Partition Heuristic RRT Algorithm of Path Planning Based on Q-learning

Zhiyong Liu ,Fei Lan ,Haibo Yang
Omniroll (guangdong) intelligent manufacturing co., LTD of Liaoning Province
Shenyang, China
liuzhiyong0919@sina.com

***Abstract*—To solve the problem of high randomness in traditional rapid exploration random tree (RRT) path planning, a partition heuristic RRT planning algorithm based on Q-learning (Q-PRRT) is proposed. Partition heuristic rules are established by designing a sampling strategy of target bias and obstacle-avoided guidance. Markov modeling is carried out for partition heuristic RRT algorithm based on Q-learning, and actions are constructed based on partition heuristic rules. Each node is evaluated by designing the global optimal path reward function with Q-learning method, retain path nodes and eliminate redundant nodes based on greedy strategies. The simulation results show that Q-PRRT algorithm guarantees the optimality of the global planning path, obtains a smoother planning path, also improves path search efficiency and obstacle avoidance ability. It has better adaptability in different obstacle environments.**

Keywords—path planning; RRT; partition heuristic; Q-learning; Markov model

I. INTRODUCTION

Path planning is the core of mobile robot research field and the basis for its control [1]. Path planning algorithms can be classified into global path planning and local path planning. There are three main categories of common global path planning algorithms, the classic graph search methods [2], including visibility graph [3], Dijkstra algorithm [4], A-star algorithm [5], etc; probability roadmap method (PRM) [8] and rapidly exploring random tree (RRT), etc. based on sampling planning method [9]; as well as intelligent methods such as genetic algorithm [6] and neural network algorithm [7], etc.These algorithms have their own advantages in different fields, however, there are also insufficiencies in computational complexity, local optimal solution, map adaptability, dynamic environment effectiveness, global and multi-objective optimization ability in general.

It became the industrial solution standard because that the path planning method based on sampling can effectively solve the problem of path planning in high-dimensional space and complex constraints [10]. Among them, rapid exploration of random tree (RRT) can efficiently traverse unknown complex obstacle space and high-dimensional dynamic environment without the environment model. It makes RRT algorithm widely applied and well researched in the field of mobile robot path planning and control [11]. However, the traditional RRT algorithm has a problem of large randomness due to blind searching. Against the existing problems of RRT algorithm, researchers have proposed many improved methods. J.J. Kuffner and S.M. LaValle proposed RRT-connect algorithm [12] and Bidirectional-RRT algorithm [13], which adopted the guidance mode of dual-tree random parallel search to improve the algorithm's convergence speed. GoalBias-RRT algorithm [14] uses target bias heuristic to accelerate the convergence speed of the algorithm, but it is easy to drop into local extrema. Jordan M and Perez A proposed Bi-RRT*[15] algorithm, which used the improved greedy algorithm as heuristic function to guide the random tree to obtain better solutions. Intelligent heuristics are also integrated with sampling algorithms. Qureshi A H and Aya Z Y. Proposed the IB-RRT* algorithm [16], which uses the heuristic idea of intelligent sample insertion to improve the algorithm's convergence speed and obtain the optimal path on the basis of dual-tree. X Zhang et al. [17] introduced a self-learning strategy and a mixed deviation sampling method to improve the planning efficiency of algorithm. These methods still have some problems in the path planning of mobile robot, such as insufficient convergence speed, low obstacle avoidance efficiency, excessive traversal, and non-smooth trajectory, which are difficult to meet the practical application.

To solve these problems, this paper proposes a partition heuristic RRT algorithm based on Q-learning. Q-learning algorithm can cause the the curse of dimensionality in large-scale and high-dimensional state space, while RRT algorithm is suitable for solving the problem of motion planning of mobile robot in multi-dimensional space, multi-degree of freedom space and complex environment. The combination of them can complement each other. In this paper, a partition heuristic rules is established by designing a target bias and obstacle avoidance guide sampling strategy. Based on the Q-learning method, establish the action set under the partition heuristic rule. Evaluate each node through the reward function of the global optimal path. Then retain path nodes and eliminate redundant nodes based on greedy strategies. Guarantees the optimality of the global planning path and obtains a smoother planning path, also improves path search efficiency and obstacle avoidance ability.

II. PARTITION HEURISTIC RULES

Traditional RRT algorithm has no heuristic mechanism and high randomicity; GoalBias-RRT algorithm adopts a target bias heuristic mechanism, which has no obstacle-avoided guidance and is easy to drop into local

978-1-7281-1907-6/19/$31.00 ©2019 IEEE

extrema. Partition heuristic rules based on target bias and obstacle-avoided guidance sampling strategy are designed to improve the traditional RRT algorithm, which can improve the guidance and obstacle-avoided ability of the path planning algorithm.

A. Local Environment Modeling of Mobile Robot

In the global path planning, the coordinate system is established based on the line from the starting point to the target point as the y-axis, and the direction of the mobile robot towards the target point is regarded as the initial direction of the coordinate, denoated as φ=0º. Mobile robot local environment partition modeling is shown in Figure 1.

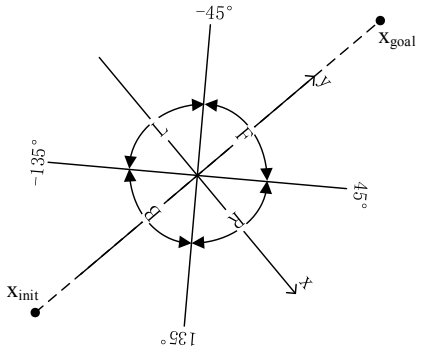

Fig. 1. Local environment model of mobile robot.

Thereinto, The Angle 0º is marked as the positive direction of the y-axis. Clockwise -45º~45º is F(Forward) region, indicates the forward region of mobile robot; 45º~135º is R(Right) region, indicates the the right side of the mobile robot; 135º~-135º is B(Backward) region, indicates the backward region of the mobile robot; -135º~-45º is L(Light) region, indicates the left area of the mobile robot.

B. Target Bias and Obstacle-avoided Guidance Sampling Strategies

The distance between obstacles and mobile robot can be obtained by measuring through its own sensors. The state of mobile robot relative to left obstacle can be represented as equation 1, d_L is the distance from the mobile robot to the left obstacle, N(Near) means that the distance between the obstacle and the mobile robot is less than or equal to the safe obstacle-avoided threshold μ. F(Far) means that the distance between obstacles and the mobile robot is greater than the safe obstacle-avoided threshold μ, and the mobile robot can ignore the influence of obstacles on it.

$$d_L = \begin{cases} \mathrm{F}, & d_L > \mu \\ \mathrm{N}, & d_L \le \mu \end{cases} \tag{1}$$

The target bias sampling strategy is oriented and improves the algorithm efficiency, but it is easy to drop into local extrema. According to the relative position of different obstacles and mobile robot, the mobile robot can be guided to avoid obstacles smoothly and reduce the probability of dropping into local extrema. The sampling strategy partition of target bias and obstacle-avoided guidance is shown in Figure 2.

Mark y-axis is the direction from the starting Angle 0º. Clockwise 0º~45º is area Ⅰ, 45º~90º is area Ⅱ, 90º~135º is area ⅠII, 135º~180º is area ⅠV, -180º~-135º is area V, -135º~-90º is area VⅠ, -90º~-45º is area VIⅠ, -45º~0º is area VIII.

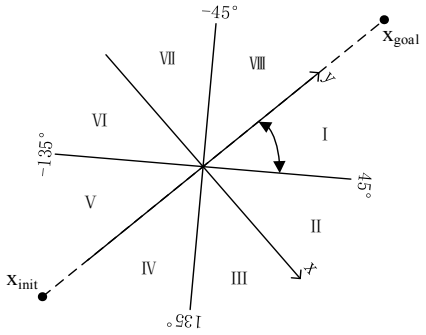

Fig. 2. The partition of target bias and obstacle-avoided guidance sampling strategy.

C. Establish Partition Heuristic Rules

According to local environment model of mobile robot and target bias and obstacle-avoided guidance sampling strategy, the specific partition heuristic rules are shown as Table 1.

TABLE I. PARTITION HEURISTIC RULES

d_F	d_B	d_L	d_R	Heuristics zone
N	N	N	N	extreme point
N	N	N	F	Ⅱ
N	N	F	N	Ⅶ
N	N	F	F	Ⅱ or Ⅶ
N	F	N	N	Ⅳ or Ⅴ
N	F	N	F	Ⅱ
N	F	F	N	Ⅶ
N	F	F	F	Ⅱ *or* Ⅶ
F	N	N	N	x_{goal}
F	N	N	F	Ⅰ
F	N	F	N	Ⅷ
F	N	F	F	x_{goal}
F	F	N	N	x_{goal}
F	F	N	F	Ⅰ
F	F	F	N	Ⅷ
F	F	F	F	x_{goal}

The rules comprehensively evaluate the information of environmental obstacles detected by the mobile robot and summarize the position of obstacles relative to the mobile robot in space. The partition heuristic rules of sampling node generation are proposed according to the relative position of the obstacle and the mobile robot. Not only the target bias strategy is cited to accelerate the convergence speed of the algorithm; but also the obstacle-avoided guidance strategy is added to avoid the random tree dropping into the local extrema.

III. Q-PRRT ALGORITHM

The Markov decision process model (MDP) of the partitioned heuristic RRT algorithm is established by

combining Q-learning and partition heuristic RRT algorithm. The action set and reward function of partitioned RRT algorithm are designed, and the value of the optimal sampling node is solved through Q value iteration, and finally the global optimal planning path is obtained.

A. MDP Modeling of Partition Heuristic RRT Algorithm

MDP modeling is to establish the learning method model of environment state to action mapping, including state space, action set, state-transition matrix, and reward function.

(1) State: each expanding node of the random tree is regarded as a state in the MDP model.The complete random tree is an n-dimensional vector, represented as T(s), and n is the number of nodes in the random tree.

(2) Action: The decision to generate the next node from the current node according to the partition heuristic rules.

(3) Reward: Evaluate the predominance and inferior position of the expanding node generated by the action, and evaluate the result as a reward value.

(4) state-transition matrix: The transition probability is updated continuously according to the reward value, and the state with a large reward value has a high transition probability.

The MDP modeling of partitioned heuristic RRT algorithm explores and decides repeatedly, updates the environmental information according to the feedback reward value, and then influences the decision process, and finally obtains the optimal solution.

B. Design of Action Set

According to the partition heuristic rules, random sampling points are obtained by exploring in the designated partition, and the action set is shown in equation 2:

$$s_{\text{rand}}=\begin{cases} s_{\text{rand}}, & a=0 \\ s_{\text{I}}, & a=1 \\ s_{\text{II}}, & a=2 \\ s_{\text{IV}}, & a=4 \\ s_{\text{V}}, & a=5 \\ s_{\text{VII}}, & a=7 \\ s_{\text{VIII}}, & a=8 \\ s_{\text{goal}}, & a=10 \end{cases} \tag{2}$$

When a=0, s_{rand} is a randomly generated point; a=1, s_{rand}=s_{I}, random sampling point is generated in I area, and so on when a=8, s_{rand}=s_{VIII}, random sampling point is generated in Ⅷ area; when a=10, s_{rand}=s_{goal}, random sampling point is the target point.

C. Design of Reward Function

The reward function is the key to determining Q-learning performance. A premium reward function will maximize the return on the problem as quickly as possible. In global path planning, the distance between mobile robot and target point and obstacles strongly associated with the planning effect. The designed reward function is:

$$R(s_{nearest},s_{new})=\begin{cases} \alpha(d_1-d_2) & ,d>\mu \\ \alpha(d_1-d_2)-\beta(d_3-d_4), & d\le\mu \end{cases} \tag{3}$$

d_1—distance between s_{nearest} to target point,
d_2—distance between s_{new} to target point,
d_3—distance between s_{nearest} to obstacles,
d_4—distance between s_{new} to obstacles,
d —distance between mobile robot to obstacles,
α and β are parameters, μ is the safe distance.

When the distance d between mobile robot to the obstacle is greater than a certain safe distance μ, the influence of distance d_3 and d_4 should be ignored, which can speed up the convergence of the algorithm.

D. Q Value Iteration

Q-learning is a value iteration algorithm. The value iteration algorithm calculates the value of each state-action pair and maximizes this value as the associated action is executed. Therefore, iterative refinement of each state value is the core of the Q-learning value iterative algorithm. Greedy strategy is to maximize the long-term reward of the action, which should not only be related to the current action feedback reward value, but also to the follow-up reward of the action. So value iteration is adopted to approximate the optimal solution.

For the partition heuristic RRT algorithm, Q-learning value iteration is used to calculate the Q-value of each node when executing actions of partition heuristic rules. The iteration equation is as follows:

$$Q_{\text{nearest}}=(1-\alpha)Q_{\text{nearest}}+\alpha\left[r(s_{nearest},a)+\gamma Q_{new}-Q_{nearest}\right] \tag{4}$$

Q_{nearest} is the Q-value in s_{nearest} state, Q_{new} is the Q-value in s_{new} state. α is the learning rate, γ is the discount factor, r(s_{neatest}, a) is the reward value obtained after snearest executes action a.

E. Q-PRRT Algorithm

TABLE II. PARTITION HEURISTIC RULES

Q-PRRT Algorithm

Input: Initialize random tree;
Create initial position s_{init}, target node s_{goal}, step size η;
Algorithm main loop:
while ($\|s_{\text{new}}$-$s_{\text{goal}}\|>\lambda$)
  Find the nearest node to the target node $s_{nearest}$;
  Obstacle detection: d_{F}、d_{B}、d_{L}、d_{R};
  while($\|\Delta\|>\theta$ or n>N)
    Execute a according to partition heuristic rules;
    Generate sampling node s';
    if obstacle_free(s_{new}, s_{nearest})
      Update Q-value:
      $Q_{nearest}=(1-\alpha)Q_{nearest}+\alpha\left[r(x_{nearest},a)+\gamma Q_{new}-Q_{nearest}\right]$;

```
        else
            return step 8;
        end
    end
    Select the action with the max Q-value;
    Generate new node s_new;
    Add s_new to the tree;
    return step 4;
end
Output: tree and path planning
```

On the basis of partition heuristic RRT algorithm, Q-learning method is adopted to establish a learning evaluation mechanism for each node, and a partitioned RRT path planning algorithm based on Q-learning is proposed. Detailed steps are shown in table 2.

In the Q-PRRT algorithm, the initial position of the mobile robot is sinit, the target position is sgoal, η represents the RRT algorithm step size, λ is the position accuracy, Δ is the Q value iteration precision, θ is a minimal constant, which determines the precision of Q-value iteration, n represents the number of iterations, and N is the maximum number of iterations.

Compared with traditional RRT algorithm, step 8 executes action a according to partition heuristic rules to generate random node s_{rand}. Step 9 explores the new node s' in step η along the straight line direction of $s_{nearest}$ to s_{rand}. Step 10 detects whether there is an obstacle between s' and $s_{nearest}$, if there are obstacles, return step 8 and re-select the action, else update Q value in step 11. In Step 7-15, loop iteration through Q-learning method until Q-value approaches the optimal solution or the iterations reaches the upper limit. Step 16-18, select the action a with the largest Q-value, generate the expanding node s_{new} through random sampling node, and add it into the random tree. Step 19 returns to step 4 to execute the next node learning generation process until the target point is found. Step 21 outputs the complete random tree to get the final planned path.

IV. SIMULATION ANALYSIS

The Q-PRRT algorithm is verified by Matlab simulation, and its performance is tested and analyzed. The simulation is executed in three types of static environments: (1) uniform obstacle environment; (2) narrow channel environment; (3) trap obstacle environment. RRT, GoalBias-RRT and Q-PRRT algorithms are used to plan the path respectively. Matlab simulation environment is designed as follows: map size is 100×100cm, step size is η =1cm, maximum number of iterations is N=100, initial starting point s_{init}=(5,5), target point is s_{goal}=(95,95). The simulation experiment planned a collision-free path from the starting point to the end point in different obstacle environments, the performance of the planned path obtained by each algorithm is evaluated and analyzed.

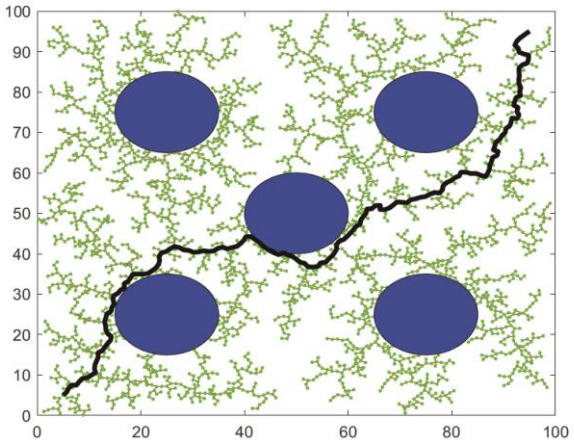

Fig. 3. The plan path of RRT in uniform obstacle environment.

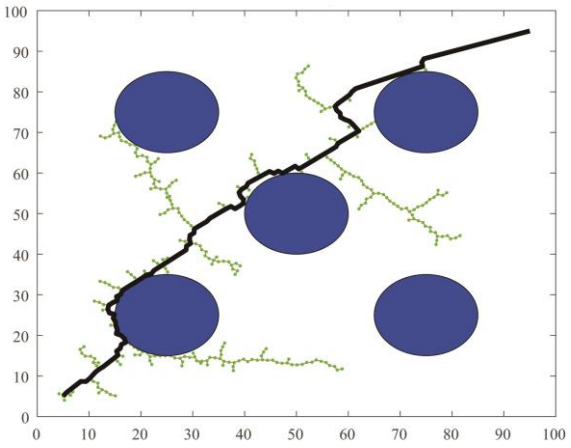

Fig. 4. The plan path of GoalBias-RRT in uniform obstacle environment.

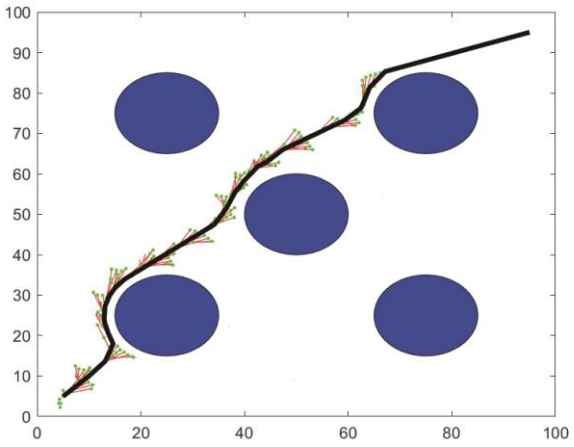

Fig. 5. The plan path of Q-PRRT in uniform obstacle environment.

Figure 3 to Figure 5 show the path planning results of RRT, GoalBias-RRT and Q-PRRT algorithms in a uniform obstacle environment. The number of RRT algorithm extension nodes is extremely high, the efficiency is extremely low, and the optimization of the planning path is really bad. The GoalBias-RRT algorithm greatly improves the efficiency of the algorithm by introducing goal-bias heuristic and the planning path is less optimized. The Q-PRRT algorithm has the best path planning, the expansion nodes are greatly reduced, and the planning path actively bypasses the obstacles, ensure the smoothness and optimality of the planning path.

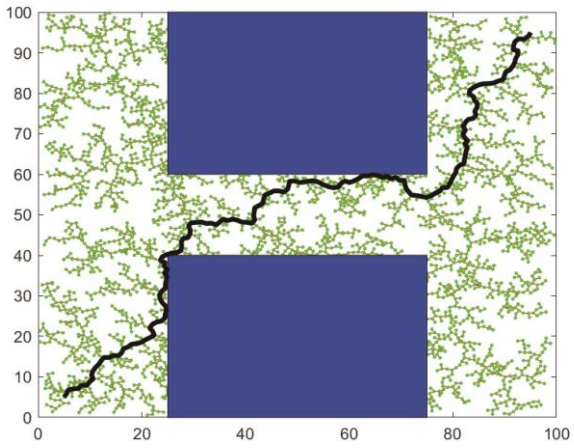

Fig. 6. The plan path of RRT in narrow channel environment.

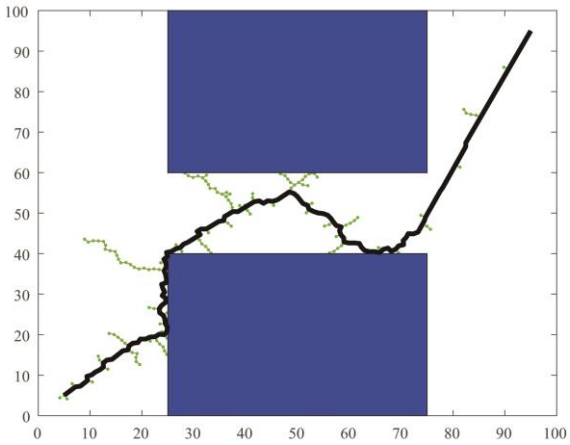

Fig. 7. The plan path of GoalBias-RRT in narrow channel environment.

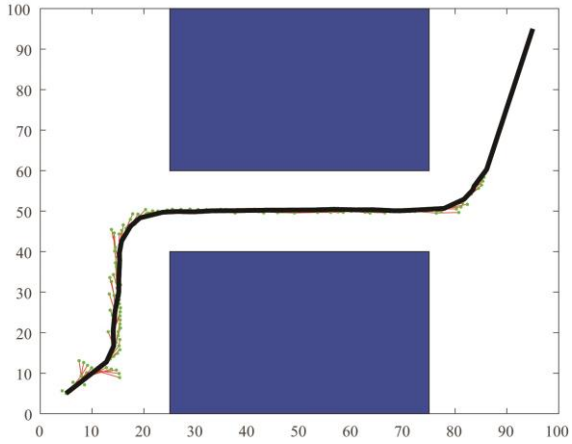

Fig. 8. The plan path of Q-PRRT in narrow channel environment.

Figure 6 to Figure 8 show the path planning results of RRT, GoalBias-RRT and Q-PRRT algorithms in the narrow channel environment. In the narrow channel environment, due to the small area of the narrow channel and the low probability of being captured by the node, the randomness of the expanding node of the RRT algorithm leads to very low exploration efficiency in the channel. Goalbiased-RRT algorithm improves the efficiency of exploration in narrow space by introducing target-bias heuristic strategy, but the efficiency is still low. The Q-PRRT algorithm is based on partition heuristic rules to obtain better quality extended nodes through learning, so that it obtains optimal and smooth path planning in narrow channels.

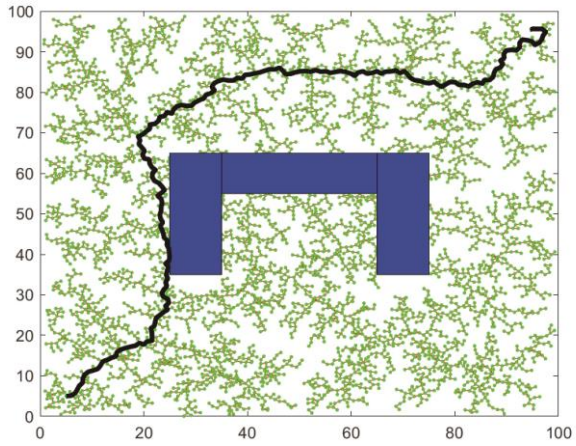

Fig. 9. The plan path of RRT in trap barrier environment.

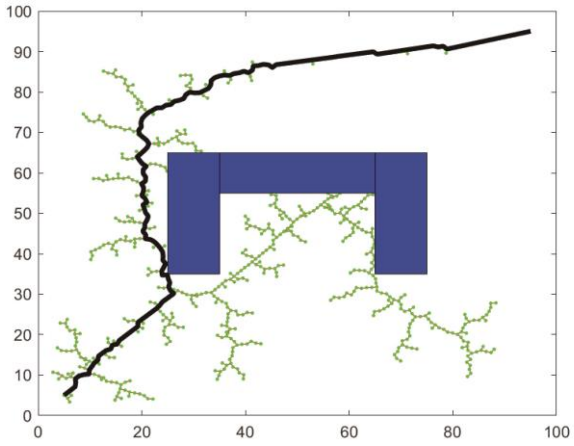

Fig. 10. The plan path of GoalBias-RRT in trap barrier environment.

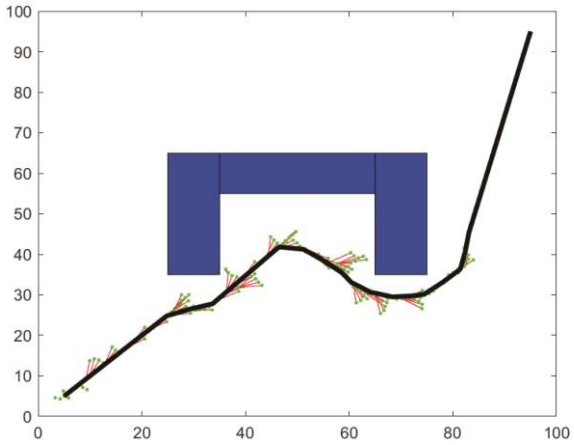

Fig. 11. The plan path of Q-PRRT in trap barrier environment.

Figure 9 to Figure 11 are path planning results of RRT, GoalBias-RRT and Q-PRRT algorithms in the trap barrier environment. RRT algorithm can get rid of traps, but the expansion efficiency and path optimization are really bad. GoalBias-RRT algorithm wastes a lot of efficiency in trapping cavity, and can only get out of the trap by the random sampling expansion mechanism with a certain probability.

Q-PRRT algorithm integrates target bias and obstacle avoidance guided sampling strategies, evaluates each node of the random tree based on q-learning method, and makes decisions on rational path planning, so as to successfully avoid traps and complete better performance path planning. Q-PRRT algorithm integrates target bias and obstacle-avoided guidance sampling strategies, the nodes of random tree are evaluated based on the Q-learning method, algorithm makes decisions and plans the path properly, successfully avoids traps and completes better performance path planning.

TABLE III. PARTITION HEURISTIC RULES

Algorithm		*Uniform obstacle*	*Narrow channel*	*Trap barrier*
Expanding nodes	RRT	4318	2840	3944
	GoalBias-RRT	438	327	705
	Q-PRRT	325	218	588
Expanding time	RRT	13.13	12.97	14.51
	GoalBias-RRT	1.66	1.52	3.45
	Q-PRRT	1.45	1.24	2.24
Path lentgh	RRT	168.06	180.19	185.11
	GoalBias-RRT	153.68	163.53	179.28
	Q-PRRT	136.63	156.06	165.04

Through the tests of three sets of simulation environments, it can be found that Q-PRRT algorithm is more advantageous than RRT and Goalbias-RRT algorithm in planning path optimality and avoiding dead zone due to the introduction of partition heuristic mechanism and learning method. Table 3 shows the performance comparison of the path planning of the three algorithms. The unit of the expanding node is (pc), the unit of the expanding time is (s), and the unit of the path length is (cm). The tabular data is averaged based on 10 simulation results.

Through analyzing and summarizing the comparison of simulation data. The performance of the RRT algorithm is highly probabilistic due to its random exploration sampling mode without heuristic mechanism, but it also has certain passing ability and escaping ability. GoalBias-RRT algorithm is based on target bias guidance with a certain probability, which is better than RRT algorithm in terms of convergence speed and expanding node number, however, it is easy to drop into the local extrema in the trap barrier environment, so it is necessary to use the random exploration sampling mechanism with a certain probability to assist the algorithm to escape the trap. Q-PRRT algorithm establishes partition heuristic rules by introducing target bias and obstacle avoided guidance sampling strategy, the reward function of the globally optimal path is applied to evaluate each node of the random tree based on Q-learning method, Then retain path nodes and eliminate redundant nodes based on greedy strategy. It improves path search efficiency and obstacle avoidance ability and also guarantees the optimality of the global planning path and obtains a smoother planning path. The performance of the number of expanding nodes, expanding time, and the length of planned path have been greatly improved.

V. CONCLUSION

A partitioned heuristic RRT path planning algorithm based on Q-learning was proposed. Partition heuristic rules based on target bias and obstacle-avoided guidance sampling strategy was established to enhance the guidance and obstacle avoidance ability of path planning. Q-learning method was applied to establish the reward function of the global optimal path. Path nodes are retained and redundant nodes are eliminated by evaluating each node of the random tree based on greedy strategy. Thereby, the global optimal planning path was guaranteed and the smoothness of the path was improved. The simulation results show that the Q-PRRT algorithm is superior to the traditional RRT and GoalBias-RRT algorithm in terms of expanding nodes, expanding time and path length.

REFERENCES

[1] Eele A J, Richards A. Path-Planning with Avoidance Using Nonlinear Branch-and-Bound Optimization[J]. Journal of Guidance, Control, and Dynamics, 2009, 32(2):384-394.

[2] Sun X, Yeoh W, Koenig S. Moving target D* Lite.[C]// International Conference on Autonomous Agents & Multiagent Systems: Volume. International Foundation for Autonomous Agents and Multiagent Systems, 2010.

[3] Welzl E. Constructing the visibility graph for n-line segments in O(n2) time[J]. Information Processing Letters, 1985, 20(4):167-171.

[4] Dijkstra E W. A Note on Two Probles in Connexion with Graphs[J]. Numerische Mathematics, 1959, 1(1):269-271.

[5] Hart P E, Nilsson N J, Raphael B. A Formal Basis for the Heuristic Determination of Minimum Cost Paths[J]. IEEE Transactions on Systems Science and Cybernetics, 1968, 4(2):100-107.

[6] Xian-Quan M, Ying-Nan Z, Qing X. Application of Genetic Algorithm in Path Planning[J]. Computer Engineering, 2008, 34(16):215-217.

[7] Glasius R, Komoda A, Gielen S C A M. Neural Network Dynamics for Path Planning and Obstacle Avoidance[J]. Neural Networks, 1995, 8(1):125-133.

[8] Kavraki L, Svest ka P, Latombe J C. Probabilistic roadmaps for path planning in high-dimensional configuration space[J]. IEEE Trans. on Robotics and Automation, 1996, 12(4):566–580

[9] LaValle S M. Rapidly-exploring random trees: A new tool path planning[R]. Ames, USA: Iowa State University, 1998.

[10] Léonard Jaillet, Porta J M. Path planning under kinematic constraints by rapidly exploring manifolds[J]. IEEE Transactions on Robotics, 2013, 29(1):105-117.

[11] LIU Huajun, YANG Jingyu, LU Jianfeng, et al. Overview of mobile robot motion planning[J]. Engineering Science, 2006, 8(1): 85-94.

[12] Kuffner J J, Lavalle S M. RRT-connect: An efficient approach to single-query path planning[C]// Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065). IEEE, 2002.

[13] LaValle S M. Randomized Kinodynamic Planning[J]. The International Journal of Robotics Research, 2001, 20(5):378-400.

[14] LaValle S M, Kuffner J J. Rapidly-exploring random trees: Progress and prospects[C]//4th International Workshop on Algorithmic Foundations of Robotics. Wellesley, USA: A K Peters, 2000: 293-308.

[15] Jordan M, Perez A. Optimal Bidirectional Rapidly-Exploring Random Trees[J]. 2013.

[16] Qureshi A H , Ayaz Y . Intelligent bidirectional rapidly-exploring random trees for optimal motion planning in complex cluttered environments[J]. Robotics and Autonomous Systems, 2015, 68:1-11.

[17] Zhang X , Felix Lütteke, Ziegler C , et al. Self-learning RRT* Algorithm for Mobile Robot Motion Planning in Complex Environments[J]. 2016.