ISSN: 2414-1895 DOI: 10.6919/ICJE.202602_12(2).0011

Global Path Planning for Unmanned Surface Vehicles based on Adaptive Parameter Q-Learning Algorithm

Haoran Wang, Shaoyi Guo

School of Navigation and Shipping, Shandong Jiaotong University, Weihai 264200, China

Abstract

Global path planning is one of the core issues in the field of Unmanned Surface Vehicle (USV) navigation. As a classic method in reinforcement learning, the Q-Learning algorithm is widely applied to USV global path planning. However, the traditional Q-Learning algorithm typically employs a fixed learning rate, greedy rate, and discount factor, which leads to issues such as redundant exploration in the late training stage and slow convergence speed. To address these limitations, this paper proposes an adaptive parameter Q-Learning path planning algorithm. The proposed algorithm dynamically adjusts the greedy rate to balance exploration and exploitation, adaptively optimizes the learning rate based on the temporal difference (TD) error, and dynamically adjusts the discount factor by combining the distance between the current state and the destination. These improvements enhance both the convergence efficiency and path planning performance of the algorithm. Comparative experiments between the improved algorithm and the traditional Q-Learning algorithm were conducted in a simulated water environment. Experimental results demonstrate that the improved algorithm increases the training convergence speed by over 30% and shortens the optimal path length by 11%, indicating a significant improvement in algorithm performance. This study provides a more efficient algorithmic scheme for USV global path planning and holds great significance for enhancing the real-time performance, reliability, and environmental adaptability of USV navigation.

Keywords

Unmanned Surface Vehicle (USV); Global Path Planning; Q-Learning Algorithm; Adaptive Parameters; Reinforcement Learning.

1. Introduction

With the rapid development of artificial intelligence technology, the extensive application of Unmanned Surface Vehicles (USVs) in fields such as ocean exploration, maritime rescue, and port operations has created an increasingly urgent demand for global path planning technologies. Path planning refers to the process of generating a navigation route from a departure port to a destination port within a mission environment containing obstacles. Whether executed before or during the voyage, it must account for surrounding meteorological factors to ensure the path safely and reliably avoids all obstacles while meeting navigation requirements [1]. Prior to departure, intelligent ships can plan an optimal route-maximizing economy, reliability, and safety-by analyzing meteorological conditions, fuel consumption, and other factors potentially affecting navigation safety along the intended course [2].Currently, mainstream ship path planning algorithms include Dijkstra [3], the A* algorithm [4], Genetic Algorithm (GA) [5], Ant Colony Optimization (ACO) [6], and Reinforcement Learning (RL) [7]. Lee H. T. et al. [8] applied the Q-Learning algorithm to ship path planning and incorporated the Douglas–Peucker algorithm to eliminate redundant waypoints, thereby shortening the navigation distance. Chen [9] proposed an improved Q-Learning path planning method that

ISSN: 2414-1895 DOI: 10.6919/ICJE.202602_12(2).0011

standardized the distances to obstacles and restricted areas relative to the rewards or penalties used for determining ship maneuvers. Shen Haiqing [10] constructed a multi-layer collision avoidance method for USVs based on Deep Q-Network (DQN). This method integrates collision avoidance regulations with evasion experience and has been successfully applied to multi-ship scenarios in complex environments, achieving autonomous collision avoidance navigation. Yoo et al. [11] proposed a Q-Learning-based method to generate optimized routes for USVs subject to environmental disturbances such as sea winds and currents. This method not only accounts for the USV's non-holonomic motion constraints but also introduces a path smoothing algorithm to significantly improve the smoothness of the generated trajectory. Wang Yinan [12] integrated the Dyna framework with the Q-Learning algorithm to propose a ship collision avoidance decision-making method. This approach significantly enhances the efficiency of value function updates and has been successfully applied to collision avoidance decisions in ship encounter situations, ultimately generating effective and safe avoidance paths. Wang Yuanhui [13] proposed an improved Neural Smoothed Fast Q-Learning (NSFQ) algorithm and employed third-order Bézier curves to smooth the initial path, which successfully reduced the path length in static obstacle environments. Chen Xinqiang [14] combined the global path planning capability of the A* algorithm with the adaptive learning mechanism of the Double Deep Q-Network (DDQN) algorithm to develop an A*-guided DDQN (A-DDQN) route planning method. A novel reward mechanism was also introduced to assist ships in identifying the optimal action strategy during route optimization.The Q-Learning algorithm in reinforcement learning does not rely on prior environmental knowledge; instead, it learns the optimal policy through continuous interaction between the agent and the environment, offering an effective solution for USV global path planning. However, the fixed parameters of the traditional Q-Learning algorithm struggle to adapt to the dynamic changes between environmental exploration and policy exploitation during training, thereby limiting algorithmic performance. Consequently, performing adaptive parameter optimization on the Q-Learning algorithm and applying it to USV global path planning holds significant theoretical value and practical importance.

2. Improved Q-Learning Algorithm

2.1 Q-Learning Algorithm

The principle of USV path planning using the Q-Learning algorithm is based on a continuous bidirectional interaction mechanism. The system state is synchronously transmitted to the USV and the environment. Based on this state, the USV executes a corresponding action upon the environment. Simultaneously, the environment conveys the interaction feedback from the previous time step to the USV. Upon receiving the current action, the environment updates its status to generate a new state for the next time step. This process iterates cyclically, constituting a continuous interaction loop, as illustrated in Figure 1.

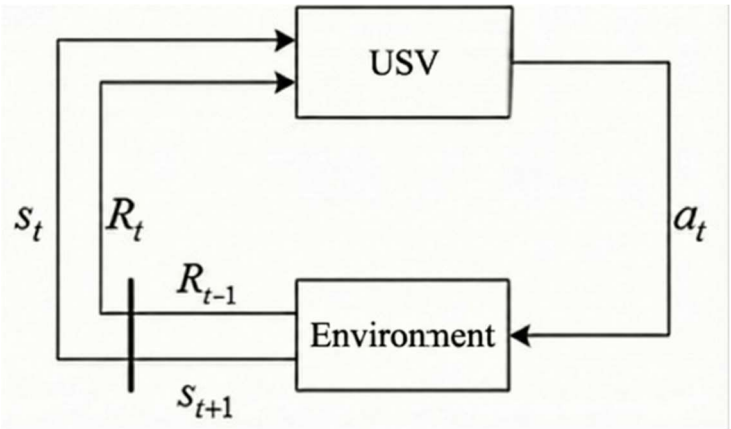

Figure 1. Principles of Reinforcement Learning

ISSN: 2414-1895 DOI: 10.6919/ICJE.202602_12(2).0011

The Q-table is the core data structure of the Q-Learning algorithm and is essentially a two-dimensional matrix. The rows of the matrix correspond to all possible states of the agent, while the columns correspond to all executable actions available in each state. Each cell in the matrix is used to store the action value for its corresponding state-action pair. The Q-table matrix is shown in Table 1.

Table 1. Three Scheme comparing

$Q-Table$	α_1	α_2	α_3
s_1	$Q(s_1,\alpha_1)$	$Q(s_1,\alpha_2)$	$Q(s_1,\alpha_3)$
s_2	$Q(s_2,\alpha_1)$	$Q(s_2,\alpha_2)$	$Q(s_2,\alpha_3)$
s_3	$Q(s_3,\alpha_1)$	$Q(s_3,\alpha_2)$	$Q(s_3,\alpha_3)$
s_4	$Q(s_4,\alpha_1)$	$Q(s_4,\alpha_2)$	$Q(s_4,\alpha_3)$

The Q-learning algorithm operates by constructing a Q-table to store the value of every state-action pair. As the agent interacts with the environment, it selects actions based on this Q-table and iteratively updates the Q-values using reward signals, ultimately forming an optimal action policy. The core formula for updating the Q-table is as follows:

$$Q(s,a)^{new} = Q(s,a)^{old} + \alpha\left[r + \gamma max_{a'}Q(s',a') - Q(s,a)^{old}\right] \quad (1)$$

Here, represents the current state-action value, denotes the learning rate, and is the immediate reward actually received. serves as the discount factor (or learning factor) used to balance the weights of immediate versus future rewards, and represents the optimal Q-value for the next state. However, the performance of the traditional Q-learning algorithm is limited by its fixed parameters, which often leads to insufficient exploration in the early stages or excessive exploration in the later stages. A fixed learning rate cannot adapt to the Q-value update requirements across different iteration phases: the update step size is insufficient when errors are large initially, and becomes redundant when errors decrease later on. Furthermore, a fixed discount factor makes it difficult to balance reward priorities depending on whether the unmanned surface vehicle (USV) is far from or near the target point. Therefore, this paper proposes an adaptive parameter Q-learning algorithm.

2.2 Adaptive Parameter Optimization

Traditional Q-Learning utilizes a fixed learning rate, which results in slow updates during the initial learning stage and instability in later stages. The adaptive learning rate adjustment formula designed in this paper is as follows:

$$\alpha_t(s,a) = \frac{\alpha_0}{1+\beta\cdot N_t(s,a)} \quad (2)$$

Traditional Q-Learning utilizes a fixed learning rate, which results in slow updates during the initial learning stage and instability in later stages. The adaptive learning rate adjustment formula designed in this paper is as follows:

$$\gamma_t = \gamma_{min} + (\gamma_{max} - \gamma_{min}) \cdot e^{-\lambda t} \quad (3)$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202602_12(2).0011

In the formula,γ_t is the decay period constant, and t is the number of training cycles. Furthermore, an improved ε-greedy strategy is adopted, wherein the exploration rate is adaptively adjusted as training progresses:

$$\varepsilon_t = \varepsilon_{min} + (\varepsilon_{max} - \varepsilon_{min}) \cdot e^{-\eta \cdot \frac{t}{T}} \quad (4)$$

Furthermore, an uncertainty-based exploration enhancement mechanism is introduced. When the Q-value of a specific state-action pair changes significantly, its exploration probability is increased:

$$\varepsilon_{boost}(s,a) = \varepsilon_t \cdot \left(1 + \sigma \cdot Var_t(s,a)\right) \quad (5)$$

Where$Var_t(s,a)$denotes the estimated variance of the Q-values,σ is the amplification factor.

3. Simulation Results and Analysis

To verify the feasibility of the improved Q-learning algorithm proposed in this paper, simulation experiments were conducted for research on intelligent ship global obstacle avoidance path planning. A 20x20 grid environment was modeled using the grid method [15], incorporating several simulated static obstacles. The simulations were performed using Python 3.9 on a computer equipped with an i8-14650H 3.6GHz processor, 16GB of memory, and an RTX 5060 graphics card. A comparative experiment between the traditional Q-learning algorithm and the improved Q-learning algorithm is shown in Figure 2, and the performance comparison is presented in Table 2:

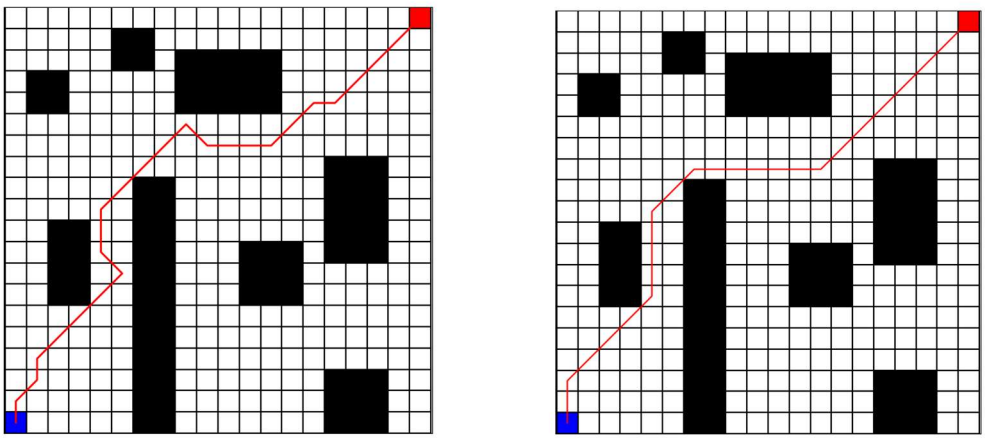

Figure 2. Comparison between traditional and improved Q-learning algorithms.

Table 2. Performance Comparison Before and After Algorithm Improvement

Algorithm	Path Length	Number of Turns	Training Time	Total Path Nodes
Traditional Q-learning	32.04	11	4	26
Adaptive Parameter Q-learning	28.28	5	3.22	15

ISSN: 2414-1895 DOI: 10.6919/ICJE.202602_12(2).0011

Through a quantitative comparative analysis of simulation data between the traditional Q-Learning algorithm and the improved adaptive parameter Q-Learning algorithm, it is evident that the improved algorithm demonstrates significant advantages in core indicators such as path planning accuracy, operational efficiency, and convergence performance. By introducing an adaptive parameter adjustment mechanism-which dynamically optimizes key parameter values according to the training progress-the improved algorithm achieves multi-dimensional performance enhancements. In terms of path optimization, the path length planned by the improved algorithm is optimized by 11% compared to the traditional algorithm, and the total number of path nodes is correspondingly decreased by 11, effectively mitigating path redundancy. Simultaneously, the number of path turns (inflection points) is substantially reduced by 55%, significantly improving the smoothness and continuity of the path. Regarding operational efficiency, the improved algorithm reduces training time by 0.78s compared to the traditional algorithm; this enhances the algorithm's real-time capability while maintaining planning accuracy, thereby meeting the timeliness requirements for path planning in dynamic environments.

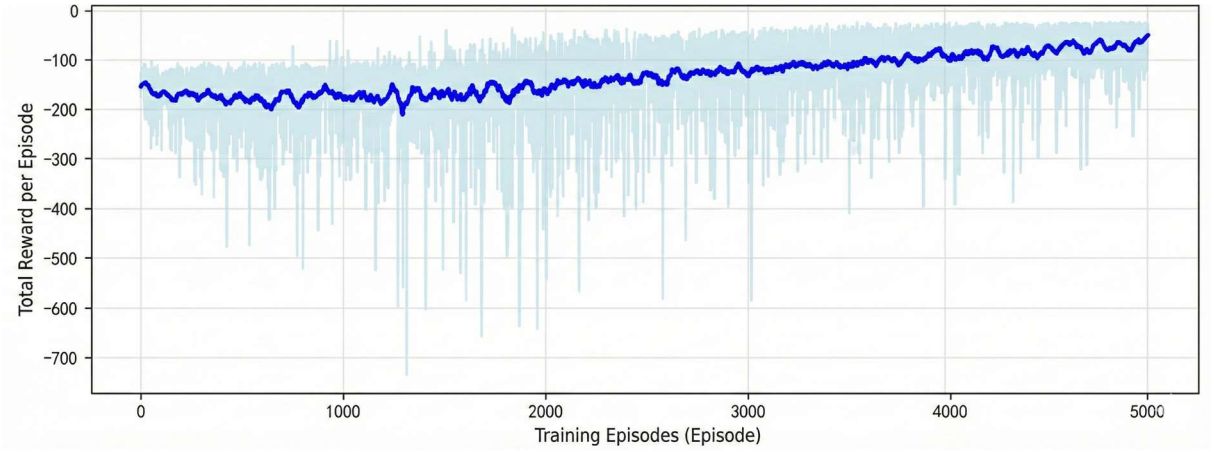

Figure 3. Traditional Q-learning algorithm

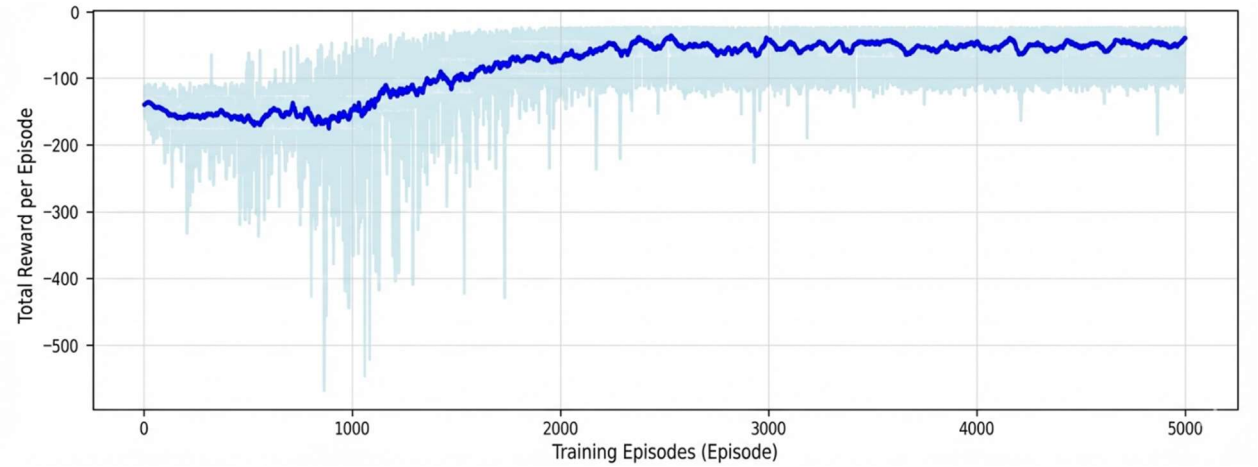

Figure 4. Improved Q-learning algorithm

Experimental results indicate that the training convergence speed of the improved algorithm is increased by over 30%. This improvement is attributed to the dynamic balance between exploration and exploitation achieved by the adaptive parameter mechanism. During the initial training phase, a higher exploration rate allows for a thorough traversal of the state space, while in later stages, the exploitation weight is gradually increased to rapidly converge to the optimal policy. This effectively prevents the issues of slow convergence or entrapment in local optima often caused by fixed

ISSN: 2414-1895 DOI: 10.6919/ICJE.202602_12(2).0011

parameters in traditional algorithms. In summary, the improved adaptive parameter Q-Learning algorithm achieves collaborative optimization of path planning quality, operational efficiency, and convergence stability through its parameter adaptive adjustment mechanism, and its overall performance is significantly superior to that of the traditional Q-Learning algorithm.

4. Conclusion

Aiming at the problem of global path planning for Unmanned Surface Vehicles (USVs), this paper proposes an adaptive parameter Q-Learning algorithm combined with a reward function. By designing a composite reward mechanism that integrates target guidance, safety constraints, navigation quality, and efficiency, and combining it with adaptive learning rates, discount factors, and exploration strategies, the algorithm is enabled to efficiently learn high-quality global paths that are safe, smooth, and energy-efficient. Simulation experiments have validated the overall superiority of this algorithm.Although this paper verified the effectiveness of the improved adaptive parameter Q-Learning algorithm in standard grid environments, limitations such as insufficient scenario adaptability remain when considering the specific characteristics of maritime scenarios for USV global path planning. Combining maritime engineering application requirements with academic research trends, future work can be further expanded and deepened in the following directions to enhance the practicality of the algorithm in actual maritime scenarios:First, strengthen adaptability verification in complex maritime scenarios. The experiments in this paper are based on static grid environments with fixed obstacle distributions, which differ significantly from the actual water environments where USVs operate. Subsequent research should adapt the algorithm to dynamic maritime scenarios by introducing dynamic interference factors such as moving ships, floating obstacles, and tidal currents. Furthermore, the environment should be expanded to three-dimensional water spaces. By incorporating the measurement error characteristics of onboard sensors such as GPS and sonar, the robustness of the adaptive parameter mechanism to the maritime environment can be optimized, ensuring the algorithm can stably generate feasible paths across diverse scenarios including nearshore areas, ports, and open waters.Second, explore multi-algorithm fusion and performance upgrading. The improved algorithm can be combined with Deep Reinforcement Learning (e.g., DQN, DDPG) to address the "curse of dimensionality" faced by traditional Q-Learning algorithms in high-dimensional state spaces, thereby adapting to larger-scale and more complex path planning tasks. Simultaneously, integrating the local search advantages of heuristic algorithms (e.g., A*, Dijkstra) could further optimize path initialization and convergence speed.

References

[1] Meng Xiangdu. Research on Path Planning Algorithms for Unmanned Vessels [D]. Tianjin: Tianjin University, 2017

[2] Zhang Daheng. Research on Autonomous Path Planning for Intelligent Ships Based on Deep Learning [D]. Dalian Maritime University 2022.DOI:10.26989/d.cnki.gdlhu.2022.000001.

[3] DIJKSTRA E. W. A note on two problems in connexion with graphs[J]. Numerische mathematik, 1959, 1(1): 269-271.

[4] Hart P E, Nilsson N J, Raphael B. A formal basis for the heuristic determination of minimum cost paths[J]. IEEE Trans-actions on Systems Science and Cybernetics, 1968, 4(2) : 100－107.

[5] HOLLAND J H. Adaptation In Natural and Artificial Systems[M]. Bradford: A Bradford Book, 1992.

[6] DORIGO M, MANIEZZO V, COLOMI A. Ant system : optimization by a colony of cooperating agents[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 1996, 26(1): 29-41.

[7] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D., 2015.

[8] Lee H T, Kim M K. Optimal path planning for a ship in coastal waters with deep Q network[J]. Ocean Engineering, 2024, 307: 118193.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202602_12(2).0011

[9] Chen, C., Chen, X.Q., Ma, F., Zeng, X.J., Wang, J., 2019a. A knowledge-free path planning approach for smart ships based on reinforcement learning. Ocean Eng. 189, 106299 https://doi.org/10.1016/j.oceaneng.2019.106299.

[10] Shen Haiqing. Collision Avoidance Navigation and Control for Unmanned Ships Based on Reinforcement Learning [D]. Dalian Maritime University, 2018.

[11] Yoo B, Kim J. Path optimization for marine vehicles in ocean currents using reinforcement learning[J]. Journal of Marine Science & Technology, 2016, 21(2): 334-343.

[12] Wang Yinan. Research on Ship Collision Avoidance Based on Navigation Rules in Q-Learning [D]. Dalian Maritime University, 2022.

[13] Wang Y, Lu C, Wu P, et al. Path planning for unmanned surface vehicle based on improved Q-Learning algorithm[J]. Ocean Engineering, 2024, 292: 116510.

[14] Chen X, Hu R, Luo K, et al. Intelligent ship route planning via an A∗ search model enhanced double-deep Q-network[J]. Ocean Engineering, 2025, 327: 120956.

[15] Yang Juncheng, Li Shuxia, Cai Zengyu. Research and Development of Path Planning Algorithms [J]. Control Engineering 2017,24(07):1473- 1480.