Generalizable Spatiotemporal Reinforcement Learning Model for Maritime Search Path Planning

Pengcheng Yang
College of Systems Engineering,
National University of Defense Technology
Changsha, China
email: yangpengcheng@nudt.edu.cn

Yingying Gao
College of Systems Engineering,
National University of Defense Technology
Changsha, China
email: 15222638242@163.com

Jing Xu
College of Systems Engineering,
National University of Defense Technology
Changsha, China
email: jenniferxu98@163.com

Qingqing Yang*
College of Systems Engineering,
National University of Defense Technology
Changsha, China
email: yqq_1982@126.com

***Abstract*—Maritime search path planning is critical for enhancing search efficiency and seizing the golden rescue time in maritime search and rescue operations. To address the insufficient generalization of existing methods, this paper presents a spatiotemporally enhanced Reinforcement Learning (RL) model. By simulating the target's probability distribution via a mixed Gaussian distribution and incorporating a Long Short-Term Memory (LSTM) network into the Proximal Policy Optimization (PPO) approach, the model's ability to extract spatiotemporal features is enhanced. Furthermore, a threshold-based scenario-switching mechanism is designed to boost training stability. Experimental results demonstrate the model's exceptional generalization and significantly improved solution quality on both training and test sets.**

Keywords-maritime search and rescue; reinforcement learning; generalization ability; path planning

I. INTRODUCTION

In recent years, maritime accidents have increased in frequency and severity, posing growing challenges to maritime rescue operations. Maritime search and rescue operations can be divided into two phases: search and rescue, with the search phase being the most time-consuming and critical for rescue success. Therefore, it is an urgent problem to plan a scientific and efficient search path for search and rescue equipment.

Existing research can be divided into traditional methods and intelligent methods [1]. Traditional methods are computationally inefficient, while heuristic and other intelligent methods, though offering some flexibility, depend heavily on expert experience and exhibit limited cross-scene generalization abilities. In contrast, RL methods can autonomously learn through interaction with the environment and master general strategies for solving a class of problems, making them particularly suitable for dynamic and uncertain environments. However, current research on RL algorithms in the field of maritime search and rescue still has limitations. Many studies [2][3] only train and test in specific scenes, failing to implement general solutions for multiple scenarios, lacking generalization ability, and violating the original intention of deep reinforcement learning.

This study aims to enhance the generalization ability of RL methods in maritime search path planning by designing a RL model that can effectively extract spatiotemporal features, thereby accelerating solution planning and improving search and rescue success rates. In Section II, we present the methodology, including the scenario generation framework based on mixed Gaussian distribution, basic components of reinforcement learning, improvement of PPO algorithm, and improvement of training process. Section III presents the experiment and results. Section IV concludes the paper.

II. METHODOLOGY

The algorithm flow of this paper is shown in Figure 1.

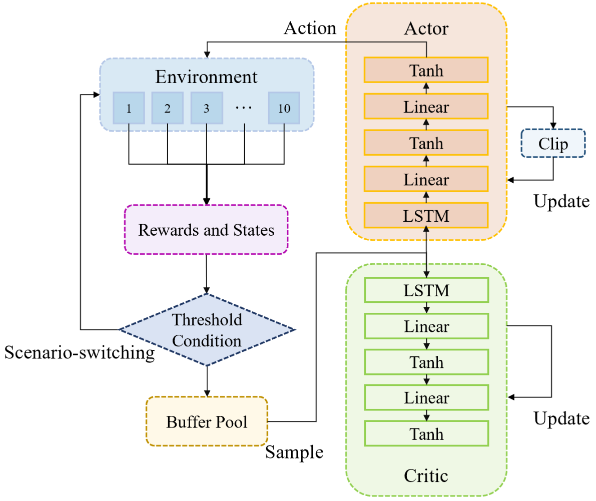

Figure 1. Algorithm Flow.

Copyright (c) IARIA, 2025. ISBN: 978-1-68558-300-2

A. Scenario Generation Framework Based on Mixed Gaussian Distribution

The initial position of a maritime search and rescue target is usually approximate. Generally, the Monte Carlo method is used to generate a large number of random particles near this coordinate. Then, real-time marine environmental data and physical oceanographic models are used to predict the approximate location of the target. This paper constructs a mixed Gaussian distribution (i.e., superimposing multiple two-dimensional Gaussian distributions) to simulate the probability distribution of the target's position over time. Each two-dimensional Gaussian distribution represents the possible position for an hour. Subsequently, the continuous space is discretized into a grid space through gridding, with the grid size being the search radius of the search and rescue equipment.

Since RL requires extensive interaction with the environment and the limited number of historical accident scenarios is insufficient to meet this requirement, this paper generates 1000 scenarios based on the mixed Gaussian distribution, 900 of which are randomly selected for initial training and 100 for later model performance testing.

B. Basic Components of Reinforcement Learning

The state space includes the observation window (the probability distribution of the agent and its surrounding square area, sized according to the perception range of the search and rescue equipment), the ratio of remaining time steps to total time steps, and the agent's current position.

The action space is defined as movement operations in four directions: east, south, west, and north. The step length of the agent in each direction is determined by the speed of the search and rescue equipment and the grid size of the environment.

The reward function consists of three parts: exploration reward (for grids with non-zero search probability), repeat penalty (for re-searching grids), and zero-value penalty (for grids with zero search probability), guiding the agent to explore effectively.

C. Improvement of PPO Algorithm

This paper improves the PPO algorithm, which is based on the Actor-Critic architecture. The Actor network generates action probability distributions, while the Critic network evaluates state value functions. As maritime search path planning is a time-series decision-making problem where steps are interrelated, the incorporation of LSTM modules into both the Actor and Critic networks enables the model to dynamically adjust attention to historical information, thereby enhancing its ability to extract spatiotemporal features.

D. Improvement of Training Process

This paper uses a vectorized parallel training framework to sample in parallel across 10 environments, quickly filling the experience replay pool and accelerating model training. Moreover, to prevent frequent scene switching from hindering model convergence during training across the 900 training scenarios, a threshold-based scene switching mechanism is designed. The threshold is determined by obtaining the optimal solution of a mixed-integer programming model using the Gurobi solver and setting 90% of this optimal solution as the threshold. The model switches to the next scene only after reaching this threshold in a training episode within the current scene, ensuring thorough learning before switching.

III. Experiment and Results

This section presents the experimental setup and results to validate the effectiveness and generalization capability of the proposed model.

A. Generalization Experiment

The model achieved an average score of 96.05% on the training set and 94.87% on the test set, indicating it has effectively learned the probability characteristics under the mixed Gaussian distribution and demonstrating strong generalization. Additionally, the average path planning time per scene was 0.55 seconds, meeting the strict timeliness requirements of maritime search tasks.

B. Ablation Experiment

To validate the effectiveness of each module in the model, we conducted ablation experiments comparing the complete model with variants lacking the observation space, LSTM module, and threshold-based switching mechanism, respectively. The results showed that the complete model outperformed the other variants in terms of average score, highlighting the importance of the designed modules in boosting model performance.

IV. Conclusion and Future Work

This paper proposed a spatiotemporally enhanced reinforcement learning model for maritime search path planning, which demonstrates strong generalization capabilities and computational efficiency. Experimental results indicate that while Gaussian distributions can effectively model target movement, they may not fully account for the complexity and unpredictability of real maritime scenarios. Future work will integrate real accident data to optimize the probability distribution model, enhancing its performance in practical rescue operations.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (72374209).

References

[1] J. Wu, L. Cheng, S. Chu, and Y. Song, “An autonomous coverage path planning algorithm for maritime search and rescue of persons-in-water based on deep reinforcement learning,” Ocean Engineering, vol. 291, pp. 116403–116423, Nov. 2023.

[2] B. Ai, et al. “Coverage path planning for maritime search and rescue using reinforcement learning,” Ocean Engineering, vol. 241, pp. 110098-110108, Dec. 2021.

[3] L. Liu, Q. Shan, and Q. Xu, “USVs Path Planning for Maritime Search and Rescue Based on POS-DQN: Probability of Success-Deep Q-Network,” Journal of Marine

Science and Engineering, vol. 12, no. 7, pp. 1158–1176, Jul. 2024.