18th International Conference on Innovative Computing, Information and Control (ICICIC2024)
Onsite (Qinhuangdao, China) and Online, September 10-13, 2024

PS-1: ICICIC2024-SS21-01

Research on Path Planning for Unmanned Surface Vessels Based on the Improved Proximal Policy Optimization Algorithm

Hongsong Zhao[1], Hongzhou Zhang[2,*] and Dawei Zhao[2]

[1]School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, P. R. China
[2]College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, P. R. China
*Corresponding author: zhz@hrbeu.edu.cn

Introduction

To effectively exploit marine resources, advances in information technology and the need for ocean exploration have given rise to the development of Unmanned Surface Vessels (USVs). As autonomous surface vessels, USVs play a key role in ocean missions, and their unmanned control systems are at the core of advancing ship technology.

Global path planning for unmanned vessels involves the application of various algorithms to automatically avoid obstacles in order to determine the best safe path in a map containing static obstacles such as reefs and islands. However, these algorithms take too long to compute and require high computational power, so it has been proposed that local path planning or dynamic obstacle avoidance focuses on real-time obstacle avoidance. In recent years, with the progress of science and technology, deep reinforcement learning algorithms have made significant breakthroughs in the field of path planning. Deep reinforcement learning algorithms have good decision-making ability and can explore and learn autonomously.

In this paper, an ICDM-PPO algorithm is proposed in order to improve the adaptability of collision avoidance algorithms to the environment and enhance the real-time planning ability of the algorithm. The main structure of this paper is as follows: firstly, we establish a surface unmanned boat motion model to study the autonomous path planning capability, and secondly, we integrate the Intrinsic Curiosity Distillation Module (ICDM) with the PPO algorithm, which drives the unmanned boat to explore the unknown state through the curiosity satisfaction value, and at the same time adopts distillation method to improve the accuracy of state prediction, and optimally adjusts intrinsic rewards by taking advantage of the uncertainty in the frequency of state exploration. The ICDM-PPO algorithm is formed. The algorithm combines strong exploration ability and high prediction ability to further improve the learning efficiency.

Research Questions

In the field of reinforcement learning, the main goal pursued by an Agent is to maximize the reward it receives through environmental feedback. However, in the face of sparse reward environments, it becomes a great challenge to construct a reward system that is both efficient and dense, which often leads the Agent to fall into a local optimum. In the unmanned boat path planning problem, the Agent obtains rewards and iteratively optimizes its strategy by interacting with the environment, but this interaction relies on blind stochastic exploration until the goal is reached.

Methodologies

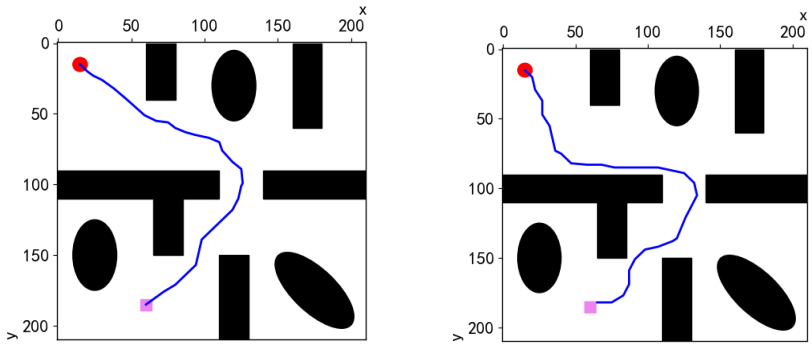
We propose an approach that combines the curiosity mechanism and the distillation module, which provides additional intrinsic incentives to facilitate the Agent's exploration of unknown environments when it encounters unexplored states. In addition, by fusing intrinsic and extrinsic rewards, the approach enables the Agent to learn and explore more efficiently, avoiding simply pursuing novelty states while neglecting the efficiency of goal-directed learning. Secondly, the intrinsic curiosity distillation module is integrated with the PPO algorithm (ICDM-PPO algorithm), which employs the distillation method to improve the accuracy of state prediction and optimally adjusts the intrinsic rewards by exploiting the uncertainty of state exploration frequency.

Tables (1)

Table 1. Parameter settings for simulation experiments

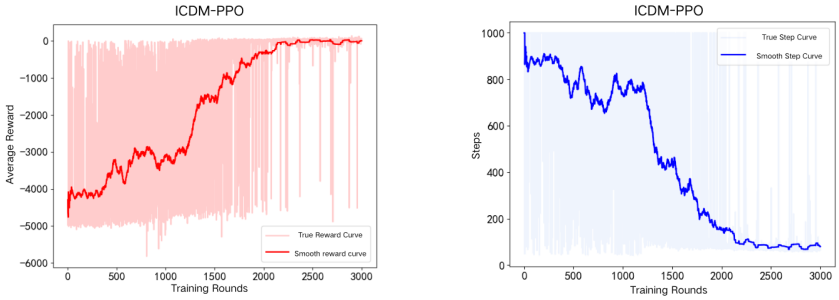
Parameters	Value
Maximum number of iteration steps per round step	1000
Maximum number of rounds episode	3000
Actor E-learning rate α	0.001
Critic E-learning rate β	0.001
Discount factor γ	0.9
Batch-size	64
Truncation factor ε	0.2
GAE parameter λ	0.95
ICDM learning rate	0.001
ICDM proportionality factor η	100
ICDM proportionality factor μ	10
ICDM loss factor	0.2

Figures



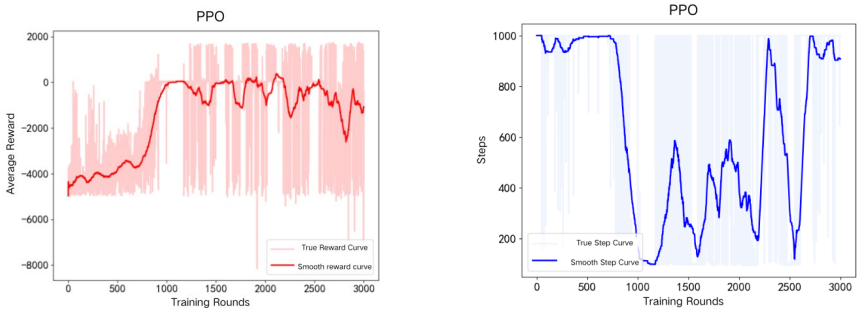
(a) ICDM-PPO algorithm path diagram (b) PPO algorithm path diagram

Figure 1. Comparison of paths for simulation experiment



(a) Incentive change curve (b) Step change curve

Figure 2. Training effect graph of ICDM-PPO algorithm



(a) Incentive change curve (b) Step change curve

Figure 3. Training effect graph of PPO algorithm

Tables (2)

Table 2. Specific data of simulation experiment

Algorithm	Path length	Step	Rewards	Success rate
ICDM-PPO algorithm	1299.1	85	0	63.3%
PPO algorithm	1453.25	unstable	unstable	37.6%

Conclusion

The full paper investigates the autonomous path planning capability of a surface unmanned boat in a static and complex environment containing special obstacles. A simulation map environment was created using Python's Tkinter module and PyTorch was applied to construct a deep neural network model as a means of evaluating the autonomy and intelligence capabilities of the unmanned boat. By integrating the ICDM module with the PPO algorithm, the ICDM-PPO algorithm was formed and validated in a specific environment with long obstacles. The experimental results show that compared with the PPO algorithm, the ICDM-PPO algorithm is able to guide the unmanned boat to avoid wandering and plan a shorter and smoother path more effectively in the special obstacle environment. In future work, collaborative multi-unmanned craft path planning can be investigated based on the results of this paper.