ISBN: 978-981-94-4198-3
2025 the 15th International Workshop on Computer Science and Engineering (WCSE 2025)
doi: 10.18178/wcse.2025.06.039

A Terminal Airspace Flight Path Planning Algorithm Based on Improved Deep Q-Network

Han Yingchao [1,2], Wei Qi [1,3,] Shen Zhiyuan [1+]

[1] College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

[2] Civil Aviation Administration of China Northwest Regional Air Traffic Management Bureau at Ningxia Subbranch, Yinchuan 750000,China

[3] China Eastern Airlines Jiangsu Co.Ltd, Nanjing 211106, China

Abstract. Due to the deteriorated weather and the activities of other airspace users, the control job at terminal airspace become complex and changeable. This situation brings more work burden for the air traffic controllers to conduct aircraft path planning. The conventional aircraft path planning methods are not consistent with the actual operation situation of air traffic control facility. In this paper, a terminal airspace flight path planning algorithm based on improved Deep Q-Network (DQN) is proposed. It use the results generated by classical path planning algorithms as priority experience to feedback the DQN model. The experimental results show the proposed algorithm is able to reduced the distance of the path planning and training time compared with the traditional path planning algorithm. It greatly improves the model training efficiency.

Keywords: Flight path planning, Deep Q-Network, Terminal airspace, restricted airspace.

1. Introduction

Balancing the safety and efficiency of air traffic control has always been a challenge. The terminal control area is the busiest and most complex area in air transportation. During diverse air traffic activities or complex weather conditions, the airspace becomes complex and unstable, affecting the operational efficiency of civil aviation flights. Therefore, it is necessary to study methods of aircraft path planning in the context of the complex and ever-changing airspace of the terminal control area.

Initially, it was mainly used geometric algorithms to plan the detour routes of aircraft. Simple network graphs[1], grid-based Dijkstra's algorithm[2], segmenting the planning of aircraft paths[3] are several methods used earlier for flight path planning. In solving the single aircraft path planning problem based on grid maps, the A* algorithm[4-6] has shown high efficiency and accuracy. Other algorithms, such as greedy algorithms[7], Markov processes and dynamic programming algorithms[8], nonlinear predictive model[9], genetic algorithms[10], simulated annealing algorithm[11], aim to consider other constraints in path planning problems simultaneously.Deep reinforcement learning methods have also used to solve path planning problems. Deep Q-network reinforcement learning[12], dueling Q-networks algorithm[13], improved the experience replay mechanism DQN[14] and Independent Deep Q-Network algorithm[15] commonly used to solve path planning problems.

This paper proposes an improved deep reinforcement learning algorithm aimed at solving the problem of aircraft path planning in terminal control areas. The article first describes the actual operational conditions of the terminal control area and uses a grid map to represent the operational environment of the terminal area and restricted airspace. Then, this paper introduces an improved prioritized experience replay mechanism for the DQN algorithm, which takes the results of traditional path planning as prioritized experiences for learning. Finally, the article compares and analyzes the results of the proposed improved DQN algorithm with those of the A* algorithm and the traditional DQN algorithm.

[+] Corresponding author. Tel.: + 13951916587; fax: +025-52119075.
E-mail address: shenzy@nuaa.edu.cn

The rest of this paper is organized as follows: Section 2 introduces actual operational conditions of the terminal control area and grid map of operational environment. Section 3 introduces an improved prioritized experience replay mechanism for the DQN algorithm. Section 4 compares and analyzes the results of the proposed improved DQN algorithm with the traditional path planning algorithm. Finally,the conclusion was given in Section 5.

2. The Operation of Terminal Control Airspace

2.1. Problem Description

In Terminal control airspace, air traffic is relatively complex. Factors that significantly affect the available airspace operations in terminal control areas mainly include: hazardous convective weather affecting aircraft operations, activities of other airspace users, general aviation flight activities, and the activities of unmanned aerial vehicles and other unknown airborne objects. These factors may pose a threat to flight safety, and even lead to danger.

If aircraft encountering dangerous weather such as thunderstorms, turbulence and wind shear, it will typically take measures to re-plan their routes to avoid. When other airspace users are active, civil aviation flights are prohibited from entering a specific airspace, civil aircraft must re-plan their routes. Within the terminal control area, when general aviation flights and UAV flight activities are conducted simultaneously, a dedicated segregated airspace must be established to ensure that civil aviation activities maintain an appropriate distance from them, thereby ensuring the safety of all types of flights.

2.2. Environmental Modeling

Converting terminal control area airspace into a grid map is a common method for aircraft path planning. Since the main object of study in this article is the arriving aircraft in the terminal control area, only the horizontal path planning is considered during path planning. Using a grid map, the complex terminal control area airspace environment can be simplified into a two-dimensional space.

Acquiring hazardous weather information mainly relies on weather radar. Under normal circumstances, when the intensity of weather radar echoes exceeds 30 dBz, it is considered that this may pose a significant threat to the flight safety of aircraft.First, we process the weather radar images by removing areas with an intensity below 30dBz. Then, we divide the images into grids of a predetermined size and treat those areas that are not completely covered by a grid as full covered.

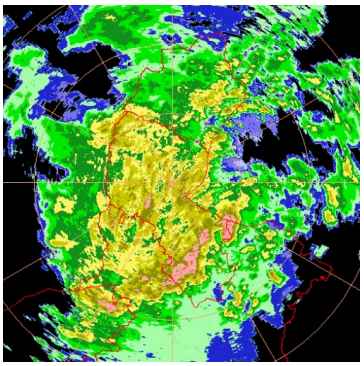

Fig. 1: Weather radar echo image.

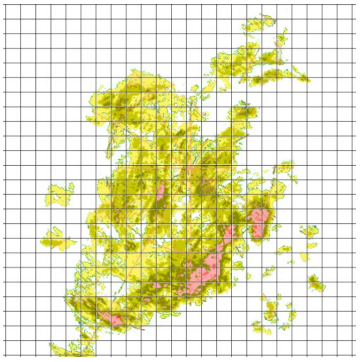

Fig. 2: Processed weather radar echo image.

Restricted airspace caused by other airspace user activities, general aviation and UAV activities is transformed into a grid according to the scope of restricted airspace, and the area that is not completely covered by one space is marked as occupying one space.

3. Flight Path Planning Algorithm Based on Improved DQN

3.1. DQN Algorithm Principle

In reinforcement learning, the Q-learning method is a tabular approach. The Q-table is a table of state-action pairs, where the data represents the expected total reward for taking a certain action in a particular state, which is the state-action value table.Q-learning is a model-free value-based reinforcement learning

algorithm that uses Q-values to determine the actions of an agent. The core of the algorithm is to store all the corresponding Q-values for states and actions in a Q-value table during the learning process, which are continuously updated as the agent explores and learns.

Deep Q-Network (DQN) refers to a Q-learning algorithm based on deep learning, which mainly combines value function approximation with neural network technology. It uses a representation method of value function approximation and employs a deep neural network to fit the Q function. Additionally, it adopts techniques such as target networks and experience replay for network training, effectively enhancing the efficiency and effectiveness of reinforcement learning.

$$Q(s,a;\theta) \approx Q^*(s,a) \tag{1}$$

Use a neural network with parameters to approximate the Q-values. Deep Q-Network (DQN) algorithm includes two neural networks, one for representing the target Q-network, and the other for generating the evaluation Q-network. Use the target network to compute the target Q-values:

$$y_t = r_t + \gamma \max_{a_j} Q'(s_{j+1}, a_j) \tag{2}$$

In a Deep Q‐network or DQN,by calculating the difference between the predicted value and the target value and defining it as the loss function.

$$L(\theta) = E\left[(y_i - Q(s,a;\theta))^2\right] \tag{3}$$

We use gradient descent to update weights and minimize the loss. After a certain number of updates to the network, it is then used to update the target network. The parameters of the target network are not directly updated during the training process, but are periodically copied from the Q-network to maintain stability and consistency in the target network.

3.2. Improve the Experience Feedback Mechanism

In Deep Q-Network (DQN) reinforcement learning, the agent interacts with the environment and stores the obtained data in a replay buffer. During the training phase, the agent randomly samples a batch of data from the replay buffer to update the Q-network. During the data extraction process, random sampling is typically used to select samples.

Random sampling mechanisms may lead to less common data not being effectively utilized, thus hindering the acceleration of the agent's learning process.In the early stages of training, the agent must perform a large number of random actions to explore viable strategies, which often leads to a significant decrease in training efficiency, as many actions do not yield effective rewards.

Path planning problems can now be solved with feasible solutions using traditional algorithms, although these solutions may not be optimal. If the results obtained from traditional algorithms are used as the initial training experience for agents, it will significantly enhance the efficiency of the initial training.

For the path planning problem presented in this article, we first use the A* algorithm to find a path for the agent to reach the destination, and then convert these paths into data for the agent's interaction with the environment, storing them in the replay buffer. And prioritize these data as priority experience.

The prior knowledge is set as follows:

(1) Assuming the path obtained through the A* algorithm is $s = [s_1, s_2, s_3 \cdots, s_n]$,Each of s_i represents the position of an agent on the grid map.

(2) The actions of an agent can be calculated using the relationship between the current position and the next position. $s_i(x_i, y_i)$ is a point in the path queue, and $s_{i+1}(x_{i+1}, y_{i+1})$ is the next point in the path queue, we calculate the difference between two coordinate points, and then convert the positional relationship of the two points into corresponding actions of the agent.

$$(x_{i+1}-x_i, y_{i+1}-y_i) == \begin{cases} (0,-1) & a=0 \\ (-1,-1) & a=1 \\ (-1,0) & a=2 \\ (-1,1) & a=3 \\ (0,1) & a=4 \end{cases} \tag{4}$$

(3) Interact with the environment using the agent's position and action information as states and actions, calculate the obtained rewards and the agent's next state and expressed as $[s,a,r,s_{next}]$, and store the data into the replay buffer.

4. Experiment and Analysis

4.1. Experiment Setting

We converted the terminal control area airspace map within a 100-kilometer range into a grid map and marked the hazardous weather areas and restricted airspace on the grid map. The airspace grid map is shown in Figure 3.

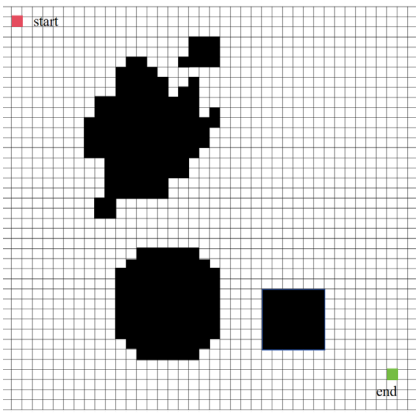

Fig. 3: The initial airspace environment.

The maximum turning angle of the agent is assumed to equal $\pi/4$.The state of the agent is represented by (x,y). The agent has five actions, each represented by $a=[0,1,2,3,4,5]$.

The reward obtained by the agent interacting with the environment is defined as follows:

(1) The agent reaches the end point and receives a reward of 100.

(2) The agent enters a restricted airspace, or goes out of the map, and receives a penalty of -100.

(3) The agent receives a penalty of -1 for each step taken.

The hyperparameters of the DQN algorithm are set as in table 1.

Table 1: Network hyperparameters

Hyperparameter	Value
Learning rate	0.0025
Experience pool size	1000
Batch size	64
Number of Episodes	2000
Discount Rate	0.99
Epsilon	1
Epsilon – Decay	0.995
Epsilon – Min	0.1

The Q-network updates the parameters at every time step; the update frequency of the Q-target network is 50 time steps.

4.2. Results and Analysis

To verify the effectiveness and practicality of the algorithm in this paper, we use the A* algorithm, the classic DQN algorithm, and the improved DQN algorithm in this paper to solve the path planning problem.

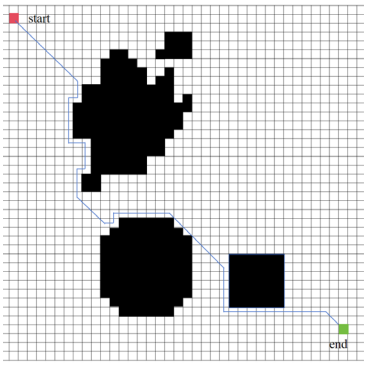

Fig. 4: A* algorithm path planning result.

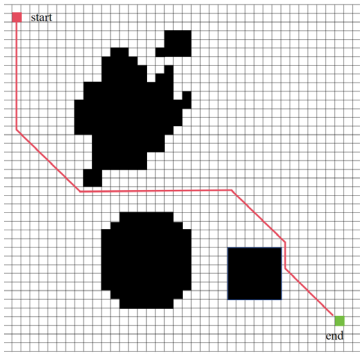

Fig. 5: Improved DQN algorithm path planning result.

Table 2: Path Planning Results Comparison

Algorithm	Turning point	Path length
A* algorithm	15	66.452
Improved DQN algorithm	5	59.866

Figure 4 and Figure 5 shows the paths obtained using the A* algorithm and the improved DQN algorithm, Table 2 compares the results of the two algorithms. The comparison of results obtained from two algorithms indicates that the path generated by the method proposed in this paper has fewer turning points and a shorter path length compared to the A* algorithm.

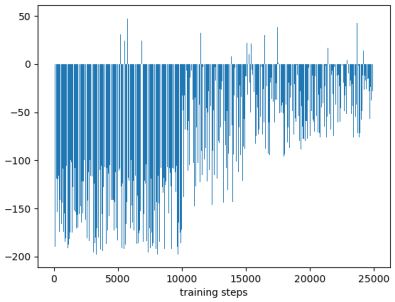

Fig. 6: Total reward of classic DQN algorithm.

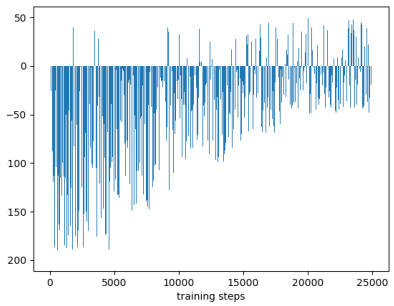

Fig. 7: Total reward of improved DQN algorithm.

To evaluate the convergence speed of the algorithm, Figure 6 and Figure 7 shows the number of training steps taken for the total reward value to converge during training for the improved DQN algorithm compared to the classic DQN algorithm. By comparing the total rewards obtained by the two algorithms, we can conclude that the improved DQN algorithm is able to learn successful strategies more quickly and also achieves a higher success rate in reaching the goal during subsequent training.

5. Conclusion

This article mainly analyzes and discusses the issue of aircraft path planning in complex terminal control areas. We describe the operational conditions of aircraft in terminal control areas. Analyzes several factors that limit the airspace in terminal control areas. The complex airspace environment of terminal control areas is represented using grid maps. The path planning results of the A* algorithm are used as priority experience replay to improve the Deep Q-Network (DQN) algorithm. Comparative experiments between the improved DQN algorithm and the traditional DQN algorithm show that the improved algorithm significantly enhances the training efficiency of the model.

6. Acknowledgements

This research has been financially supported by Northwest Regional Air Traffic Management Bureau at Ningxia subbranch.

7. References

[1] Y. Chiang, J. T. Klosowski, C. Lee, and J. S. B. Mitchell, "Geometric algorithms for conflict detection/resolution in air traffic management," *IEEE*, vol. 2, pp. 1835-1840, 1997.

[2] Krozel, Jimmy, et al. "Terminal area guidance incorporating heavy weather." *Guidance, Navigation, and Control Conference*. 1997.

[3] Sridhar, Banavar, et al. "Integration of traffic flow management decisions." *AIAA Guidance, Navigation, and Control Conference and Exhibit*. 2002.

[4] Gianazza, David, Nicolas Durand, and Nicolas Archambault. "Allocating 3D-trajectories to air traffic flows using A* and genetic algorithms." CIMCA 2004, *international conference on Computational Intelligence for Modelling, Control and Automation*. 2004.

[5] Prete, Joseph, and Joseph Mitchell. "Safe routing of multiple aircraft flows in the presence of time-varying weather data." AIAA *Guidance, Navigation, and Control Conference and Exhibit*. 2004, pp. 4791.

[6] Bojorquez, Oliver, Nathan Dolan, and Jun Chen. "Aircraft rerouting under risk tolerance during space launches." AIAA *Scitech 2020 Forum*. 2020, pp. 1589.

[7] Krozel, Jimmy, et al. "Comparison of algorithms for synthesizing weather avoidance routes in transition airspace." AIAA *Guidance, Navigation, and Control Conference and Exhibit*. 2004, pp. 4790.

[8] d'Aspremont, Alexandre, et al. "Optimal path planning for air traffic flow management under stochastic weather and capacity constraints." 2006 *International Conference onResearch, Innovation and Vision for the Future. IEEE*, 2006, pp. 1-6.

[9] Pannequin, Jessica, et al. "Multiple aircraft deconflicted path planning with weather avoidance constraints." AIAA *Guidance, Navigation and Control Conference and Exhibit*. 2007, pp. 6588.

[10] Ayo, Babatope S. Data-driven flight path rerouting during adverse weather: Design and development of a passenger-centric model and framework for alternative flight path generation using nature inspired techniques. Diss. University of Bradford, 2019.

[11] Taylor, Christine, and Craig Wanke. "Dynamically generating operationally acceptable route alternatives using simulated annealing." *Air Traffic Control Quarterly*, 2012, 20(1): 97-121.

[12] Lv, Liangheng, et al. "Path planning via an improved DQN-based learning policy." *IEEE Access.* 2019, pp. 67319-67330.

[13] Rybak, L. A., et al. "Development of an algorithm for managing a multi-robot system for cargo transportation based on reinforcement learning in a virtual environment." *IOP Conference Series: Materials Science and Engineering,* 2020, 945(1): 012083.

[14] Yang, Yang, Li Juntao, and Peng Lingling. "Multi‐robot path planning based on a deep reinforcement learning DQN algorithm." *CAAI Transactions on Intelligence Technology*, 2020, 5(3): 177-183.

[15] Dong, S. U. I., X. U. Wei**, and Kai Zhang. "Study on the resolution of multi-aircraft flight conflicts based on an IDQN." *Chinese Journal of Aeronautics*, 2022, 35(2): 195-213.