中国舰船研究
CHINESE JOURNAL OF SHIP RESEARCH
CSCD收录期刊 Scopus收录期刊
中文核心期刊 JST收录期刊
中国科技核心期刊 DOAJ收录期刊

基于改进DDPG算法的无人船自主避碰决策方法

关巍 郝淑慧 崔哲闻 王淼淼

Autonomous decision-making method of unmanned ship based on improved DDPG algorithm

GUAN Wei, HAO Shuhui, CUI Zhewen, WANG Miaomiao

在线阅读 View online: https://doi.org/10.19693/j.issn.1673-3185.03929

本文网址: http://www.ship-research.com/cn/article/doi/10.19693/j.issn.1673-3185.03929 **期刊网址:** www.ship-research.com

引用格式: 关巍, 郝淑慧, 崔哲闻, 等. 基于改进 DDPG 算法的无人船自主避碰决策方法 [J]. 中国舰船研究, 2025, 20(1): 172–180.

GUAN W, HAO S H, , et al. Autonomous decision-making method of unmanned ship based on improved DDPG algorithm[J]. Chinese Journal of Ship Research, 2025, 20(1): 172–180 (in Chinese).

基于改进 DDPG 算法的无人船自主避碰决策方法

扫码阅读全文

关巍*, 郝淑慧, 崔哲闻, 王淼淼

大连海事大学 航海学院, 辽宁 大连 116026

摘 要: [**目的**] 针对传统深度确定性策略梯度(DDPG)算法数据利用率低、收敛性差的特点, 改进并提出一种新的无人船自主避碰决策方法。[**方法**] 利用优先经验回放(PER)自适应调节经验优先级, 降低样本的相关性, 并利用长短期记忆(LSTM)网络提高算法的收敛性。基于船舶领域和《国际海上避碰规则》(COLREGs), 设置会遇情况判定模型和一组新定义的奖励函数, 并考虑了紧迫危险以应对他船不遵守规则的情况。为验证所提方法的有效性, 在两船和多船会遇局面下进行仿真实验。[**结果**] 结果表明, 改进的 DDPG 算法相比于传统 DDPG 算法在收敛速度上提升约 28.8%, [**结论**] 训练好的自主避碰模型可以使无人船在遵守 COLREGs 的同时实现自主决策和导航, 为实现更加安全、高效的海上交通智能化决策提供参考。

关键词: 无人船; 深度确定性策略梯度算法; 自主避碰决策; 优先经验回放; 国际海上避碰规则; 避碰

中图分类号: U664.82 **文献标志码:** A **DOI:** 10.19693/j.issn.1673-3185.03929

0 引 言

船舶作为全球贸易、物流和经济发展中至关重要的运输工具, 承担着连接世界各地货物和资源的重要使命。近年来, 随着海洋资源的不断开发, 传统船舶在海洋运输、深海探测等多个领域显现出一些局限性。国际海事组织报告指出, 海上 80% 的船舶碰撞事故是由人为失误引起的[1]。实现船舶的智能化与自主避碰已经成为国际海事的首要关注议题之一。

目前, 解决船舶自主避碰主要有传统方法和深度强化学习方法两类。传统方法包括启发式算法以及基于概率模型和数理模型的算法[2]。詹小飞等[3]提出一种基于多策略改进的麻雀搜索算法(multi-strategy improved sparrow search algorithm, MISSA), 该算法在路径距离、转向角度与次数等关键性能指标上具有优异表现。Mu 等[4]提出一种基于人工势场(artificial potential field, APF)的避障运动规划算法, 利用 APF 切换逻辑设计运动规划框架, 保证了交叉口意外障碍物场景下的安全避障。张一帆等[5]设计 APF 引导的双向快速扩展随机树(bidirectional rapidly-exploring random trees, Bi-RRT)算法, 实现无人船路径规划, 优化后的路径更适用于无人船的跟踪控制, 满足海上实际航行需求, 但并未考虑动态障碍物与《国际海上避碰规则》(international regulations for preventing collisions at sea, COLREGs)。宁君等[6]提出基于混合粒子群算法的无人船避碰决策方法, 基于模糊综合评价策略构建无人船碰撞危险度模型, 实现基于 COLREGs 多船会遇态势下的避碰路径规划。Guan 等[7]改进 A-star 算法设计无人水面船(unmanned surface vessel, USV)路径规划方法, 并引入改进的动态窗口方法(improved dynamic window approach, IDWA)以避免碰撞。然而, 传统方法面对复杂未知环境问题时表现出了一定的局限性。

在深度强化学习(deep reinforecement learning, DRL)方法方面, 2016 年, 谷歌 DeepMind 的 AlphaGo 战胜世界冠军李世石, 展示了 DRL 技术

收稿日期: 2024–05–14 **修回日期:** 2024–08–11 **网络首发时间:** 2025–01–15 08:03

基金项目: 国家自然科学基金资助项目(51409033, 52171342)

作者简介: 关巍, 男, 1982 年生, 博士, 教授, 博士生导师。研究方向: 船舶运动控制理论, 船舶自主避障决策。E-mail: gwwtxdy@dlmu.edu.cn

郝淑慧, 女, 2000 年生, 硕士生。研究方向: 船舶自主避障决策。E-mail: hjkhsh@163.com

***通信作者:** 关巍

在解决决策挑战方面的成效[8]。Guan 等[9]提出一种基于 Q 学习(deep Q network, DQN)算法训练的广义行为决策模型, 利用目标障碍区(obstacle zone by target, OZT)来计算未来碰撞的面积, 并将虚拟网格传感器作为观测状态的输入, 在开阔水域测试其性能优势。Shen 等[10]将深度 Q 学习算法应用于多船避碰, 并开发一种限制水域的测试方法。然而, DQN 存在 Q 值高估和动作空间离散的问题。Fan 等[11]提出一种符合 USV 机动性的强化学习防撞(reinforcement learning collision avoidance, RLCA)算法, 采用 double-DQN 方法减少动作值函数的高估。Chu 等[12]提出一种基于双深度 Q 网络 (double deep Q network, DDQN) 的 DRL 路径规划方法, 增强了欠驱动自主水下航行器(autonomous underwater vehicle, AUV)在未知环境下受洋流干扰的路径规划能力。宋利飞等[13]针对深度确定性策略梯度(deep deterministic policy gradient, DDPG)算法收敛速度慢和训练时容易出现局部最优的情况, 分离成功经验池与失败经验池, 但这种分类方式相对固定, 可能会忽略一些中间状态或非典型但重要的经验。胡正阳等[14]在 DDPG 算法基础上, 以 COLREGs 为基准设计相应的奖励函数, 引入势能回报塑形的思想来引导智能体学习最佳策略。作为行动者−评价者(actorcritic, AC)算法的改进, Wang 等[15]基于异步优势行动者−评价者(asynchronous advantage actor-critic, A3C)算法进行未知环境下的无人潜航器避障规划, 保证了规划路径的实时性和高效性。然而, A3C 的异步更新机制虽然加速了训练, 但也可能导致模型训练不稳定。相比之下, DDPG 算法通过经验回放和软性更新目标网络, 提高了样本利用效率和训练稳定性, 更适合处理连续动作空间的任务。

在前人研究的基础上[13, 14], 本文将提出一种基于改进 DDPG 算法的无人船自主避碰决策方法。在传统 DDPG 算法中引入优先经验回放(prioritized experience replay, PER)和长短期记忆(long short-term memory, LSTM)网络, 以增强算法的收敛性能。同时, 构建无人船三自由度模型与会遇局面判定模型, 将船舶领域与 COLREGs 融入奖励函数中, 并考虑紧迫危险下背离规则的情况。仿真实验包括两船会遇局面实验和多船会遇局面实验, 通过奖励函数曲线对比和路径对比验证所提方法是否具有优越性。

1　模型选择

1.1　无人船运动数学模型

建立无人船运动的数学模型旨在更好地分析无人船在航行中的操纵特性。本文选取包括艏摇、前进和横漂 3 个自由度的无人船平面运动数学模型[16]。无人船模型参数如表 1 所示。

表 1　无人船相关参数

Table 1　The principle parameters of the unmanned vehicle

参数	数值
船长/m	52.5
船宽/m	8.6
吃水深度/m	2.29
最大指令舵角/(°)	35
额定速度/kn	14

无人船运动数学模型方程为[17]

$$\begin{bmatrix} m-X_{\dot{u}} & 0 & 0 \\ 0 & m-Y_{\dot{v}} & mx_{\mathrm{C}}-Y_{\dot{r}} \\ 0 & mx_{\mathrm{C}}-N_{\dot{v}} & I_{zz}-N_{\dot{r}} \end{bmatrix} \begin{bmatrix} \Delta\dot{u} \\ \dot{v} \\ \dot{r} \end{bmatrix} = \begin{bmatrix} X_u & 0 & 0 \\ 0 & Y_v & Y_r-mu_0 \\ 0 & N_v & N_r-mx_{\mathrm{C}}u_0 \end{bmatrix} \begin{bmatrix} \Delta u \\ v \\ r \end{bmatrix} + \begin{bmatrix} 0 \\ Y_\delta \\ N_\delta \end{bmatrix} \delta \tag{1}$$

式中: m 为无人船质量; $[X_*, Y_*, N_*]$为流体动力学导数; u_0为无人船在等速直线运动时的速度; x_{C} 为无人船质心坐标; u, v, r 为无人船在 3 个自由度的速度; δ为无人船舵角。建立无人船平面运动的线性数学模型后, 无人船的加速度、速度、位置等参数可以通过实际的舵角来计算。

1.2　船舶领域模型

1963 年, 藤井首次提出了船舶领域的概念[18]。本研究采用该概念计算碰撞风险区域, 并参考 Śmierzchalski 简化的六边形域模型[19], 将其进一步简化为圆形, 如图 1 所示。一旦他船(target ship, TS)进入该圆形区域, 则表示本船 (own ship, OS) 遭遇紧迫危险。圆的半径计算公式为

$$d_5 = L \cdot V^{1.26} + 30V + u^* \tag{2}$$

式中: L 为船长; V 为船速; u^*为无人船航行时的位置误差。

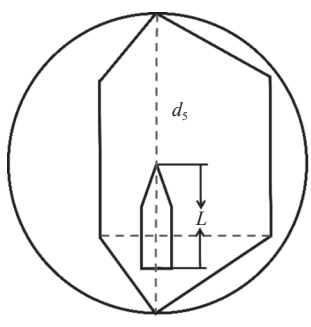


图 1　船舶领域模型

Fig. 1　Ship domain model

1.3 COLREGs 模型

COLREGs 旨在规范各类船舶在海上交通中的行为。COLREGs 第二章第 14～17 条明确规定了当探测到与他船有碰撞危险时，他船和本船之间的会遇情况判断方法以及相应的防撞措施[20]。

本文依据相对方位划分了 6 种情况，在每一种相对方位下，依据相对航向将会遇划分为 7 种局面，如图 2 所示。在对遇局面下，OS 与 TS 均为让路船，需右转从对方右舷通过。在右交叉局面下，OS 为让路船，应右转从 TS 右侧通过，TS 为直航船，应保速保向。与右交叉局面不同，在大角度右交叉局面下，OS 应左转避让 TS。在追越局面下，OS 速度大于 TS，OS 为让路船，应右转从 TS 前侧通过，TS 保速保向。在左交叉和被追越局面下，OS 为直航船，TS 为让路船。其中，被追越局面要求 OS 速度小于 TS。安全航行局面意味着 OS 与 TS 均不需要采取任何避碰行动[2]。

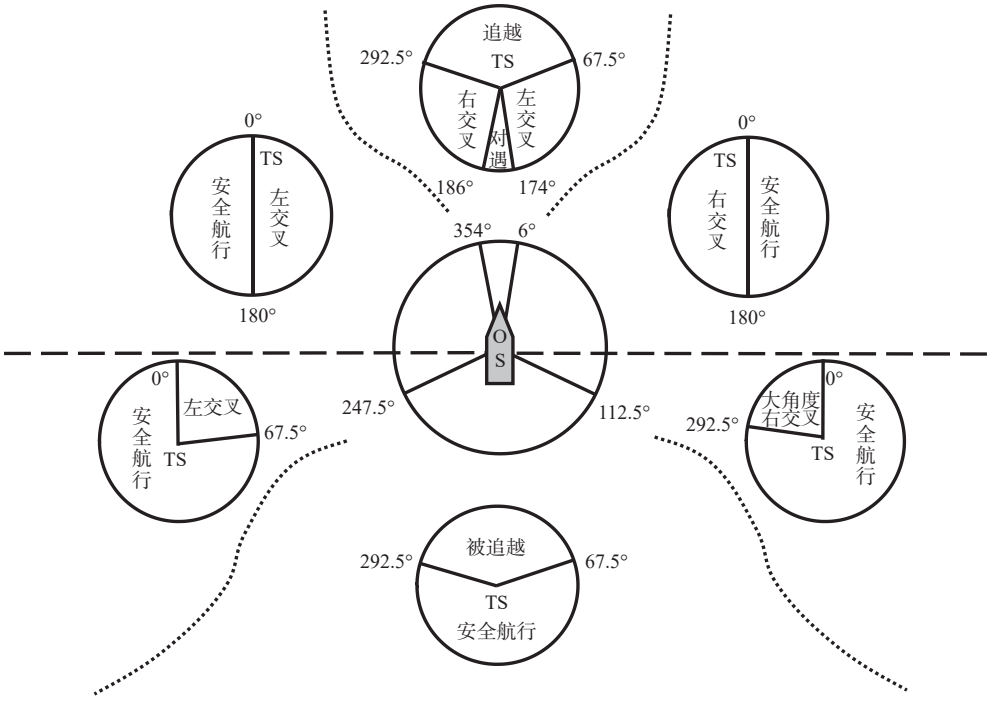


图 2 典型会遇局面情况分类

Fig. 2 The classification of typical encounter situations

1.4 会遇局面判定模型

在使用本文提出的方法进行无人船自主避碰决策之前，需要建立一个无人船会遇局面判定模型，该模型的工作流程如图 3 所示。

首先计算本船船舶领域。当 OS 雷达线探测到他船时，若他船进入本船船舶领域，OS 陷入紧迫危险，有必要背离 COLREGs 第二章第 14～17 条，直接被判定为让路船。若他船未侵犯本船船舶领域，模型将判断 OS 是否处于两船会遇局面。

若 OS 处于两船会遇局面，模型将根据 COLREGs 第二章第 14～17 条进行具体会遇局面划分及责任认定。若 OS 处于多船会遇局面，模型将表现得像一艘让路船，根据雷达线探测的距离远近依次避让。

2 改进的 DDPG 算法与奖励函数设计

2.1 改进的 DDPG 算法

传统 DDPG 算法基于行动者−评价者（actor-critic, AC）框架，输出确定性动作来提升算法的收敛性和稳定性。本文对 DDPG 算法在两个方面进行改进：引入优先经验回放机制；采用长短期记

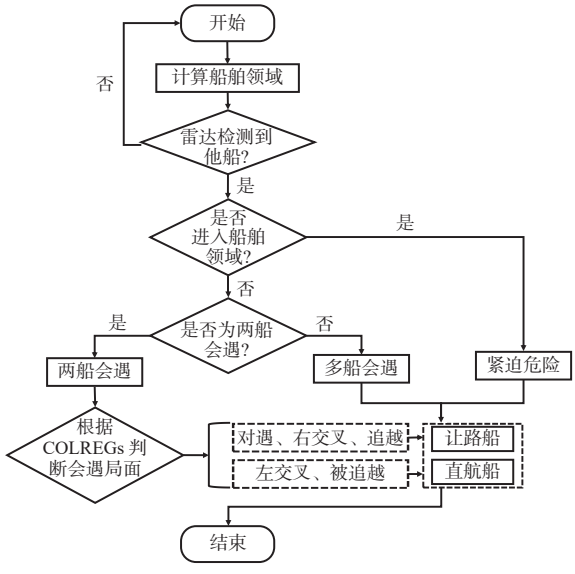


图 3 会遇局面判定流程

Fig. 3 The determining process for the ships encounter situations

忆网络。

2.1.1　优先经验回放

传统 DDPG 算法在借鉴 DQN 算法思想的基础上, 采用均匀经验回放。然而, 这种方法导致部分较差经验的利用率较高, 不利于算法的收敛。为提高传统 DDPG 算法的训练效率和收敛速度, 本研究将其与优先经验回放机制相结合, 数据的优先级取决于经验的价值。改进的 DDPG 算法利用时序差分(temporal difference-error, TD-error)评估经验的价值, 其中 TD-error 的公式为:

$$Q(s,a|\theta_Q)=r+\gamma Q'(s',a'|\theta_{Q'}) \tag{3}$$

$$T=r+\gamma Q'(s',a'|\theta_{Q'})-Q(s,a|\theta_Q) \tag{4}$$

式中: $Q(s,a|\theta_Q)$为当前状态下动作的价值函数; $Q'(s',a'|\theta_{Q'})$为下一个状态下动作的价值函数; s 为状态值; a 为动作值; θ_Q为价值网络参数; r 为奖励值; γ 为折扣因子; T 为当前经验的 TD-error, TD-error 值越大, 表示该经验的价值越高, 被抽取的优先级越高。因此, 数据的优先值定为:

$$P(i)=p_i/\sum p_i \tag{5}$$

$$p_i=|T_i+\varepsilon| \tag{6}$$

式中: $P(i)$为第 i 个数据的优先值; p_i 为第 i 个数据的价值; T_i 为第 i 个数据的 TD-error; ε为一个常量, ε的引入是为了防止某些经验的概率为 0。

每次产生新数据后, 经验池内数据的优先级将被更新, 为减少排序的工作量, 采用二叉树思想进行数据采样。

二叉树思想可分为存储优先值和基于优先值采样两步。在模型训练之前, 为所有经验赋予一个初始优先值, 并将其分配给二叉树的底层节点。树的最底层存储着经验的优先值, 而树的最顶层存储所有经验优先值之和。在开始训练后每次计算 TD-error 时, 经验的优先值被更新并被存入二叉树中。随后, 改进的 DDPG 算法在一定范围内进行随机采样, 其中经验的优先级越高, 被采样到的概率也越大。经过优先经验回放, 改进 DDPG 算法每次采样到高价值经验的概率被提升, 策略网络和价值网络的损失函数为:

$$L_\mu=-Q(s,\mu(s|\theta_\mu)|\theta_Q) \tag{7}$$

$$L_Q=[r_t+\gamma(1-D)Q'(s',a'|\theta_{Q'})-Q(s,a|\theta_Q)]^2 \tag{8}$$

式中: μ代表策略网络; θ_μ为策略网络参数; r_t为 t 时刻的奖励值; D 为 1 或 0, 分别代表 USV 是否发生碰撞。

2.1.2　长短期记忆网络

传统 DDPG 算法使用全连接网络结构, 为解决一般递归神经网络存在的长时间依赖问题并增强对先前经验的记忆性, 本文采用 LSTM 网络对状态信息进行预处理。改进 DDPG 算法的交互过程和网络结构如图 4 所示。图中: h_t 表示时间步 t 的隐藏状态; c_t 表示时间步 t 的记忆状态; $\tilde{a}$表示策略网络重新选择的动作; a'表示策略目标网络重新选择的动作; 变量 s_t, a_t 在后文解释。其

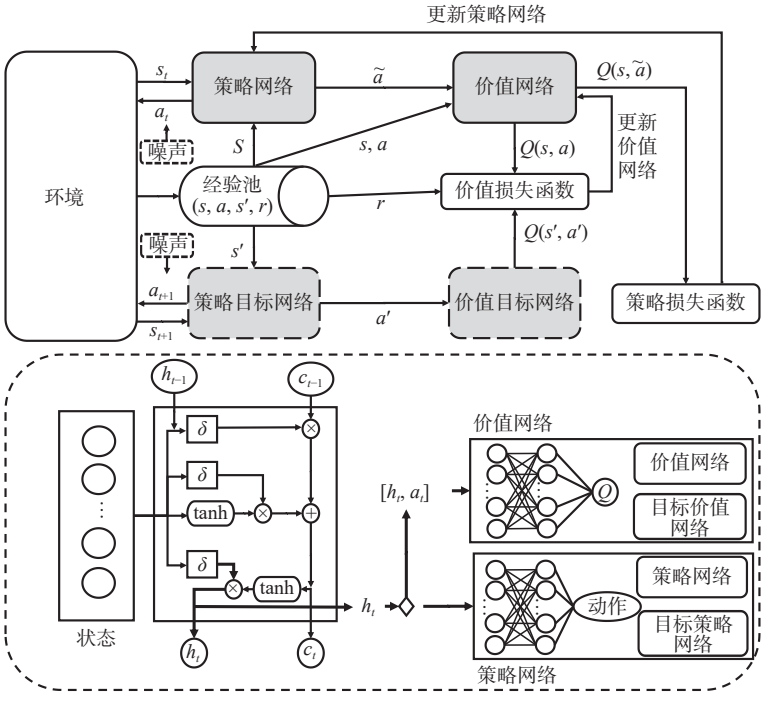


图 4　改进 DDPG 算法交互过程及网络结构

Fig. 4　The interaction process and network structure of the improved DDPG algorithm

中，改进 DDPG 算法具有 2 个策略网络和 2 个价值网络：策略网络输入状态信息，经过 LSTM 网络后输出正态分布的均值和方差，并通过抽样获得具体动作；价值网络输入 LSTM 网络处理后的状态信息和动作信息，输出当前状态下的动作价值函数。

在改进的 DDPG 算法中，状态信息设置为

$$s_t = [(l_1, l_2, \cdots, l_{37}), d, \varphi] \tag{9}$$

式中：$(l_1, l_2, \cdots, l_{37})$为 37 根雷达线探测到的数据；$d$ 为无人船到终点的距离；φ为无人船到终点的相对方位角。

本文动作空间定义为

$$a_t = [\delta_t] \tag{10}$$

式中，δ_t为无人船的指令舵角，取值范围为 [−25°, 25°] 的连续变量。

2.2 奖励函数设计

强化学习是一种通过最大化奖励函数来实现目标的方法。改进 DDPG 算法的训练成功与否取决于奖励函数的设计。Christiano 等[21] 在文献中探讨了主线奖励与辅助奖励间的关系。主线奖励大且直观，智能体通过试错获取，达成任务。然而，在面对复杂任务时，由于正样本的稀缺性，这一策略往往面临稀疏奖励的挑战。因此，需将任务分解为子目标，并配以适当奖惩以提高任务成功率，即信用分配[22-23]。

本研究中，终点奖励作为主线奖励，旨在确保智能体正确行动，即避障与到达终点；辅助奖励则用于优化奖励机制，包括制导奖励、航向修正奖励及 COLREGs 奖励，以平衡整体表现。

2.2.1 终点奖励

终点奖励R_g旨在引导无人船到达终点并实现避碰。当无人船到达终点时，给予较大正奖励；当无人船发生碰撞时，给予较大负奖励，定义为

$$R_g = \begin{cases} -20, & D = 1 \\ 200, & G = 1 \end{cases} \tag{11}$$

式中：$D = 1$表示无人船发生碰撞；$G = 1$表示无人船到达终点。

2.2.2 制导奖励

制导奖励R_d旨在鼓励无人船驶向终点。当无人船接近终点时，制导奖励为正；当无人船远离终点时，制导奖励为负，定义为

$$R_d = \lambda_g\left(\sqrt{(x_p - x_g)^2 + (y_p - y_g)^2} - \sqrt{(x - x_g)^2 + (y - y_g)^2}\right) \tag{12}$$

式中：λ_g为制导奖励权重；(x_p, y_p)为无人船上一时刻位置坐标；(x, y)为无人船当前时刻位置坐标；(x_g, y_g)为终点位置坐标。

2.2.3 航向修正奖励

实际航行中，无人船会因为避碰措施等偏离预计航线，航向修正奖励R_{yaw}旨在引导无人船返回预计航线，定义为

$$R_{yaw} = \lambda_{yaw} e^{(\varepsilon|\varphi_c|)^2 \sqrt{(x - x_g)^2 + (y - y_g)^2}} \tag{13}$$

式中：λ_{yaw}为航向修正奖励权重；ε为角度误差权重；φ_c为终点方向和无人船实际航向的差值。

2.2.4 COLREGs 奖励

COLREGs 奖励R_C鼓励无人船依据 COLREGs 采取防撞行动。当雷达线未探测到周围他船或障碍物信息时，COLREGs 奖励为正，鼓励探索；当雷达线探测到他船或障碍物，但距离仍大于船舶领域时，无人船需依据 COLREGs 实现两船避碰或多船避碰；当他船或障碍物侵犯船舶领域时，无人船陷入紧迫危险，COLREGs 奖励为负，此时无人船决策需考虑背离 COLREGs 第二章第 14～17 条。该奖励函数定义为：

$$R_C = \begin{cases} 0.2, & r \geqslant r_l \\ \lambda_C \sqrt{(x - x_g)^2 + (y - y_g)^2}, & r_d < r \leqslant r_l \text{且} C = \text{True} \\ 0, & r_d < r \leqslant r_l \text{且} C = \text{False} \\ -0.2, & r \leqslant r_d \end{cases} \tag{14}$$

式中：λ_C为 COLREGs 奖励权重；r_l为雷达线长度；r_d为船舶领域半径；C 为布尔值，True 表示无人船遵守 COLREGs，False 表示无人船背离 COLREGs。

总奖励R_{total}定义为上述奖励之和：

$$R_{total} = R_g + R_d + R_{yaw} + R_C \tag{15}$$

3 实验分析

3.1 仿真设置

为全面验证本文所提方法在复杂海洋环境中的性能，设计并实施了一系列仿真实验。具体而言，选择 ROS 作为仿真平台，在 Gazebo的框架下，首先构建 1.1 节所述的三自由度无人船模型；然后，搭建一个具有风、浪干扰的三维海洋仿真环境，风浪干扰具体参数如表 2 所示。

算法构建方面，采用改进后的 DDPG 算法训练模型，训练过程中的详细参数如表 2 所示。发生碰撞或超出训练最大步数均代表一次训练轮次的结束，算法输出迭代奖励值并进行参数更新，

表 2　改进 DDPG 算法相关参数及奖励函数参数

Table 2　The parameters of improved DDPG algorithm and reward functions

参数	数值	参数	数值
折扣率γ	0.99	制导奖励权重λ_g	5
策略网络学习率	0.000 3	航向修正奖励权重λ_{yaw}	−2.5
价值网络学习率	0.000 3	COLREGs奖励权重λ_C	6.5
LSTM 隐藏层数量	256	雷达线长度r_l/n mile	3.5
软更新系数	0.01	船舶领域半径r_d/n mile	1.4
批次大小	256	角度权重ε	0.23
波风向/(°)	31.4	有效波高/m	0.17
波浪周期/s	6.2	平均风速/(m·s^{-1})	0.38

直到迭代奖励值随着训练轮次的增加趋于稳定。

实验设计方面, 分别在两船会遇局面和多船会遇局面下进行仿真测试, 并在多船会遇局面下与传统 DDPG 算法进行对比实验。其中, OS 和 TS 均使用 1.1 节中描述的船舶运动数学模型。OS 使用如表 1 所示的船舶模型参数。TS 为有人船, 其船长、船宽和吃水深度与表 1 中的参数相同。有关速度的详细信息参见每组实验中的初始化信息表格。

3.2　算法奖励函数曲线对比实验

为验证改进的 DDPG 算法的性能相比于传统 DDPG 算法是否有所提升, 进行奖励函数对比实验, 如图 5 所示。其中, 两种算法训练时均采用相同的训练参数和奖励函数。在 4 500 轮次左右的训练过程中, 改进的 DDPG 算法的收敛速度相较于传统 DDPG 算法提升 28.8% 左右, 收敛曲线更加平滑稳定, 收敛值也相对较高。

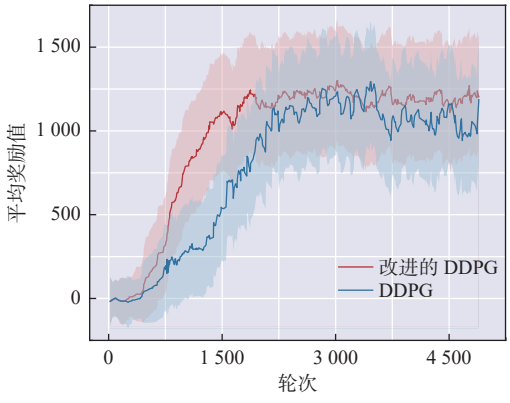


图 5　奖励函数对比曲线

Fig. 5　Reward function comparison curves

3.3　仿真实验

3.3.1　两船会遇局面

本小节进行两船会遇局面的仿真实验, 其初始信息如表 3 所示, 避碰过程、本船舵角变化及两船间距离变化如图 6 所示。由于安全航行会遇局面不涉及任何避碰操作, 加入实验中无明显意义, 因此只给出 6 种会遇局面。

表 3　两船会遇过程初始信息

Table 3　The initial data of two-ship encounter situation

会遇局面	无人船	起始位置/n mile	目标位置/n mile	速度/kn
对遇	OS	(−1.0, −1.0)	(5.0, 6.0)	14.00
	TS	(3.0, −3.0)	(−2.0, −2.0)	9.18
右交叉	OS	(1.0, −4.0)	(−1.0, 6.0)	14.00
	TS	(4.0, 0)	(−4.0, 0)	9.60
左交叉	OS	(−4.0, 0)	(1.0, 8.0)	14.00
	TS	(−6.0, 4.0)	(2.0, 4.0)	9.60
追越	OS	(−1.0, −4.5)	(−.01, 8.0)	14.00
	TS	(−1.0, −2.0)	(−1.0, 4.0)	6.00
大角度右交叉	OS	(−2.2, −3.1)	(0.8, 2.7)	14.00
	TS	(0.5, −3.4)	(−0.2, 1.3)	9.90
被追越	OS	(4.5, 3.6)	(4.5, 5.9)	14.00
	TS	(4.5, 2.6)	(4.5, 6.8)	19.00

1) 在对遇局面下, 当两船间距小于 3.5 n mile, 雷达线探测到 TS 时, OS 有碰撞危险, 并被判定为让路船。OS 需向右转, 直到避碰结束后方驶回计划航线。

2) 在右交叉局面下, TS 位于 OS 的右舷 [6º, 112.5º] 范围内。OS 为让路船, TS 为直航船。此时, OS 需向右转向以避开 TS, 直到避碰结束后方驶回计划航线。

3) 在左交叉局面下, TS 位于 OS 左舷 [247.5º, 355º] 范围内。OS 为直航船, TS 为让路船。由于 TS 不改变航向和速度, 违反了 COLREGs。OS 需保持速度和航向, 直到 TS 入侵船舶领域, 而 OS 直接被判定为让路船, 向右转向以避开 TS。

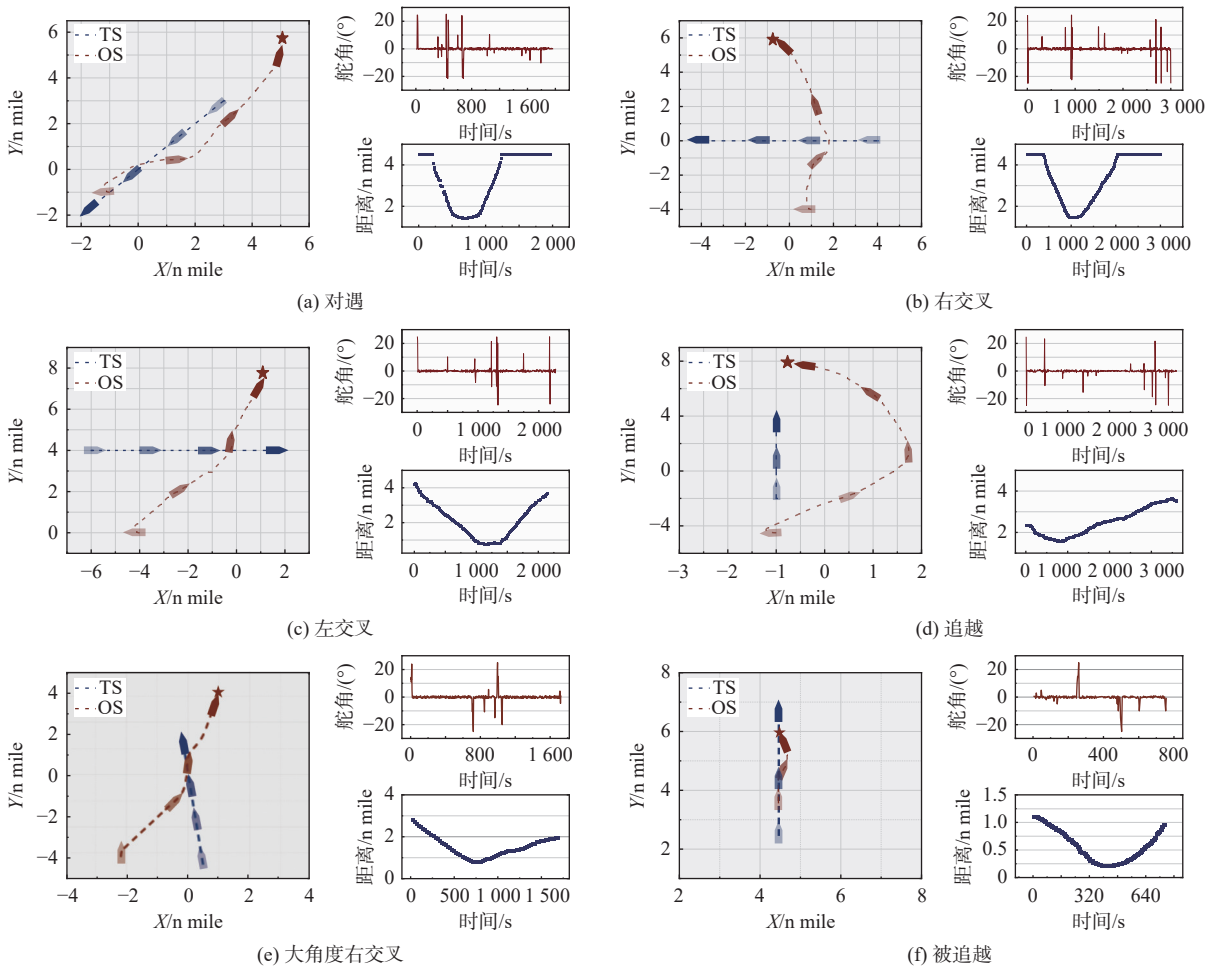
4) 在追越局面下, OS 速度大于 TS。当两艘船之间的距离小于 3 n mile 时, OS 位于 TS 的右舷 [112.5º, 247.5º] 范围内。OS 为让路船, 需向右转以避开 TS。

5) 在大角度右交叉局面下, TS 位于 OS 的右舷 [90º, 112.5º] 范围内。OS 为让路船, 需向左转以避开 TS。

6) 在被追越局面下, TS 速度大于 OS, TS 位于 OS[112.5º, 247.5º] 范围内。OS 为直航船, TS 为让路船。由于 TS 不改变航向和速度, 违反了 COLREGs。TS 入侵船舶领域后, OS 直接被判定为让路船, 向右转向以避开 TS。

3.3.2　多船会遇局面

本小节进行了三船会遇局面的仿真实验, 三



(a) 对遇 (b) 右交叉

(c) 左交叉 (d) 追越

(e) 大角度右交叉 (f) 被追越

图 6 两船会遇局面下无人船轨迹、舵角及两船间距离

Fig. 6 The trajectory, rudder angle and distance under the two-ships encounter situations

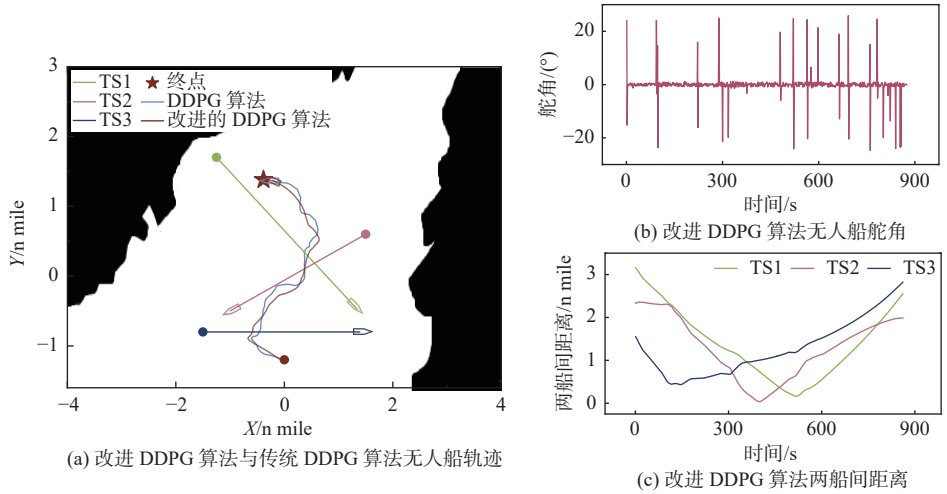
船会遇过程初始信息如表 4 所示，避碰过程、本船舵角变化及两船间距离变化如图 7 所示。

在 OS 雷达线探测到周围三艘船后，经模型判定，处于多船会遇局面，OS 应为让路船，并需要依据雷达线检测的顺序依次避碰。当 TS3 与 OS 形成左交叉局面且未按 COLREGs 要求航行时，OS 首先需要保速保向，当遭遇紧迫危险时需

表 4 三船会遇过程初始信息

Table 4 The initial information of there-ship encounter situation

无人船	起始位置/n mile	目标位置/n mile	速度/kn
OS	(0, −1.20)	(−0.40, 1.40)	14.00
TS1	(−1.25, 1.70)	(1.35, −0.40)	13.75
TS2	(1.50, 0.60)	(−1.00, −0.50)	11.20
TS3	(−1.50, −0.80)	(1.40, −0.80)	11.90



(a) 改进 DDPG 算法与传统 DDPG 算法无人船轨迹

(b) 改进 DDPG 算法无人船舵角

(c) 改进 DDPG 算法两船间距离

图 7 三船会遇局面下无人船轨迹，舵角及两船间距离

Fig. 7 The trajectory, rudder angle and distance under there-ships encounter situation

向右转向以避开 TS3。当 TS2 与 OS 形成右交叉局面时, OS 向右转向从 TS2 后方通过。成功避开 TS2 后, TS3 与 OS 形成了左交叉局面且未按 COLREGs 要求航行。OS 保速保向直到 TS3 侵犯船舶领域, 此时 OS 向右转向以避开 TS3。会遇局面结束后, OS 向左转向驶回计划航线。

为验证本文所提改进 DDPG 算法在决策能力方面与传统 DDPG 算法的区别, 在同样环境对比其的轨迹, 如图 7 所示。为避免单次仿真数据的偶然性, 基于改进 DDPG 算法进行 5 次实验, 航行轨迹如图 8 所示。针对路径长度、决策时间和最小会遇距离, 分别计算改进 DDPG 算法 5 次仿真数据的平均值, 并与传统 DDPG 算法进行对比。如表 5 所示, 在相同的环境下, 改进 DDPG 算法的路径长度更短, 曲线更平滑, 决策时间更短, 效果更好。

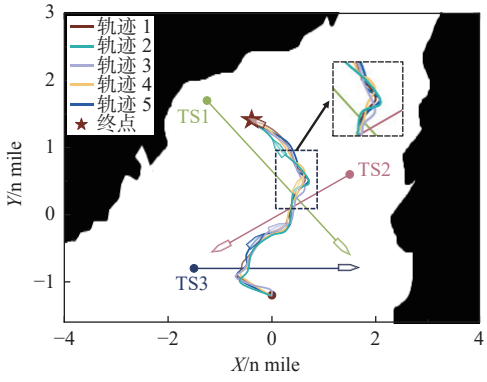


图 8　改进 DDPG 算法多次实验轨迹图

Fig. 8　The experiment trajectories of improved DDPG algorithm

表 5　两种算法轨迹对比

Table 5　The trajectories comparison for two algorithms

算法	实验次数	路径长度/n mile	决策时间/s	最小会遇距离/n mile
改进DDPG算法	1	3.403	875.0	0.760
	2	3.352	862.0	0.900
	3	3.348	861.0	0.650
	4	3.316	853.0	0.730
	5	3.432	882.0	0.840
传统DDPG算法		3.596	924.6	0.620
平均值		3.370	866.6	0.776

4　结　语

本文提出了一种基于改进 DDPG 算法的无人船自主避碰决策方法, 利用 PER 与 LSTM 网络提高算法的性能。此外, 结合船舶领域概念和 COLREGs, 建立一个会遇局面判定模型, 设计一组奖励函数, 并考虑紧迫危险, 以应对他船不遵守规则的情况。

基于训练好的模型, 设置了两组仿真实验, 分别在两船和多船会遇局面下验证模型的有效性。最终实验结果表明, 模型在符合 COLREGs 的情况下实现两船和多船避碰, 到达预定终点, 并能及时应对他船不遵守规则的情况。通过与传统 DDPG 算法进行对比实验, 证明了改进 DDPG 算法在奖励函数曲线收敛性和自主避碰决策方面的优势, 对海上交通领域的智能化决策具有一定的参考价值。

参考文献:

[1] MOKHTARI A H, DIDANI H R K. An empirical survey on the role of human error in marine incidents[J]. TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation, 2013, 7(3): 363–367.

[2] 关巍, 崔哲闻, 罗文哲. 基于改进 PPO 算法的船舶自主避碰决策 [J]. 大连海事大学学报, 2023, 49(4): 28–36.
GUAN W, CUI Z W, LUO W Z. Ship autonomous collision avoidance decision based on improved PPO algorithm[J]. Journal of Dalian Maritime University, 2023, 49(4): 28–36 (in Chinese).

[3] 詹小飞, 赵红, 王宁, 等. 基于多策略改进麻雀搜索算法的无人艇路径规划 [J]. 大连海事大学学报, 2024, 50(1): 1–10.
ZHAN X F, ZHAO H, WANG N, et al. Multi-strategy improved sparrow search algorithm-based path planning of unmanned surface vehicle[J]. Journal of Dalian Maritime University, 2024, 50(1): 1–10 (in Chinese).

[4] MU R, YU W H, LI Z X, et al. Motion planning for autonomous vehicles in unanticipated obstacle scenarios at intersections based on artificial potential field[J]. Applied Sciences, 2024, 14(4): 1626.

[5] 张一帆, 史国友, 徐家晨. 基于人工势场法引导的 Bi-RRT 的水面无人艇路径规划算法 [J]. 上海海事大学学报, 2022, 43(4): 16–22.
ZHANG Y F, SHI G Y, XU J C. Path planning algorithm of unmanned surface vehicles based on Bi-RRT guided by artificial potential field[J]. Journal of Shanghai Maritime University, 2022, 43(4): 16–22 (in Chinese).

[6] 宁君, 黄寓旸, 尤恽, 等. 基于混合粒子群算法的船舶避碰决策 [J]. 大连海事大学学报, 2023, 49(1): 34–43.
NING J, HUANG Y Y, YOU Y, et al. Ship collision avoidance decision based on hybrid particle swarm algorithm[J]. Journal of Dalian Maritime University, 2023, 49(1): 34–43 (in Chinese).

[7] GUAN W, WANG K. Autonomous collision avoidance of unmanned surface vehicles based on improved a-star and dynamic window approach algorithms[J]. IEEE Intelligent Transportation Systems Magazine, 2013, 15(3): 36–50.

[8] SILVER D, HUBERT T, SCHRITTWIESER J, et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play[J]. Science, 2018, 362(6419): 1140–1144.

[9] GUAN W, ZHAO M Y, ZHANG C B, et al. Generalized behavior decision-making model for ship collision

avoidance via reinforcement learning method[J]. Journal of Marine Science and Engineering, 2023, 11(2): 273.

[10] SHEN H Q, HASHIMOTO H, MATSUDA A, et al. Automatic collision avoidance of multiple ships based on deep Q-learning[J]. Applied Ocean Research, 2019, 86: 268–288.

[11] FAN Y S, SUN Z, WANG G F. A novel reinforcement learning collision avoidance algorithm for USVs based on maneuvering characteristics and COLREGs[J]. Sensors, 2022, 22(6): 2099.

[12] CHU Z Z, WANG F L, LEI T J, et al. Path planning based on deep reinforcement learning for autonomous underwater vehicles under ocean current disturbance[J]. IEEE Transactions on Intelligent Vehicles, 2023, 8(1): 108–120.

[13] 宋利飞, 许传毅, 郝乐, 等. 基于改进 DDPG 算法的无人艇自适应控制 [J]. 中国舰船研究, 2024, 19(1): 137–144.
SONG L F, XU C Y, HAO L, et al. Adaptive control of unmanned surface vehicle based on improved DDPG algorithm[J]. Chinese Journal of Ship Research, 2024, 19(1): 137–144 (in Chinese).

[14] 胡正阳, 王勇. 基于深度确定性策略梯度的船舶自主航行避碰方法 [J]. 指挥控制与仿真, 2024(5): 37–44.
HU Z Y, WANG Y. A deep deterministic policy gradient method for collision avoidance of autonomous ship[J]. Command Control & Simulation, 2024(5): 37–44 (in Chinese).

[15] WANG H J, GAO W, WANG Z, et al. Research on obstacle avoidance planning for UUV based on A3C algorithm[J]. Journal of Marine Science and Engineering, 2024, 12(1): 63.

[16] GUAN W, PENG H W, ZHANG X K, et al. Ship steering adaptive CGS control based on EKF identification method[J]. Journal of Marine Science and Engineering, 2022, 10(2): 294.

[17] PERERA L P, OLIVEIRA P, GUEDES SOARES C. System identification of nonlinear vessel steering[J]. Journal of Offshore Mechanics and Arctic Engineering, 2015, 137(3): 031302.

[18] DAVIS P V, DOVE M J, STOCKEL C T. A computer simulation of marine traffic using domains and arenas [J]. Journal of Navigation, 1980, 33(2): 215–222.

[19] ŚMIERZCHALSKI R. Ships' domains as collision risk at sea in the evolutionary method of trajectory planning[M]// SAEED K, PEJAŚ J. Information Processing and Security Systems. Boston: Springer, 2005: 411−422.

[20] CUI Z W, GUAN W, LUO W Z, et al. Intelligent navigation method for multiple marine autonomous surface ships based on improved PPO algorithm[J]. Ocean Engineering, 2023, 287: 115783.

[21] CHRISTIANO P F, LEIKE J, BROWN T B, et al. Deep reinforcement learning from human preferences[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc. , 2017: 4299−4307.

[22] ZHENG Z Y, OH J, SINGH S. On learning intrinsic rewards for policy gradient methods[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal: Curran Associates Inc., 2018: 4644−4654.

[23] ZHENG Z Y, OH J, HESSEL M, et al. What can learned intrinsic rewards capture?[C]//Proceedings of the 37th International Conference on Machine Learning. PMLR, 2019: 1060.

Autonomous decision-making method of unmanned ship based on improved DDPG algorithm

*GUAN Wei**, *HAO Shuhui*, *CUI Zhewen*, *WANG Miaomiao*

Navigation College, Dalian Maritime University, Dalian 116026, China

Abstract: ［**Objectives**］To enhance the safety and efficiency of maritime traffic, this paper proposes an autonomous collision avoidance decision-making method for unmanned ships based on an enhanced Deep Deterministic Policy Gradient (DDPG) algorithm. ［**Methods**］In order to address the issues of low data utilization and poor convergence in traditional DDPG algorithms, we employ Priority Experience Replay (PER) to dynamically adjust experience priority, reduce sample correlation, and utilize the Long Short-Term Memory (LSTM) network to improve the algorithm convergence. Based on the domain knowledge of ships and adhering to the International Regulations for Preventing Collisions at Sea (COLREGs), a model for determining meeting situations and a novel set of reward functions that consider urgent scenarios when other ships fail to comply with the COLREGs are introduced. Generalization experiments are conducted involving two-ship and multi-ship encounters to validate the effectiveness of the proposed method. ［**Results**］As the experimental results demonstrate, compared to traditional DDPG algorithms, our improved approach enhances the convergence speed by approximately 28.8%. ［**Conclusions**］The trained model enables autonomous decision-making and navigation while ensuring compliance with the COLREGs, thereby providing valuable insights for intelligent decision-making in the field of maritime transportation.

Key words: unmanned vehicles; deep deterministic policy gradient (DDPG) algorithm; autonomous collision avoidance decision-making; priorit-ized experience replay (PER); international regulations forpreventing collisions at sea (COLREGs); collision avoidance