Industrial Engineering and Innovation Management (2022)
Clausius Scientific Press, Canada
DOI: 10.23977/ieim.2022.050105
ISSN 2522-6924 Vol. 5 Num. 1

# *Embed Behavior Decision Making into Ship Collision Avoidance Path Planning Based on Ant Colony and Q-Learning Algorithm*

**Jiayin Hu, Duowen Yan, Jian Zheng**

*Transport and Communications College, Shanghai Maritime University, Shanghai 201306, China*

***Keywords:*** Dynamic collision avoidance, Path optimization, Improved ant colony algorithm

***Abstract:*** In order to solve the problems of ship path planning and slow convergence speed of traditional ant colony algorithm under different encounter situations in the dynamic complex water surface, a ship turns decision model was proposed based on Q-learning algorithm, and the transfer probability formula of ant colony algorithm was improved. The improved ant colony algorithm can quickly calculate the safe and realistic collision avoidance path. This algorithm uses Q-learning to plan the optimal collision avoidance path by ship improved ant colony algorithm. Simulation results show that the algorithm can converge quickly and calculate the optimal collision avoidance path. This method can be effectively applied to ship collision avoidance path planning in complex waters.

## 1. Introduction

Nowadays, more than 85% of goods in international trade are transported by sea. With the development of the shipping industry, ships are becoming more extensive and more specialized, and the sailing speed is getting faster and faster. In addition, the maritime traffic environment is becoming increasingly complex, and the traffic density is increasing [1]. It is an effective way to deal with fast collision avoidance efficiently and correctly to realize automatic collision avoidance by improving collision avoidance means and reducing dependence on subjective judgment. Therefore, ship collision avoidance decision systems research is significant to ship safety.

At present, path planning methods mainly fall into the following three categories: traditional path planning methods, path planning methods based on biological intelligence, and path planning methods based on reinforcement learning. The first type of path planning methods includes the A* algorithm [2], Dijkstra algorithm [3], artificial potential field method [4]. All these algorithms have the problem of low search efficiency in a complex environment. The second type of path planning methods based on biological intelligence mainly include particle swarm algorithm [5], ant colony algorithm [6], genetic algorithm [7], neural network [8]. Although these algorithms have good global optimization performance, they quickly fall into local optimization errors due to their large number of parameters and slow convergence. The traditional ant colony algorithm still has many shortcomings in path planning. Zhang Cheng et al. proposed an improved ant colony algorithm, which can solve the shortcomings of traditional ant colony algorithm in path planning, such as poor convergence rate, local optimization, and poor solution quality [9]. You et al. designed a new dynamic search induction operator through the dynamic search model to improve the performance of the ant colony algorithm by reducing the operation time and improving the solution accuracy [10].

The third type of reinforcement learning-based methods includes the Q-learning algorithm [11], Sarsa algorithm [12]. Compared with the first two kinds of algorithms, the algorithm based on reinforcement learning has the advantage of solid adaptability in a complex environment. In recent years, machine learning has attracted more and more attention from scholars in ship collision avoidance planning. For example, In 2018, Professor Guo Chen's team used the A* algorithm and deep competitive Q-learning algorithm to design collision avoidance strategies for ships in restricted waters [13]. In 2018, Professor Zhang Xinyu's team adopted the collision avoidance decision method of profound reinforcement learning disabilities to solve ships' intelligent collision avoidance problems in uncertain environments [14]. The above research results reflect that the deep learning method has a good effect on learning international Regulations for Collision Avoidance at Sea and the decision-making of collision avoidance.

This paper proposes an improved ant colony algorithm combined with the Q-learning principle based on the careful consideration of previous studies and shortcomings analysis. This algorithm uses q-learning to plan the optimal collision avoidance path by ship improved ant colony algorithm.

## 2. Environmental Modeling

### 2.1 Grid Method

The raster method was used to establish the environment model for the simulation operation environment. G was denoted as the limited movement area of the ship on the two-dimensional plane. The raster numbers within the area and the generated raster map environment are shown in Figure 1. Mark and number the grid map from top to bottom and from left to right, taking the lower-left corner of G as the origin of coordinates, the horizontal axis as X-axis, and the vertical axis as Y-axis. There are a limited number of obstacle grids located in the relevant area, represented by black grids in Figure 1 and free grids by white grids.

### 2.2 Dynamic Barrier Environment Change Setting

After the static and dynamic obstacle model is established based on the raster chart, an update mechanism is needed to ensure the regular operation of the model to reflect the obstacle's change. To meet the dynamic path planning in an unknown environment, the characteristics of this section is to design a kind of time step as the index of dynamic obstacles changes rule, i.e., in a particular time step area, the dynamic obstacles as static obstacles for obstacle avoidance process, within the global path planning can have multiple time zones for obstacle collision avoidance. Assuming that the time required for local planning is t, then the time required for obstacle environment change is $t \leqslant n * T$. The value of n is related to real-time. The higher the real-time requirement is, the smaller the value of n is.

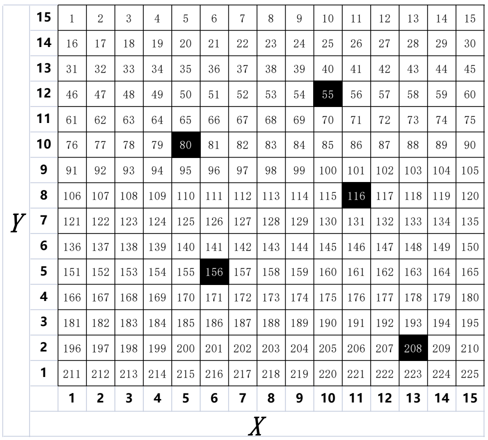


*Fig .1 Grid Map*

## 3. Ship Behavior Decision into Collision Avoidance

### 3.1 Behavior Decision Based on q-Learning Algorithm

According to the *Convention on the International Regulations for Preventing Collisions at Sea*(COLREGS), when two motor ships meet, there are four types of encounter: head-on, crossing from the starboard, crossing from the port, overtaking. When the ship needs to take action to avoid collision, it usually achieves the effect of collision avoidance by turning, and the rules of collision avoidance and turning are shown in Figure 2.

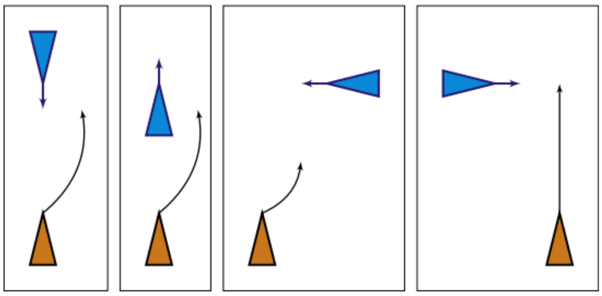

*Fig.2 Rules for Avoiding Collision*

In the complex maritime navigation environment, to enable ASV to have the ability of autonomous obstacle avoidance, this paper adopts the steering collision avoidance path planning model based on Q-learning to achieve autonomous obstacle avoidance of ASV. Q-learning as the collision avoidance algorithm can quickly feedback results and avoid the increase of collision risk caused by an excessive delay caused by repeated decisions.

First of all, we according to the complexity of the collision scene, set a limited quantity of state and collision avoidance behavior, to state the number of columns, the number of collision avoidance behavior as the line, get the initial Q (s, a) table, which s is ASV time state, a is the action strategies such as turn left, turn right, forward, and so on, Q (s, a) is of ASV said the status and action of the matrix. The Q value represents the reward after the current action is performed.

Secondly, the setting of behavioral decision mechanisms includes action strategy and state reward and punishment. Rewards were added to the virtual training environment, and ASV updated the environment after each reward and repeated the next training round. Finally, Q table output, according to the size of the value of the last column of Q table, makes a decision. The decision formula is as follows, where Pss' is the transition probability from state s to s' and represents the discount factor:

$$\pi_{st} = arg\,max\,P_{ss'} \left[\tau_{ij} + \gamma Q(s',a)\right]$$

Execute the decision action, observe the state at the next moment, judge the task stage and the surrounding environment, select the return mode suitable for the current state, and get the return value r. The formula is as follows:

$$Q(s,a) \leftarrow Q(s,a) + \alpha\left[r + \gamma max_{a'} Q(s',a') - Q(s,a)\right]$$

where $\alpha$ is the learning efficiency, which determines the learning percentage of the current error; $Q(s,a)$ is the estimated value of Q in the current state, and $Q(s',a')$ is the actual value of Q at the next moment in the current state.

## 3.2 Parameter Setting of q Learning Algorithm

We added the ASV reward value to the training environment based on the relevant collision avoidance requirements of COLREGS.

As in COLREGS, when two motorized vessels meet in opposite or near opposite directions and pose a risk of collision, they shall each pass to the right or to the port side of the other vessel.

In the algorithm should be set to: in the case of head-on, setting the ship to the right of the reward value equal to $r_{a1}$, left to $r_{a2}$;in the case of crossing from port, there is a ship coming from starboard, the incentive value for turning right is set to $r_{b1}$, and turning left is set to $r_{b2}$. This method is used to define the course reward value in the other two cases to construct the decision-making behavior criterion of the unmanned ship.

*Table 1 the Setting Of Direction Reward*

| situation | action1 reward**:** | action2 reward**:** |
|---|---|---|
| a**.** Head-on | $r_{a1}$ | $r_{a2}$ |
| b**.** Crossing from port | $r_{b1}$ | $r_{b2}$ |
| c**.** Crossing from starboard | $r_{c1}$ | $r_{c2}$ |
| d**.** Overtaking | $r_{d1}$ | $r_{d2}$ |

# 4. Ship Path Planning Based on Ant Colony Algorithm

## 4.1 Traditional Path Planning Model Based on Ant Colony Algorithm

Ant colony algorithm is a probabilistic algorithm that simulates the foraging process of ants. It has the characteristics of heuristic search and positive feedback information. It can find the shortest path from the origin through several demand points back to the origin, widely used in the traveling salesman problem. Let the number of ants be m, the number of points is n, and the distance between points be $d_{ij}$(i,j=1, 2,...,n),the pheromone on the connection between i and j at time t is $\tau_{ij}(t)$,the distance heuristic function is $\eta_{ij}(t)$. Therefore, the transfer probability of the No.k ant from i to j at time t is:

$$p_{ij}^{k}(t) = \begin{cases} \dfrac{(\tau_{ij}(t))^{\alpha} * (\eta_{ij}(t))^{\beta}}{\Sigma(\tau_{il}(t))^{\alpha} * (\eta_{il}(t))^{\beta}}, ?j \in a_m \\ 0??j \notin a_m \end{cases}$$

$$a_m = \{1,2\ldots,n\} - B_m$$

Where $a_m$ is the node-set that the ant can select, ; $B_m$ is taboos table (the point ants walk); α is the pheromone heuristic factor, and represents the degree of influence of pheromone on subsequent ant path finding; β is the expectation heuristic factor, represents the influence of heuristic information on ants; $\eta_{ij}$ is the heuristic function, represents the heuristic information of the spacing from j to target E in the current ant domain;

$$\eta_{ij}\left(t\right)=\frac{1}{d_{jE}}$$

Heuristic information and pheromone concentration are the key criteria for ant path selection. Since pheromone concentration will disappear with time, the algorithm will update local and global pheromones. Pheromone update formula after iteration is:

$$\tau_{ij}\left(t+1\right)=\left(1-\lambda\right)\tau_{ij}\left(t\right)+\tau_0$$

$$\tau_{ij}\left(t+1\right)=\left(1-\rho\right)\tau_{ij}\left(t\right)+\sum_{k=1}^{m}\Delta\tau_{ij}^{k}\left(t\right)$$

$$\Delta\tau_{ij}^{k}\left(t\right)=\begin{cases}\dfrac{Q}{L_m},?Though\,node\,ij\,at\,time\,t\\ 0?,?Others\end{cases}$$

Where $1-\lambda$ is the local pheromone residual coefficient; $\rho$ is the global pheromone volatility coefficient; $\tau_{ij}\left(t\right)$ is the pheromone between iteration nodes ij at time t; $\tau_0$ is the initial pheromone concentration, is a small constant; $\Delta\tau_{ij}^{k}\left(t\right)$ is the pheromone increment after one iteration of ant colony; Q is pheromone intensity; $L_m$ is the path length of ant M in this cycle. According to the above equation, the pheromone intensity left between nodes ij is closely related to path optimization. The shorter the path, the higher the pheromone concentration left, and the ant colony will follow the path with higher concentration.

## 4.2 Improved Ant Colony Algorithm

Traditional ant colony algorithm is mostly used to solve the shortest path problem, which is difficult to realize the limit of course and speed. It is more difficult to deal with the problem of ship navigation safety based on COLREGS in complex waters. Therefore, this paper combines the ant colony algorithm and the Q-learning algorithm to construct the model of ship behavior decision making and collision avoidance.

Moreover, the path planned by ant colony algorithm in grid environment has many turning times and large cumulative turning Angle. To solve these problems, Liu Xinyu et al. [15] proposed the midpoint smoothing mechanism method, which smoothed the Angle of the planned path, so that the path was relatively smooth. Inspired by this, the midpoint smoothing mechanism can be used to avoid large angle turning behavior during navigation, which is closer to the real experimental situation.

The main steps of the algorithm are as follows:

Step 1: According to the static environment, the grid method is used to model the ASV operating environment, and the starting point and target end point are set according to the task;

Step 2: Generate the globally optimal path through ant colony algorithm according to the starting point and target end point coordinates of ASV.

Step 3: ASV makes a decision on ship behavior based on Q-Learning algorithm when the dynamic obstacles are detected in the safe distance.

Step 4: The distance between the detection and the dynamic obstacle is used to judge whether the obstacle avoidance is finished or not

Step 5: After obstacle avoidance, it returns to the original path and moves to the target point.

## 5. Simulation and Analysis

The deepwater channel of the Yangtze Estuary is one of the waters with the most intensive ship flow, the most complicated navigation conditions, and the most challenging management in the world. It is the only throat for large ships to enter and exit the Yangtze Estuary. Therefore, the decision-making problem of intelligent collision avoidance is a crucial problem to be solved urgently. Therefore, this paper will make simulation experiment analysis in the Yangtze Estuary deepwater channel to provide a feasible solution to ships' intelligent collision avoidance decision problem. The actual sea conditions of the deepwater channel in the Yangtze Estuary are shown in Fig. 4. Obstacle information can be obtained after rasterizing environmental information in the electronic chart and AIS system, as shown in Fig. 5. Four related static electronic charts are presented according to the dynamic obstacle updating mechanism, as shown in Fig. 6.

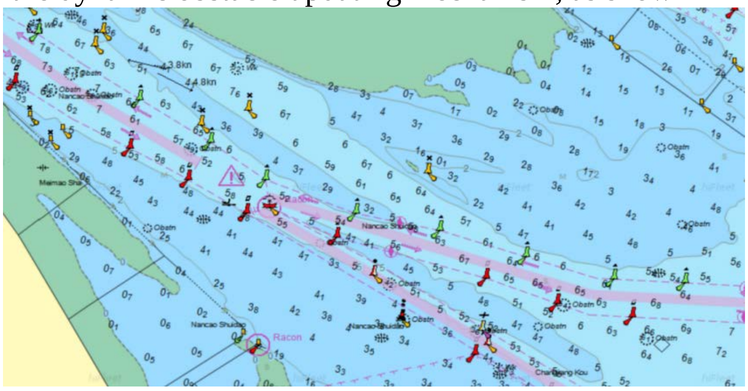

*Fig.4 Field Map of Sea*

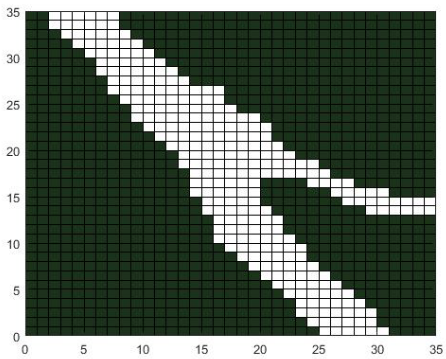


*Fig.5 Simplified Static Grid Map*

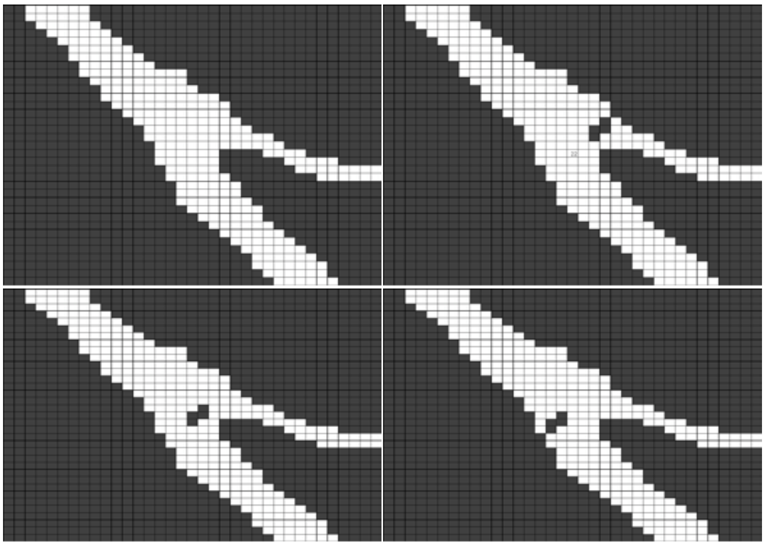

*Fig.6 Simplified Grid Map*

The simulation takes the south Channel of Yangtze River Estuary as the experimental scene (regional scope: 122.106833°E~122.286833°E,29.817667°N ~29.893333°N).

The experiment assumes that the ship starts at (121°55.28E, 31°7.34N) and ends at (122°6.04 E, 31°1.52N). The ship is 30m long, with an initial heading of 225° and the ship speed is 9 kn. The environmental information is provided by electronic chart. The water depth in this area meets the requirements of intelligent ship. The ship visual field constraint is $\theta$=45°, and the maximum search step is N=4. The default estimated collision point is marked in green.

Table 2 Shows the Related Parameters of the Improved Ant Colony Algorithm. Table 3 Shows the Related Parameters of q-Learning Algorithm.

*Table 2 the Setting Of Ant Colony Parameter*

| The parameter name | The parameter value |
|---|---|
| Pheromone factor weight | 1 |
| Heuristic factor weights | 7 |
| Pheromone evaporation coefficient | 0.3 |
| Pheromone concentration enhancement coefficient | 1 |

*Table 3 the Setting Of q-Learning Parameter*

| The parameter name | The parameter value |
|---|---|
| number of states | 10 |
| epsilon | 0.9 |
| alpha | 0.1 |
| lambda | 0.9 |
| refresh time | 0.01 |
| episodes | 10 |

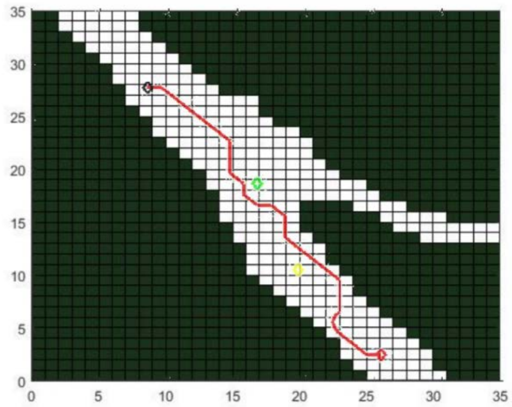


*Fig .7 Simulation Result*

Figure 7 shows all dynamically planned paths after all paths are planned. It can be seen from the figure that the algorithm can avoid static and dynamic obstacles and realize the combination of dynamic planning and autonomous obstacle avoidance. The obstacle avoidance mechanism causes some fluctuations in the planned paths. At the same time, when the path display near the collision point is set, the ship can turn right to avoid the target ship according to COLREGS rules, and the collision avoidance behavior decision of the q-learning algorithm is in line with the actual navigation scene. The collision simulation scene effectively verifies the feasibility of the collision avoidance system. This algorithm can give a collision-avoidance path considering the actual navigation rules and ship safety. Traditional ant colony algorithms cannot match the complexity of the external offshore environment, resulting in low planning success rate and large planning time. The algorithm designed in this paper can adapt to different environmental complexity, improve the planning success rate, and reduce time. In addition, by analyzing the experimental results, the environmental information provided by the electronic chart. AIS is transformed into a grid map after discrete processing, which significantly simplifies the simulation environment and can be used for the improved ant colony algorithm to realize planning in an extensive range, select the optimal route, and accurately avoid static and dynamic obstacles to ensure the safe navigation of ships.

## 6. Conclusion

This paper presents an improved ant colony dynamic path planning method based on ant colony combined with Q-learning algorithm in dynamic path planning in a complex dynamic environment. The Q-learning algorithm is used to learn the collision avoidance behavior of ships. The collision avoidance decision of ships conforming to the actual ship navigation rules is realized, and the situation inconsistent with the international collision avoidance rules is avoided. The effectiveness of the improved algorithm is verified by simulation. In the follow-up study, we will conduct the subsequent research on efficiently and safely avoiding a collision when multiple ships meet.

## Acknowledgement

This research is supported by the Project of Science and Technology of Shanghai(No. 21692193000)

## References

*[1] KANG Yu tao, ZHU Da qi. Review of research on collision avoidance path planning for ships. [J]. Ship and Ocean Engineering, 2013, 42(5):141- 145.*
*[2] WANG Dianjun.Indoor mobile-robot path planning based on an improved A* algorithm. [J]. Journal of Tsinghua University (Science and Technology), 2012,52(8):1085-1089.*
*[3] STENTZ A. The D* algorithm for real-time planning of optimal traverses [J]. 1994(2):586-593.*
*[4] XU B,ZHANG J,WANG C. A real-time obstacle avoidance method for multi-AUV cluster based on artificial potential field [J]. Chinese Journal of Ship Research,2018,13(6):66-71.*
*[5] JIA Huiqun, WEI Zhonghui, HEXin.Path planning based on improved particle swarm optimization algorithm.[J].Journal of agricultural machinery,2018,49(12):371-377.*
*[6] FU Zhenqiu, JI Guang, YANG Ying.AUV three-dimensional path planning method based on improved ant colony optimization and particle swarm optimization[J].Ship Science and Technology,2018,40(1):72-77.*
*[7] TSOU M, KAO S, SU C. Decision support form genetic algorithm for ship collision avoidance route planning and alerts [J]. Journal of Navigation, 2010,63(1):167-182.*
*[8] Yang S.X., Meng M.Q.-H. Real-time collision-free motion planning of a mobile robot using a Neural Dynamics-based approach[J].Transactions on Neural Networks,2003,14(6):p.1541-1552.*
*[9] Zhang ChengLing, YouzhuChen, Mengyuan. Path planning of mobile robot based on an improved ant colony algorithm[J]. Journal of Electronic Measurement and Instrumentation, 2016, 30(11):7.*
*[10] YOU Xiaoming, LIU Sheng, LU Jinqiu. Ant colony algorithm based on dynamic search strategy and its application on path planning of robot[J]. Control and Decision, 2017, 32(3):5.*
*[11] Wen, Shuhuan, Chen, et al. The Q-learning obstacle avoidance algorithm based on EKF-SLAM for NAO autonomous walking under unknown environments[J]. Robotics and Autonomous Systems, 2015.*
*[12] MA Pengwei, PAN Dilin. SARSA(λ) Algorithm Based on Heuristic Function[J]. Computer and Digital Engineering, 2016, 44(5):4.*
*[13] SHEN Haiqing, GUO Chen, LI Tieshan.. An intelligent collision avoidance and navigation approach of unmanned surface vessel considering navigation experience and rules[J]. Journal of Harbin Engineering University, 2018, 39(6):8.*
*[14] WANG Chengbo, ZHENG Xinyu, ZHANG Jiawei. Method for intelligent obstacle avoidance decision-making of unmanned vessel in unknown waters[J]. Chinese Journal of Ship Research,2018,13(06):72-77.*
*[15] LIU X Y, TAN L M, YANG C X, et al. Self-adjustable dynamic path planning of unknown environment based on ant colony-clustering algorithm [J]. Journal of Frontiers of Computer Science and Technology, 2019, 13(5): 846-857.*