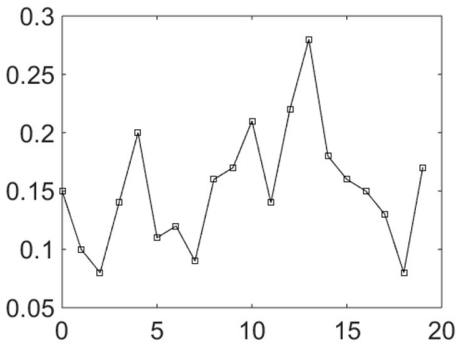


FIGURE 5. 20 Groups of planning time

In order to further validate the algorithm ability to adapt to the environment, the simulation experiment is done in the obstacles in a more complex environment, as shown in Figure 6, the mobile robot in a complex environment can also be planning a collision free path, indicate that this algorithm has strong adaptability to various environments.

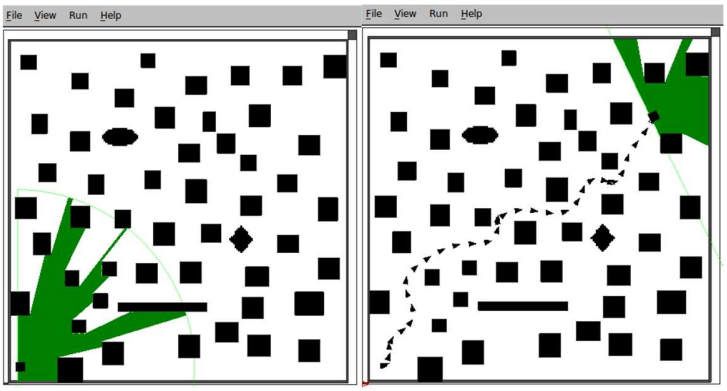


FIGURE 6. Complex environment

CONCLUSION

Although the traditional Q algorithm can find a collision free path to reach the target point in the unknown environment, but because of the lack of prior knowledge, the iteration speed is slow, especially in the environment of increasing scale, the state space will lead to sharp increase of dimension disaster. Aiming at the above problems, based on the traditional Q learning, according to the actual situation, put forward opposite Q virtual sub goal-based learning, the algorithm through the establishment of two monocatenarian to update the Q value and backtracking by setting the virtual sub target will complex environment into simple local environment. After these measures are taken, the real-time performance of the algorithm and the adaptability to the environment are greatly improved. The experimental results show that the algorithm is simple, fast and adaptable to the environment, especially when the environment is large, it can better reflect the superiority of the algorithm.

ACKNOWLEDGMENTS

At the time of this thesis, I would like to express my deep appreciation to all those who have taken care and help in learning and living during the Master's degree.

First of all, we would like to thank Associate Professor Lin. Can be successfully completed the writing of the paper, all embodies the teacher's effort and sweat. The teacher in the paper topics, research programs to determine and the specific implementation process have given careful guidance, their rigorous attitude and systematic research ideas I benefit for life.

REFERENCES

1. Li Lopez A, Paredes R, Quiroz D, et al. Robotman: A security robot for human-robot interaction[C]// International Conference on Advanced Robotics. IEEE, 2017:7-12.
2. Klidbary S H, Shouraki S B, Kourabbaslou S S. Path planning of modular robots on various terrains using Q-learning versus optimization algorithms[J]. Intelligent Service Robotics, 2017, 10(2):121-136.
3. Ogunniyi S. Energy efficient path planning: the effectiveness of Q-learning algorithm in saving energy[J]. 2014..
4. Hwang H J, Viet H H, Chung T C. Q(λ) Based Vector Direction for Path Planning Problem of Autonomous Mobile Robots[J]. Speech Communication, 2011, 17(3–4):249-262.
5. Wen S, Chen X, Ma C, et al. The Q-learning obstacle avoidance algorithm based on EKF-SLAM for NAO autonomous walking under unknown environments[J]. Robotics & Autonomous Systems, 2015, 72(C):29-36.
6. Pieters M, Wiering M A. Q-learning with experience replay in a dynamic environment[C]// Computational Intelligence. IEEE, 2017.