ISSN 2616-5767 Vol.7, Issue 6: 155-161, DOI: 10.25236/AJETS.2024.070622

Three Dimensional Path Planning System for Unmanned Aerial Vehicles Based on Reinforcement Learning Algorithm

Zhengxiang Huang[1,*], Yong Tian[1]

[1]*College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, 211106, China*
[*]*Corresponding author*

Abstract: *In response to the problem of low efficiency and weak obstacle avoidance ability in finding the optimal or suboptimal path for 3D path planning of drones in complex dynamic environments, this paper uses Q-learning algorithm to complete the 3D path planning of drones, aiming to improve their path planning and obstacle avoidance capabilities. Firstly, by constructing a three-dimensional gridded environment model, the system calculates the reward for each state under the influence of natural environment and obstacles, and then guides the drone to avoid obstacles and find the optimal path. The system uses the ε - greedy strategy for exploration and learning, optimizing decisions by continuously updating the Q-table value table. The experimental results show that the drone has a success rate of 93.3% in obstacle avoidance in complex and multi obstacle scenes. Moreover, in terms of average path length, the Q-learning algorithm has shortened it by approximately 20.00%, 11.45%, and 40.39% compared to ant colony algorithm, A* algorithm, and RRT algorithm, respectively. In dynamic wind speed environments, the Q-learning algorithm reduces the path length by about 4% to 11% compared to other algorithms, further demonstrating its effectiveness and advantages in complex environments.*

Keywords: *Unmanned Aerial Vehicle (UAV), Reinforcement Learning, Q-learning, 3D Path Planning, Markov Decision Process (MDP), Autonomous Flight, Obstacle Avoidance System*

1. Introduction

In recent years, with the rapid development of drone technology, its application in military and civilian fields has become increasingly widespread. Due to their small size, flexible maneuverability, and multi-sensor integration, drones have become an ideal platform for executing complex tasks. However, when performing autonomous flight and mission planning in dynamic and complex 3D environments, drones face many challenges, especially in terms of flight constraints, obstacle avoidance capabilities, and energy consumption. Traditional path planning algorithms such as ant colony algorithm, A* algorithm, and RTT algorithm can solve some problems in static scenarios, but their adaptability and efficiency are limited when facing complex environments.

Traditional path planning algorithms such as ant colony algorithm, A* algorithm, and RRT algorithm perform well in ordinary static environments, but their adaptability and efficiency are limited when dealing with a large number of obstacles [1-2]. In contrast, Q-learning algorithm is adept at handling complex, adaptive, and multi-step optimization problems, particularly suitable for complex dynamic obstacle avoidance and path planning tasks. By continuously exploring and learning the values of states and actions in the environment, Q-learning can autonomously find the optimal path and dynamically adjust obstacle avoidance strategies. This demonstrates significant advantages in its application in autonomous drone flight.

The organizational structure of the paper is as follows: Firstly, the introduction elaborates on the research background and limitations of traditional methods in the field of unmanned aerial vehicle path planning; the second part introduces relevant research in the field of drone path planning, providing a theoretical basis for the method proposed in this paper; the third part elaborates on the design and implementation of a 3D path planning system based on Q-learning, including system architecture design, construction of a 3D grid environment model, and design of a reward mechanism; the fourth part verifies the effectiveness of the proposed method through experiments and compares its performance with ant colony algorithm, A* algorithm, and RRT algorithm; finally, the conclusion section summarizes the

ISSN 2616-5767 Vol.7, Issue 6: 155-161, DOI: 10.25236/AJETS.2024.070622

research findings, explores the potential of Q-learning in drone path planning, and proposes future research directions.

2. Related Work

Various methods have been proposed for three-dimensional path planning of unmanned aerial vehicles (UAV), with a focus on optimization based techniques, uncertainty modeling, and the application of reinforcement learning algorithms. For example, Kiani et al. proposed a multi UAV 3D path planning method based on the Grey Wolf Algorithm, which addresses path planning problems through the Incremental Grey Wolf Optimization (I-GWO) Algorithm and Extended Grey Wolf Optimization (Ex-GWO) Algorithm, significantly improving path cost and convergence speed, with a 36.11% increase in path cost compared to other algorithms [3]. Lv et al. proposed a hybrid algorithm (HGEOGWO) combining the Golden Eagle optimizer and the Grey Wolf optimizer, which was applied to multi UAV 3D path planning in power inspection and showed significant optimization performance in different test cases [4]. This type of path planning method based on heuristic algorithms shows good efficiency, but has weak adaptability in dynamic environments.

On the other hand, A* and its variant algorithms also demonstrate good performance in three-dimensional environments, especially in obstacle avoidance and path length optimization. Mandloi et al. conducted a detailed study on 3D path planning using the A* algorithm and its extensions such as Theta and Lazy Theta, demonstrating that these algorithms can provide relatively optimal paths in different dimensions and obstacle complexities, but have limitations in real-time response capabilities [5]. Zammit and van Kampen studied real-time path planning in dynamic uncertain environments. By comparing A* and RRT algorithms, they found that A* performed better than RRT in complex scenarios and had a higher success rate in path safety and real-time performance [6].

Reinforcement learning algorithms have shown great potential in path planning in dynamic environments, especially in drone path optimization and obstacle avoidance. The improved Q-learning algorithm proposed by Wang et al. has made significant progress in dynamic obstacle avoidance. The algorithm improves convergence speed and path accuracy by introducing priority weights, and can more effectively handle complex multi constraint environments [7]. Shang et al. combined RRT algorithm and Q-learning to achieve dynamic step size adjustment in path planning algorithm, improving the speed of initial path generation, and achieving path smoothing through bidirectional pruning and B-spline curves [8]. Bonny et al. combined Q-learning with the bee algorithm and validated the robustness and effectiveness of the method through experiments in both static and dynamic environments [9].

In addition, some studies have further improved the adaptability and efficiency of Q-learning in different applications. The mild conservative Q-learning (MCQ) proposed by Lyu et al. performed well in offline reinforcement learning, particularly in balancing value function generalization and overestimation problems. The algorithm demonstrates good transfer performance from offline to online [10]. Tran et al. applied Q-learning in cloud computing virtual machine migration, effectively balancing cost and service quality [11]. These studies provide methodological support for the application of reinforcement learning based path planning in the three-dimensional environment of unmanned aerial vehicles.

Overall, although traditional heuristic algorithms and A* algorithms have advantages in path optimization, reinforcement learning algorithms, especially Q-learning and its improved methods, demonstrate superior adaptability and flexibility in complex dynamic environments, helping to improve the obstacle avoidance ability and path efficiency of drones. Based on this, this paper utilizes the advantages of Q-learning algorithm in UAV 3D path planning, aiming to enhance the path planning and obstacle avoidance capabilities of UAV 3D path planning system in complex dynamic environments.

3. Method

3.1 System Architecture

The overall architecture of the unmanned aerial vehicle 3D path planning system is shown in Figure 1:

ISSN 2616-5767 Vol.7, Issue 6: 155-161, DOI: 10.25236/AJETS.2024.070622

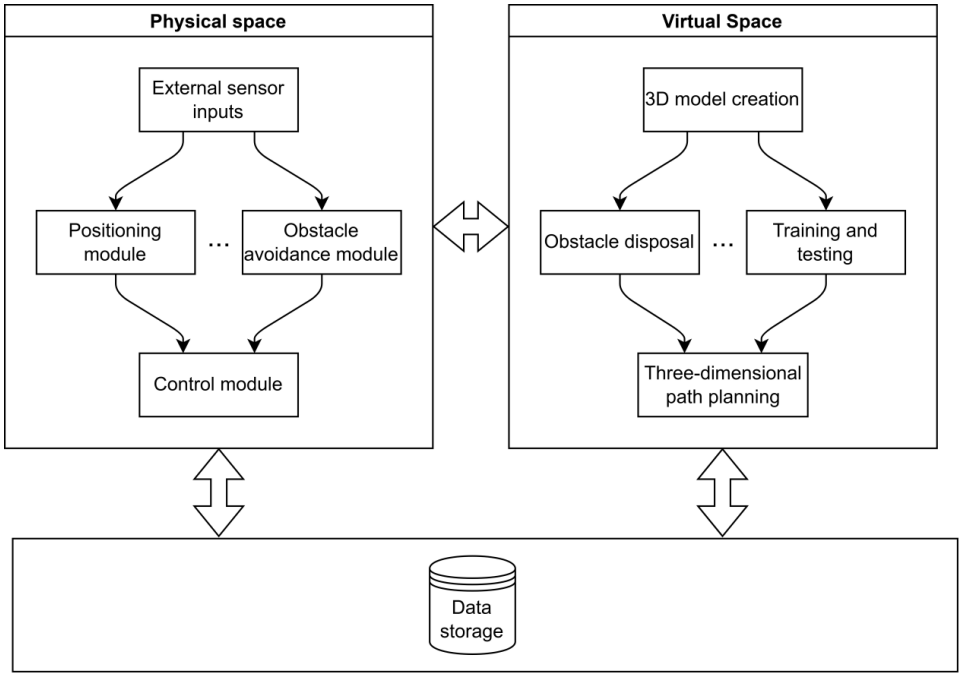

Figure 1: Overall system architecture

From Figure 1, it can be seen that the three-dimensional path planning system architecture of the drone consists of multiple modules, and path optimization is performed through reinforcement learning algorithms. The system obtains environmental data through external sensor input modules. After receiving sensor data, the information enters the obstacle avoidance module and the positioning module respectively. The obstacle avoidance module processes obstacle information to ensure that the drone can recognize and avoid obstacles during flight; the positioning module determines the location of the drone. The information from both is transmitted to the control module, which is responsible for integrating obstacle avoidance and positioning data to control the path selection of the drone, ensuring that the drone can move along a safe and efficient path. At the same time, the system generates a three-dimensional array model of the drone flight environment based on sensor data through a three-dimensional model creation module, and uses this information for path planning and obstacle handling. Based on the 3D model, the system further processes obstacle information and hands it over to the obstacle processing module to optimize obstacle avoidance decisions. The training and testing module is the reinforcement learning core of the system, which continuously trains and tests the drone to optimize its decision-making ability in complex dynamic environments. Based on the above modules, the 3D path planning module generates the optimal path and makes real-time adjustments according to changes in the environment. All data of the system, including sensor information, 3D models of the environment, and parameters generated during the learning process, will be stored in the data storage module for subsequent use and analysis. The entire system achieves autonomous obstacle avoidance and path planning for drones in three-dimensional space through the collaboration of multiple modules.

3.2 3D Model Creation

To achieve Q-learning based 3D path planning for drones, the system first needs to construct a 3D gridded environment model to represent the feasible space and obstacle positions during drone flight. Specifically, the entire flight environment is discretized into a three-dimensional grid of M * M * M, where each grid represents a spatial unit. This gridded model can effectively simulate the three-dimensional environment in reality, transforming complex continuous spaces into discrete state spaces, which facilitates path planning using reinforcement learning algorithms [12].

In each grid, the system assigns a value to the spatial unit through an environmental reward function. This return function takes into account the flight targets, obstacle distribution, and environmental factors during the flight of the drone. The return value of the grid reflects the passability and flight cost of the spatial unit, for example, paths that avoid obstacles may have higher positive returns, while paths that approach obstacles or pose risks may have negative returns [13]. In this way, the system can provide reward feedback for drones in different states, helping them find the optimal flight path through learning. When creating a 3D model, the system needs to update environmental information in real-time. The

ISSN 2616-5767 Vol.7, Issue 6: 155-161, DOI: 10.25236/AJETS.2024.070622

dynamic environmental data input from external sensors will continuously adjust the gridded model to ensure that the drone can still effectively avoid obstacles and plan paths when facing complex and changing environments.

3.3 Implementation of 3D Path Planning for Drones

In order to effectively plan the path of the drone in a three-dimensional environment, key parameters as shown in Table 1 have been set:

Table 1: Training parameter settings

Parameter	Value	Unit
Movement Direction i	1-26	-
Greedy Coefficient G	0.2	-
Reward Coefficient R	0.8	-
Training Iterations	3000	times
Maximum Movement Steps	80	steps
Map Side Length M	5 (8)	grid
Number of Maps	4 (10)	-

Q-table is a table used during the training process to store expected rewards for executing different actions in various states. For each fixed map, the state of the drone can be represented by a three-dimensional vector (x, y, z) to indicate its position. In a three-dimensional environment, there are 26 selectable actions for each state, so Q-table is a four-dimensional array represented as Q [x] [y] [z] [i], where i represents the action selected at a certain moment. Through learning, the values in the Q-table will continuously update, enabling the drone to gradually optimize its path selection. The core of path optimization for drones is to update the Q-table using the Q-learning algorithm. The updated formula is:

$$Q[x][y][z][i] = Q[x][y][z][i] + \alpha(r[x_1][y_1][z_1] + \gamma \max(Q[x_1][y_1][z_1][i]) - Q[x][y][z][i]) \quad (1)$$

In this formula, after executing action i in the current state (x, y, z), the new state obtained is (x_1, y_1, z_1). At this point, the reward for the drone is updated as: the reward obtained in the current state plus the expected return of the best future action. The α in the formula is the learning rate used to adjust the pace of learning, where α=0.8. In addition, to prevent the drone from falling into local optima during training, the greedy coefficient G is set to 0.2. Under the greedy strategy, the drone has an 80% probability of choosing the current optimal behavior, but also retains a 20% probability of exploring and trying new paths. This balance ensures that the drone can quickly converge to the optimal solution while avoiding local optima. The goal of path planning is to find the optimal path that satisfies the flight constraints. The construction of the return function takes into account various influencing factors in the environment, such as wind force, terrain, temperature, lighting, etc. The expression form of the return function is:

$$\text{Reward}=w_1 \times f_1 + w_2 \times f_2 + w_3 \times f_3 \quad (2)$$

Among them, w_1 to w_3 are weighting coefficients, and the sum is 0.5, used to balance the impact of different environmental factors on returns. The environmental parameters f_1 to f_3 are used to quantify the quality of each environmental parameter, with values ranging from 1 (indicating good environment) to -1 (indicating poor environment). During the training process, whenever the drone passes through a grid, the system calculates a score based on the environmental factors and reward function of the grid as a reward for that location. The drone accumulates these rewards, continuously adjusts its path, and ultimately finds the optimal path from the starting point to the endpoint. After 3000 rounds of training, an accurate Q-table was finally formed, enabling the drone to find the optimal route in actual testing.

4. Results and Discussion

4.1 Comparison of the Effectiveness of Multi Obstacle Path Planning

The experiment tested the path planning effect of Q-learning algorithm on unmanned aerial vehicles in multi obstacle maps. After 3000 training sessions, the drone is able to autonomously plan obstacle avoidance paths in complex environments, effectively avoid obstacles, and find the optimal route. Figure 2 shows the Matlab simulation results of path planning for a drone in a multi obstacle map, where the drone can reach the target position in a short number of steps.

ISSN 2616-5767 Vol.7, Issue 6: 155-161, DOI: 10.25236/AJETS.2024.070622

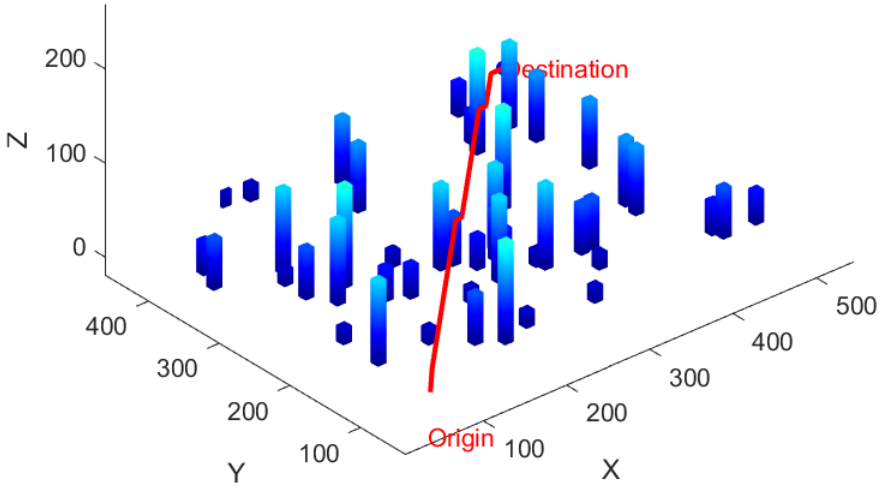

Figure 2: Path planning results of unmanned aerial vehicles in complex and multi obstacle environments

4.2 Efficiency of 3D Path Planning

Table 2 shows the average efficiency comparison of ant colony optimization algorithm (ACO), A* algorithm, fast extended random tree (RRT), and Q-learning in three-dimensional path planning in complex multi obstacle environments.

Table 2: Comparison of efficiency in 3D path planning

Metric	ACO	A*	RRT	Q-learning
Search Time (s)	28.5576	1.1181	0.0541	0.0434
Search Return Rate	0.6373	0.7791	0.7826	0.7231
Maximum Turning Angle (degrees)	109.4712	180	96.7966	90.1241
Number of Turns Exceeding 45 Degrees	57.2452	346.1241	11.1414	5.6546

In terms of search time, Q-learning performs the best, with an average search time of 0.0434 seconds, slightly better than RRT's 0.0541 seconds. ACO has the longest search time, reaching 28.5576 seconds, significantly lagging behind other algorithms. The search time of the A * algorithm is 1.1181 seconds, which is moderate and reflects the balance of its efficiency. In terms of search return rates, the A* algorithm and RRT perform the most reliably, with values of 0.7791 and 0.7826, respectively. Although Q-learning has high efficiency, its return rate is slightly lower at 0.7231, while ACO has the lowest return rate at only 0.6373. In terms of path smoothness, the maximum turning angle of Q-learning is only 90.1241°, and the number of turns exceeding 45° is the least, only 5.6546 times, indicating that its path is smoother and more efficient. RRT also performs well, with a maximum turning angle of 96.7966° and 11.1414 turns exceeding 45 degrees. The path generated by the A* algorithm has a large turning angle, with a maximum turning angle of 180 °, and the number of turns exceeding 45 degrees is as high as 346.1241 times, indicating its limitations in generating smooth paths. ACO also performs poorly, with a maximum turning angle of 109.4712 ° and 57.2452 turns exceeding 45 degrees.

4.3 Comparison of Path Length

The experimental results also compared the path length performance of ant colony algorithm, A* algorithm, RRT algorithm, and Q-learning algorithm in multi obstacle environments and dynamic wind speed environments, and calculated the path optimization percentage of Q-learning algorithm relative to

ISSN 2616-5767 Vol.7, Issue 6: 155-161, DOI: 10.25236/AJETS.2024.070622

other algorithms. The specific data is shown in Figure 3.

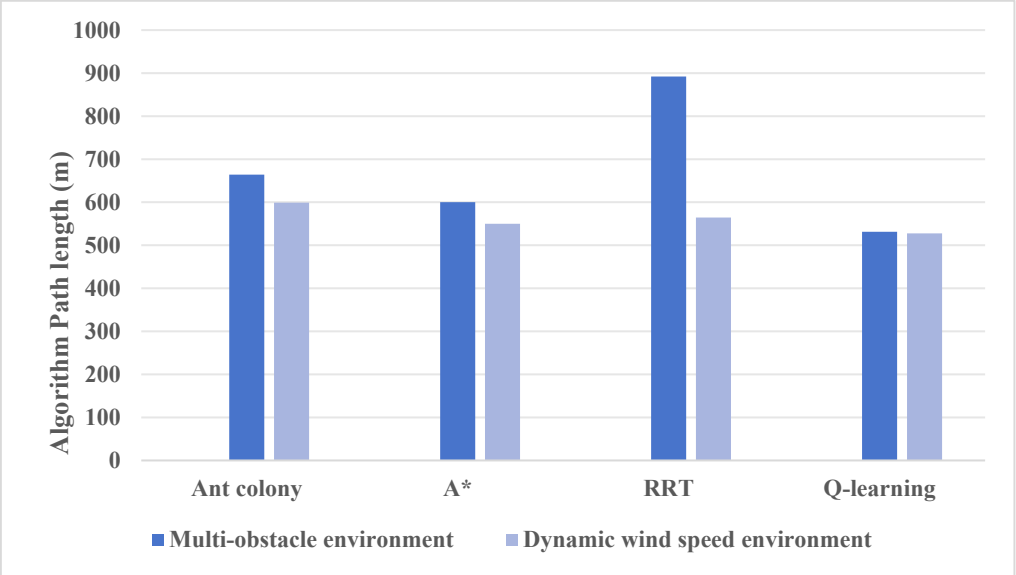

Figure 3: Comparison of algorithm path length in different testing environments

In a multi obstacle environment, the Q-learning algorithm achieves the shortest path length (531.4562 meters), which is about 20.00% shorter than the ant colony algorithm (664.2807 meters); compared to the A* algorithm (600.2351 meters), it has shortened by about 11.45%; compared to the RTT algorithm (892.0390 meters), it has shortened by about 40.39%. These data indicate that the Q-learning algorithm has significant path optimization capabilities in complex obstacle environments.

In a dynamic wind speed environment, the Q-learning algorithm has a path length of 527.5467 meters, which is approximately 11.95% shorter than the ant colony algorithm (599.1645 meters); compared to the A* algorithm (549.8595 meters), it has shortened by about 4.06%; compared to the RTT algorithm (564.5649 meters), it has shortened by about 6.56%. This further indicates that the Q-learning algorithm not only performs well in complex environments, significantly reducing path length and improving the flight efficiency of drones.

5. Conclusion

This paper presents a three-dimensional path planning system for unmanned aerial vehicles based on Q-learning reinforcement learning algorithm. By constructing a three-dimensional environment model and using Q-learning algorithm, the system can continuously optimize path decisions by maximizing cumulative rewards. Under extensive training, drones can effectively avoid obstacles and find the optimal path. The system uses a greedy strategy to balance exploration and utilization, avoiding the problem of local optimal solutions. The experimental results show that after about 2000 training sessions, the system's path planning tends to be stable and can achieve autonomous flight in complex environments. Through the evaluation of path smoothness and obstacle avoidance ability, the optimized path reduces unnecessary turns and redundant nodes, significantly improving flight efficiency. Compared with traditional algorithms, Q-learning performs significantly in path optimization. In multi obstacle environments, the path length is reduced by 20.00% compared to ant colony algorithm, 11.45% compared to A* algorithm, and 40.39% compared to RRT algorithm. In addition, under dynamic wind speed conditions, Q-learning reduces the path length by 4% to 12% compared to other algorithms. These results indicate that the Q-learning algorithm has significant advantages in handling complex tasks and multiple times, and is particularly suitable for path planning and obstacle avoidance systems of unmanned aerial vehicles. Overall, the Q-learning algorithm has demonstrated strong potential in 3D path planning for drones, providing an effective solution for their autonomy in unknown and complex environments. Future research can further integrate deep reinforcement learning techniques to address more complex and unpredictable environments and enhance the real-time decision-making capabilities of drones.

ISSN 2616-5767 Vol.7, Issue 6: 155-161, DOI: 10.25236/AJETS.2024.070622

References

[1] Ramasamy M, Ghose D. Learning-based preferential surveillance algorithm for persistent surveillance by unmanned aerial vehicles[J].IEEE, 2016.DOI:10.1109/ICUAS.2016.7502678.
[2] Mandloi D, Arya R, Verma A K. Unmanned aerial vehicle path planning based on A algorithm and its variants in 3d environment [J]. International Journal of System Assurance Engineering and Management, 2021, 12(5): 990-1000.*
[3] Kiani F, Seyyedabbasi A, Aliyev R, et al. 3D path planning method for multi-UAVs inspired by grey wolf algorithms[J]. Journal of Internet Technology, 2021, 22(4): 743-755.
[4] Lv J X, Yan L J, Chu S C, et al. A new hybrid algorithm based on golden eagle optimizer and grey wolf optimizer for 3D path planning of multiple UAVs in power inspection[J]. Neural Computing and Applications, 2022, 34(14): 11911-11936.
[5] Maboudi M, Homaei M R, Song S, et al. A Review on Viewpoints and Path Planning for UAV-Based 3-D Reconstruction [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023, 16: 5026-5048.
[6] Zammit C, van Kampen E J. Real-time 3D UAV path planning in dynamic environments with uncertainty [J]. Unmanned Systems, 2023, 11(03): 203-219.
[7] Wang C, Yang X, Li H. Improved Q-learning applied to dynamic obstacle avoidance and path planning [J]. IEEE Access, 2022, 10: 92879-92888.
[8] Shang Y, Liu F, Qin P, et al. Research on path planning of autonomous vehicle based on RRT algorithm of Q-learning and obstacle distribution[J]. Engineering Computations, 2023, 40(5): 1266-1286.
[9] Bonny T, Kashkash M. Highly optimized Q-learning-based bees approach for mobile robot path planning in static and dynamic environments[J]. Journal of Field Robotics, 2022, 39(4): 317-334.
[10] Lyu J, Ma X, Li X, et al. Mildly conservative q-learning for offline reinforcement learning[J]. Advances in Neural Information Processing Systems, 2022, 35(1): 1711-1724.
[11] Tran C H, Bui T K, Pham T V. Virtual machine migration policy for multi-tier application in cloud computing based on Q-learning algorithm[J]. Computing, 2022, 104(6): 1285-1306.
[12] Chintala P, Dornberger R, Hanne T. Robotic path planning by Q learning and a performance comparison with classical path finding algorithms[J]. International Journal of Mechanical Engineering and Robotics Research, 2022, 11(6): 373-378.
[13] Hu J. A novel deep learning driven robot path planning strategy: Q-learning approach[J]. International Journal of Computer Applications in Technology, 2023, 71(3): 237-243.