[引用格式] 侯玉立, 王宁, 邱赤东, 等. 无人艇集群路径规划研究综述: 深度强化学习 [J]. 水下无人系统学报, 2025, 33(2): 194-203.

无人艇集群路径规划研究综述: 深度强化学习

侯玉立[1], 王 宁[2*], 邱赤东[1], 翁永鹏[1]

(1. 大连海事大学 船舶电气工程学院, 辽宁 大连, 116026; 2. 大连海事大学 轮机工程学院, 辽宁 大连, 116026)

摘 要: 无人艇(USV)集群在复杂海洋任务中展现出显著优势, 但其路径规划面临高维、动态以及多约束等挑战。传统路径规划算法因协同机制薄弱与适应性不足, 难以满足日渐复杂的需求, 而深度强化学习(DRL)技术的发展为 USV 集群路径规划提供了新的研究方向。文中系统综述了基于 DRL 的 USV 集群协同路径规划技术框架及典型算法。首先, 梳理了 USV 集群路径规划的技术演进脉络与多维约束条件, 分析了集中式和分布式决策框架的适用场景与局限性; 其次, 探讨了多种典型 DRL 算法的原理、应用场景及改进方向, 分析了其优势与不足; 最后, 总结了该领域面临的主要挑战和发展方向, 旨在为基于 DRL 的 USV 集群协同路径规划研究提供参考。

关键词: 无人艇集群; 协同路径规划; 深度强化学习

中图分类号: TJ630.32; U674.941 **文献标识码:** R **文章编号:** 2096-3920(2025)02-0194-10

DOI: 10.11993/j.issn.2096-3920.2025-0034

A Review of Research on Path Planning of Unmanned Surface Vessel Swarm: Deep Reinforcement Learning

HOU Yuli[1], *WANG Ning*[2*], *QIU Chidong*[1], *WENG Yongpeng*[1]

(1. Marine Electrical Engineering College, Dalian Maritime University, Dalian 116026, China; 2. Marine Engineering College, Dalian Maritime University, Dalian 116026, China)

Abstract: An unmanned surface vessel(USV) swarm has shown significant advantages in complex marine missions, but its path planning faces high-dimensional, dynamic, and multi-constraint challenges. Traditional path planning algorithms are difficult to meet increasingly complex needs due to weak coordination mechanisms and insufficient adaptability, while the development of deep reinforcement learning(DRL) technology provides a new research direction for the path planning of USV swarms. This paper systematically reviewed the technical framework and typical algorithms for collaborative path planning of USV swarms based on DRL. Firstly, the technical evolution context and multi-dimensional constraints of path planning of USV swarms were sorted out, and the applicable scenarios and limitations of centralized and distributed decision frameworks were analyzed. Secondly, the principle, application scenarios, and improvement directions of various typical DRL algorithms were discussed, and their advantages and disadvantages were analyzed. Finally, the main challenges and development directions in this field were summarized. This paper aims to provide a reference for the research on DRL-based collaborative path planning of USV swarms.

Keywords: unmanned surface vessel swarm; collaborative path planning; deep reinforcement learning

收稿日期: 2025-02-27; **修回日期:** 2025-03-14; **录用日期:** 2025-03-18.

基金项目: 国家自然科学基金项目(U23A20680, 52271306); 国家拔尖人才专项支持计划项目(SQ2022QB00329); 辽宁省领军人才项目(XLYC2202005); 大连市科技创新基金重大基础研究项目(2023JJ11CG009); 中央高校基本科研业务费专项资金资助(3132023501).

作者简介: 侯玉立(1998-), 男, 在读硕士, 主要研究方向为无人艇集群路径规划技术.

*** 通信作者简介:** 王 宁(1983-), 男, 博士, 教授, 主要研究方向为智能海洋机器人、绿色智能船舶及海洋人工智能.

0　引言

与有人船相比, 无人艇(unmanned surface vessel, USV)凭借其安全性高、机动性强及使用成本低的优势, 近年来在军事作战[1]、科学研究[2]、物流运输[3]以及灾害救援[4]等领域得到了广泛的应用。随着任务复杂度的增加, 单 USV 的作业能力出现瓶颈。由多个 USV 根据特定规则共同完成任务的 USV 集群可以通过协同规划实现资源整合与效率提升, 成为突破性能瓶颈的关键技术[5-7]。

路径规划是实现自主导航、避障和目标追踪等功能的关键技术, 旨在通过算法确定 USV 从起点到终点的无碰最优路径。早期的路径规划方法主要针对路径长短进行优化, 但在存在风浪和洋流等复杂扰动的海洋环境中, USV 的最短距离路径不一定是耗时最少或能耗最低的路径。如果再考虑路径的平滑性和算法的运行速度, USV 的最优路径规划无疑是一个需要综合考虑多种评价指标的多目标优化问题。当路径规划的对象从单 USV 扩展到 USV 集群时, 规划的路径不仅要避免多目标冲突, 还需要统筹协调各 USV, 满足各种协同约束, 以实现超越单纯增加 USV 数量的协同效果, 这对路径规划算法提出了更高的要求。传统路径规划算法协同机制薄弱, 依赖目标分配-路径规划的双层框架实现协同, 在应对复杂环境变化时适应能力不足[8-10]。

随着人工智能技术的发展, 基于深度强化学习(deep reinforcement learning, DRL)的智能化路径规划方法为 USV 集群协同路径规划问题带来了新的解决思路。基于 DRL 的路径规划方法能够通过试错训练获取适应复杂环境并满足协同规则的策略, 展现出了巨大的潜力。虽然目前已经有学者对 USV 制导控制技术进行了总结和分类[11-13], 但是针对 USV 集群协同路径规划的专项总结仍然较少。为进一步提高 USV 集群协同路径规划的智能性与规划效率, 探索更先进的路径规划技术, 有必要对现有技术进行梳理, 对技术体系进行分类, 分析现有技术的优势与不足, 从而为 USV 集群协同路径规划技术的进一步发展提供参考。

为此, 文中首先梳理了 USV 集群路径规划技术的发展背景, 介绍了协同路径规划技术的演进过程; 进而重点关注基于 DRL 的智能化协同路径规划方法, 梳理了现有的技术体系, 介绍了当前常用的几种典型算法的最新进展; 最后, 结合现有技术进展和挑战, 探讨了未来研究方向。

1　USV 集群路径规划技术背景

USV 集群协同路径规划的目标是在考虑 USV 模型、环境、任务与协同等多维约束的前提下, 为 USV 集群规划最优路径, 其本质是一个高维、非线性且强耦合的多目标优化问题。从 USV 集群协同路径规划约束条件与技术演进脉络 2 个维度出发, 梳理 USV 集群路径规划技术发展背景。

1.1　USV 集群协同路径规划约束

USV 集群协同路径规划的约束条件可归纳为 USV 自身约束、环境约束、任务约束和时空间协同约束 4 类, 如图 1 所示。其中, USV 自身约束不仅包含动力系统限制带来的运动学和动力学约束(如最大航速、最小转弯半径和加速度阈值等), 还涉及通信系统的带宽、传输延迟和抗干扰能力等通信能力限制, 以及能源系统的容量、消耗速率等续航能力制约。环境约束方面, 除传统静态障碍物(如岛礁、禁航区等)和动态障碍物(如移动船舶、海上平台等)的空间分布外, 还需考虑水文气象条件(如风浪流扰动、能见度等)对航行稳定性的影响。对于不同的任务类别, 也存在各自的特殊约束, 如覆盖搜寻类任务中的作业精度约束、追踪拦截类

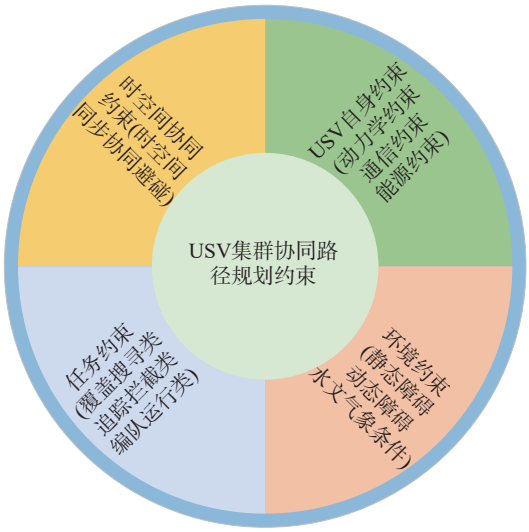

图 1　USV 集群协同路径规划约束图

Fig. 1　Collaborative path planning constraints of USV swarm

任务中的实时性指标、编队运行类任务中的队形保持指标等。时空间协同约束则表现为 USV 间的时空同步(如 USV 编队的队形同步、拦截任务中的时空间同步等)以及与时间耦合的协同避碰规则(如 USV 优先级、安全距离保持等)等综合性约束。

各种约束之间相互耦合, 形成了复杂的非线性耦合系统, 而且约束的增加和耦合会使可行解空间呈指数级收缩, 进而易导致优化算法陷入局部最优。针对这一难点, 目前的研究大多从任务约束出发, 将整个路径规划问题分解为多个子问题, 再针对子问题中需要考虑的特殊约束进行分层优化。例如: 针对覆盖搜寻类任务, 可以分解为区域/目标分配、USV 动态避障路径规划等子任务[9, 14-15]; 针对追踪拦截类任务, 可以分解为目标运动预测、威胁评估和协同规则等[16-17]; 针对编队运行类任务, 可以将协同路径规划问题分解为编队设计、队形保持与重构等[18-19]。

1.2 USV 协同路径规划算法演进过程

面向不同的任务需求与约束条件, USV 集群协同路径规划算法经历了从基于几何规则的传统算法到群体智能优化算法、神经网络优化算法再向 DRL 驱动的人工智能算法演进的 3 个阶段。

基于几何规则的传统算法(如 A^*、D^*算法及人工势场法等)主要针对静/动态障碍约束下的路径长度优化问题。其中, A^*算法虽能通过全局信息计算最优路径, 但在解决需要动态避障的 USV 集群路径规划问题时计算效率低, 难以及时处理 USV 集群运行过程中的突发情况。尽管通过划分各 USV 通行优先级[20] 及引入时间约束[21] 进行重规划可以在理论上解决 USV 集群内的避碰问题, 但其重规划耗时特性仍严重制约实际部署。相比 A^*算法, 人工势场法不依赖全局信息, 可以根据障碍物信息及时调整期望路径, 具有更强的实时性, 更适用于 USV 集群[18, 22]。但是, 人工势场法也存在易陷入局部最优的缺陷, 因此常与全局路径规划算法配合使用[23-24]。虽然通过路径平滑算法可以使传统算法规划的路径满足 USV 基本动力学约束[20], 但传统算法在解决需要考虑更多约束的多目标协同优化问题时仍存在技术瓶颈。

针对多目标优化问题, 受生物群体行为启发的智能优化算法(如遗传算法、粒子群算法以及蚁群算法等)展现出更高维度的优化能力[25-26]。此类算法通过群体协同搜索机制, 可处理包含能耗、时间和安全距离等多约束的复杂优化问题, 灵活性更高。然而在高维、动态和复杂海洋环境中, 该类算法仍面临收敛速度滞后、实时性不足和难以应对不确定性等挑战[15, 27]。此外, 基于神经网络架构的智能优化算法也为复杂时变环境下的路径规划提供了新思路[28-29]。该方法通过将环境信息映射到神经网络中, 并根据任务指标动态调整神经元之间的激励, 建立任务需求与神经元激励强度的动态耦合机制, 从而动态生成最优路径。然而, 此类方法仍存在对预设激励的依赖性。

针对上述算法存在的问题, DRL 技术通过环境交互式学习机制, 展现出了传统方法难以企及的优势。基于 DRL 的路径规划技术具有自适应性强、适用于多目标优化问题等优势, 能够显著提升路径规划的效率和鲁棒性[30-31]。DRL 可以实现从原始传感器数据输入到控制命令输出的端到端感知-决策框架, 降低算法对高精度先验地图的依赖, 从而使 USV 可以更好地应对环境的不确定性, 提高任务完成效率[32]。

2 DRL 路径规划技术框架

从优化机制层面分析, DRL 通过“感知-决策-奖励-优化”的闭环机制优化智能体的路径规划策略, 为解决 USV 集群协同路径规划问题提供了系统的技术框架。从路径规划系统架构维度分析, 完全去中心化的分散式决策框架受限于局部环境感知能力, 难以实现多 USV 间的任务协同。鉴于此, 当前主流 USV 集群协同路径规划技术主要呈现为 2 种范式: 基于全局信息的集中式决策框架和兼顾自主性与协调性的分布式决策框架。表 1 对比了 2 种决策框架在不同维度的特点。

2.1 集中式决策框架

如图 2 所示, 集中式决策框架使用单一智能体为 USV 集群规划路径, 通过全局状态观测实现协同路径规划[33-34]。这一框架的优势在于能够直接优化全局目标函数, 获得全局最优策略。由于需要一个中心式的路径决策器获取全局状态并分发各 USV 的决策, 这一技术框架对 USV 集群的通信带宽和延迟提出了较高要求。同时, USV 集群中的

表 1　集中式与分布式决策框架特点对比

Table 1　Comparison of features between centralized and distributed decision-making frameworks

对比维度	集中式	分布式
可靠性	单一中心节点统一决策, 故障风险高	多节点自主决策, 容错性高
扩展性	扩展困难, 需重构中心架构	扩展灵活, 通过增加节点实现扩展
通信需求	各USV与中心节点频繁交互, 需具有足够的通信带宽	节点间通信, 需协调调度
优化能力	基于全局状态可获取全局最优解	基于局部状态获取局部最优解
资源消耗	中心节点计算、存储压力较大	计算分散至各节点, 负载均衡

各 USV 状态相互耦合, 随着 USV 数量的增加, 状态和动作组合数的规模将呈指数增长。这种增长会使单智能体的计算复杂度膨胀到超出可行范围, 导致“维度爆炸”。因此, 集中式决策框架并不适用于大规模 USV 集群。

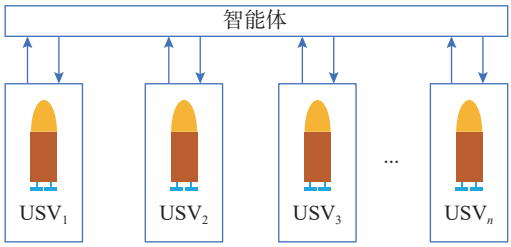

图 2　集中式决策框架

Fig. 2　Centralized decision framework

从理论层面看, 基于集中式决策框架的单智能体 DRL 算法通过合理构建状态空间、动作空间及奖励函数机制, 可扩展至小规模 USV 集群的协同路径规划场景。具体而言, Zhao 等[33] 基于编队速度同步误差与位置误差构建了适用于编队运行任务的奖励函数, 采用深度确定性策略梯度(deep deterministic policy gradient, DDPG)算法实现了 USV 动力输出策略的优化; Luis 等[34] 针对同构 USV 集群协同巡逻任务特性, 设计了可扩展的多头集中式深度 Q 网络(deep Q-network, DQN)算法, 并在多头 Q 网络中嵌入卷积神经网络(convolutional neural networks, CNN)模块作为全局状态提取器, 相比分散式的独立 DQN 算法训练速度得到大幅提升。

2.2　分布式自主决策框架

如图 3 所示, 分布式自主决策框架的核心特征在于各 USV 配备独立智能体, 通过局部观测与邻域通信交互实现自组织协同路径规划。从系统特性维度来看, 分布式决策框架赋予了 USV 自主决策能力, 有效降低了 USV 对中心节点的依赖性, 更适应实际海洋场景中通信受限的作业条件, 在可扩展性方面也展现出显著优势。然而, 受限于局部信息处理机制, 该框架难以严格保证决策的全局最优性。

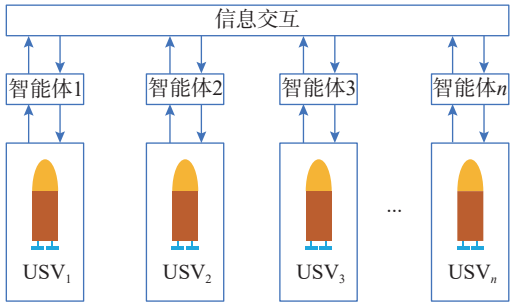

图 3　分布式决策框架

Fig. 3　Distributed decision framework

值得注意的是, 区别于本质上仍属于单 USV 路径优化范畴的分散式决策框架, 分布式决策框架虽不需要全局通信支持, 但仍需设计合适的协同通信机制, 以实现 USV 间的状态估计与协同决策优化[35-36]。

当前基于 DRL 的分布式决策框架主要采用“集中训练、分布执行”(centralized training with decentralized execution, CTDE)的多智能体(multi-agent, MA)DRL 算法典型架构, 并通过任务层级分解、奖励函数设计和算法结构优化等方式提升最优策略的收敛效率与稳定性。根据面向的任务, 表 2 总结了现有基于 DRL 的 USV 集群协同路径规划研究文献。

表 2　集中式与分布式决策框架应用

Table 2　Centralized and distributed decision framework applications

决策框架	编队运行	覆盖搜寻	追踪拦截
集中式	[33]	[34]	—
分布式	[37-42]	[36][43-44]	[17][45-48]

3　DRL 路径规划典型算法

根据环境模型依赖性, DRL 算法可分为基于模型与无模型两大范式。鉴于 USV 集群协同路径

规划任务的环境建模存在显著复杂性, 当前研究主要聚焦于无需先验模型、通过试错学习实现策略优化的无模型 DRL 算法。根据算法的优化目标是价值函数(Critic)还是策略函数(Actor), 现有算法可进一步分为基于价值的方法、基于策略梯度的方法以及基于 Actor-Critic 架构的方法三类。其中, 基于价值的算法(如 DQN 等)通过价值迭代更新 Actor 网络参数, 收敛速度和稳定性不高, 对超参数敏感且只适用于离散动作空间[34]。直接优化 Actor 网络的基于策略梯度的方法(如近端策略优化(proximal policy optimization, PPO)算法等)能够有效处理连续动作空间, 基于策略本身的随机性对环境进行探索, 探索能力较强, 为了平衡新旧策略更新设计的“裁剪机制”提高了算法的稳定性[46-47]。然而, 在线策略更新导致 PPO 对样本的利用效率不高, “裁剪机制”的存在也导致其收敛速度较慢。基于 Actor-Critic 架构的算法(如 DDPG、柔性 Actor-Critic 算法(soft actor-critic, SAC)等)通过 Critic 网络评估动作并指导 Actor 网络的参数更新, 平衡了策略稳定性与收敛速度, 离线策略更新机制可以实现对样本经验的高效复用。然而该类算法也存在对超参数敏感及计算成本高等不足[38-39, 42]。为清晰地突出各算法的特点, 表 3 从 DQN、PPO、DDPG、SAC 及它们基于 CTDE 的 MA 变体的维度对现有 USV 集群协同路径规划研究结果、特点及适用场景进行了分类总结。由于 DQN 局限性较大, 相关研究较少, 文中重点围绕 PPO、DDPG、SAC 三类典型算法展开体系化论述。

表 3　基于不同 DRL 典型算法的 USV 集群路径规划特点

Table 3　Characteristics of USV swarm path planning based on different typical DRL algorithms

基线算法	应用	收敛速度	稳定性	样本效率	适用场景
DQN	[34]	中等	中等	中等	只适用于USV集群离散决策场景, 如基于栅格化地图的协同搜寻类任务
PPO	[37][43][46][47]	较慢	较高	较低	适用于动态环境中的拦截与编队运行类任务
DDPG	[17][33][36][40-41][45][48]	较快	中等	较高	适用于同构USV集群编队运行类任务
SAC	[42][44]	较快	较高	较高	适用于复杂动态环境中的编队运行类任务

3.1　DDPG

DDPG 是一种属于 Actor-Critic 架构的无模型离线策略算法, 通过引入经验回放机制提升数据利用效率, 通过目标网络缓解状态动作值(Q 值)的估计误差。基于 CTDE 范式的 MADDPG 算法[49]因具有实现简单且支持连续动作空间输出的特性, 在 USV 集群路径决策研究领域受到了广泛关注。如图 4 所示, 图中橙色部分为集中式训练, 蓝色部分为分布式执行。在 MADDPG 算法中, 每个 USV 对应的智能体有自己的 Actor 网络和 Critic 网络, 利用经验回放缓冲区的历史数据, Critic 网络可以访问其他智能体的状态信息形成全局状态进行集中式训练, 而 Actor 网络仅需使用 USV 自身局部观察信息, 做出路径决策动作, 实现分布式执行。

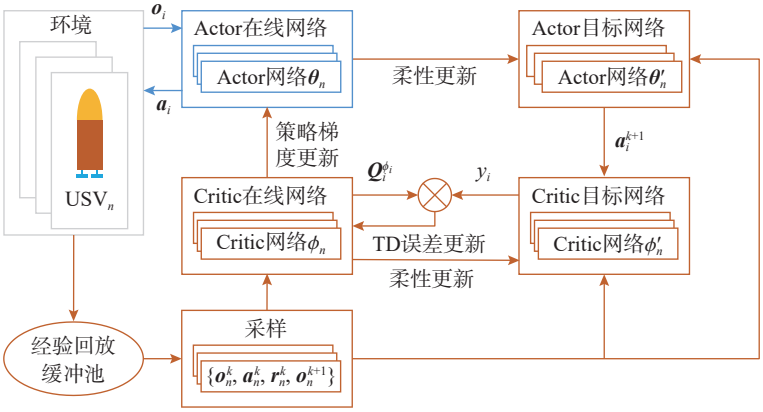

图 4　MADDPG 算法

Fig. 4　MADDPG algorithm

在USV总数为n的任务中，USV$_i$智能体的Actor在线网络参数为$\boldsymbol{\theta}_i$, 目标网络参数为$\boldsymbol{\theta}'_i$, Critic在线网络参数为ϕ_i, 目标网络参数为ϕ'_i。Critic在线网络的目标是最小化时序差分(temporal difference, TD)误差, 从而准确估计Q值, 其损失函数为

$$L(\phi_i)=\mathbb{E}\left[y_i^k-Q_i^{\phi_i}\left(\boldsymbol{s}^k,\boldsymbol{a}_1^k,\boldsymbol{a}_2^k,\cdots,\boldsymbol{a}_n^k\right)\right]^2 \tag{1}$$

式中: k为第k个时间步; $Q_i^{\phi_i}$为Critic在线网络输出的状态动作值; $\boldsymbol{s}$为全局状态信息, 为了简化状态空间设计, 通常由各USV的局部观察信息$\boldsymbol{o}_i$叠加组成; $\boldsymbol{a}_i(i=1,\cdots,n)$为USV$_i$的策略动作; y_i^k为目标状态动作值, 由Critic目标网络输出的状态动作值$Q_i^{\phi'_i}$和通过采样获取的奖励值r_i组成, 即

$$y_i^k=r_i^k+\gamma Q_i^{\phi'_i}\left(\boldsymbol{s}^{k+1},\boldsymbol{a}_1^{k+1},\boldsymbol{a}_2^{k+1},\cdots,\boldsymbol{a}_n^{k+1}\right) \tag{2}$$

式中: $\boldsymbol{a}_i^{k+1}=f_{\theta' i}\left(\boldsymbol{o}_i^{k+1}\right)$为Actor目标网络的输出; γ为折扣系数。Actor在线网络的目标是最大化Q值的期望, 即选择能产生最大Q值的动作, 通过策略梯度更新参数, 损失函数为

$$L(\boldsymbol{\theta}_i)=-\mathbb{E}\left[Q_i^{\phi_i}\left(\boldsymbol{s}^k,\boldsymbol{a}_1^k,\boldsymbol{a}_2^k,\cdots,\boldsymbol{a}_n^k\right)\right] \tag{3}$$

式中, $\boldsymbol{a}_i^k=f_{\theta i}\left(\boldsymbol{o}_i^k\right)$为Actor在线网络的输出。通过梯度下降法可以实现Actor在线网络和Critic在线网络的更新, 即

$$\boldsymbol{\theta}_i\leftarrow\boldsymbol{\theta}_i-\lambda_{\theta i}\nabla_{\theta i}L(\boldsymbol{\theta}_i) \tag{4}$$

$$\phi_i\leftarrow\phi_i-\lambda_{\phi i}\nabla_{\phi i}L(\phi_i) \tag{5}$$

式中, $\lambda_{\theta i}$和$\lambda_{\phi i}$为学习率。Actor目标网络和Critic目标网络通过柔性更新法更新参数, 即

$$\boldsymbol{\theta}'_i\leftarrow\tau\boldsymbol{\theta}_i+(1-\tau)\boldsymbol{\theta}'_i \tag{6}$$

$$\phi'_i\leftarrow\tau\phi_i+(1-\tau)\phi'_i \tag{7}$$

其中, τ为一个较小的常数。

对于USV集群编队运行类任务, 不同于传统算法优先在运行过程中保持严格的编队构型, 基于MADDPG的决策框架可以通过设计复合奖励函数, 在满足集群行为“聚集、分离、速度一致”三原则的前提下[50], 允许个体USV通过偏离策略实现碰撞规避, 同时保持编队宏观形态稳定性。这种具有容错特征的编队模式被称为柔性编队结构, 常通过领航者机制校准编队的参考位置[38-41]。对于追踪拦截类任务, 于长东等[45]的研究表明, 可通过围捕半径与夹角的量化建模构建奖励函数获取围捕策略, 但是文中逃逸USV与围捕USV并未放在不同算法框架下进行训练, 策略的实用性不强。Song团队[17, 48]提出对抗进化训练框架, 通过为追逃双方分别设计DRL策略并加入信用分配机制解决协同中的贡献度问题, 具有更高的效率和泛化能力。目前, MADDPG算法在探索阶段需要通过人为设计的噪声扰动来改变动作输出, 噪声的大小和持续时间依赖人员经验, 通常需要多次尝试调整, 这导致探索与利用的均衡性难以有效控制。虽然可以利用贪婪策略在一定程度上平衡探索与利用[40], 但在高复杂度的动态环境中, MADDPG仍难以获得符合多重约束条件的最优策略[51]。基于此特性, 目前MADDPG主要应用于同构USV集群, 以规避异构系统复杂的动力学特性引发的收敛效率问题。

3.2　PPO

PPO是一种基于策略梯度的DRL算法, 其核心思想是通过限制策略更新的幅度来平衡探索与利用, 从而提升训练稳定性[52]。PPO通过引入裁剪机制约束策略更新步长, 避免了因策略突变导致的训练崩溃问题, 使其适用于解决高维、动态问题。PPO是一种在线策略算法, 使用最新一批的数据进行更新, 目标函数为

$$L^{\text{clip}}(\boldsymbol{\theta})=\mathbb{E}\left[\min\left(\begin{array}{l}\rho^k(\boldsymbol{\theta})A^k,\\ \text{clip}\left(\begin{array}{l}\rho^k(\boldsymbol{\theta}),\\ 1-\epsilon,1+\epsilon\end{array}\right)A^k\end{array}\right)\right] \tag{8}$$

式中: $\rho^k(\boldsymbol{\theta})=f_\theta\left(\boldsymbol{a}^k|\boldsymbol{o}^k\right)\Big/f_{\theta'}\left(\boldsymbol{a}^k|\boldsymbol{o}^k\right)$为新旧策略在同一状态下选择动作的概率比, $f_\theta(\cdot)$和$f_{\theta'}(\cdot)$分别为新、旧策略网络函数; A^k为优势函数, 通常由广义优势估计的方法近似; ϵ为一个超参数; clip为裁剪算子, 用来将$\rho^k(\boldsymbol{\theta})$限制在区间$[1-\epsilon,1+\epsilon]$内, 防止新旧策略变化过大。

在USV集群协同路径规划中, PPO和基于CTDE范式的MAPPO凭借其超参数鲁棒性强的优势, 在USV集群协同路径决策场景中展现出广泛适用性。该类算法在编队运行与追踪拦截类任务中均展现出良好的工程适用性[43, 47]。针对算法优化路径, 现有研究主要沿2个技术路线推进: 融合传统路径规划算法的先验知识构建引导机制提

高探索效率[53] 以及改进神经网络架构提高信息处理能力[47]。例如, Li 等[46] 将速度障碍法引入奖励函数指导 USV 集群躲避障碍物, 同时采用双向门控循环单元实现变长观测序列的固定维度特征编码, 通过课程学习策略实现从稀疏奖励场景到密集干扰场景的渐进式训练, 加速策略收敛。Xia 等[47] 则提出了一种特征嵌入块, 通过列最大池化和列平均池化压缩观测维度, 提升了网络对输入变化的鲁棒性。尽管 PPO 在动态场景中表现优异, 但其在线学习机制存在对训练数据质量和数量的强依赖性, 且难以复用过往经验数据。在多种仿真任务场景中, PPO 都需要近百万步的交互训练才能获得稳定策略[43, 47], 这一特性制约了 PPO 算法在现实场景中的应用可行性。

3.3　SAC

SAC 是 DDPG 的改进版本, 其核心思想是通过最大化策略的熵来增强探索能力, 适用于复杂动态海洋环境中 USV 集群的协同路径规划。SAC 通过联合优化策略熵与累计奖励, 鼓励智能体探索尝试多样化动作, 避免陷入局部最优。

与 DDPG 不同, 基于 SAC 的智能体只需维护 1 个 Actor 网络, 该网络的基本结构为如图 5 所示的双头输出神经网络, 输入观察信息$\boldsymbol{o}^k$后同时输出动作$\boldsymbol{a}^k$以及选择该动作的概率$\boldsymbol{\mu}^k\left(\boldsymbol{a}^k|\boldsymbol{o}^k\right)$, 其损失函数为

$$L(\boldsymbol{\theta})=\mathbb{E}\left[\alpha\log\boldsymbol{\mu}^k\left(\boldsymbol{a}^k|\boldsymbol{o}^k\right)-Q^{\phi}\left(\boldsymbol{o}^k,\boldsymbol{a}^k\right)\right] \tag{9}$$

式中: α为熵温度系数; $\boldsymbol{o}^k$通过经验回放缓冲池采样获取; $\boldsymbol{a}^k$和$\boldsymbol{\mu}^k\left(\boldsymbol{a}^k|\boldsymbol{o}^k\right)$通过 Actor 网络获取。Critic 网络损失函数为

$$L(\phi)=\mathbb{E}\left[y^k-Q^{\phi}\left(\boldsymbol{o}^k,\boldsymbol{a}^k\right)\right]^2 \tag{10}$$

其中, $\boldsymbol{a}^k$由 Actor 网络输出;

$$y^k=r^k+\gamma[Q^{\phi'}(\boldsymbol{o}^{k+1},\boldsymbol{a}^{k+1})-\alpha\log\boldsymbol{\mu}^k(\boldsymbol{a}^{k+1}|\boldsymbol{o}^{k+1})] \tag{11}$$

通过梯度下降法和柔性更新法即可更新 Actor 网络和 Critic 网络。

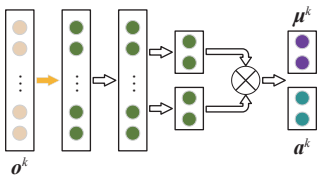

图 5　SAC 基本 Actor 网络结构

Fig. 5　Basic actor network structure of SAC

SAC 通过最大熵优化机制平衡智能体行为随机性, 减少无效采样, 显著提升了智能体在复杂动态环境中的探索效率。得益于 SAC 的深度探索特性, USV 集群可以实现多约束条件下的编队重构优化[42] 与动态避障策略生成[44], 具有较高的扩展潜力。例如, 针对存在编队构型约束与外部环境扰动的编队运行任务, Jin 等[42] 通过将任务拆解为目标追踪、动态避障以及队形保持等子任务, 利用 SAC 算法同时优化各 USV 面向多个任务的策略, 并在训练过程中融入策略共享机制, 使得 USV 在部分可观测条件下自主平衡各任务优先级, 实现复杂环境下的编队运行。相应地, 这需要构建包括拓扑关系、运动学约束和环境耦合特征的多维奖励函数体系, 对奖励函数的设计和网络的数据处理能力提出了较高的要求。Yao 等[44] 则充分发挥 SAC 算法的探索性强优势, 不直接生成路径规划策略, 而是对人工势场法的关键参数进行动态调节, 使这一传统算法能自适应复杂海洋环境。然而, 由于该方法的策略框架为人工势场法的固有框架, 制约了 USV 集群的自主性。

4　研究挑战与未来方向

4.1　研究挑战

基于 DRL 的 USV 集群协同路径规划技术虽然展现出巨大潜力, 但目前的大多数研究仍处于虚拟仿真阶段, 相关结果无法被信任用于实船航行, 研究不充分、不全面, 且面临以下关键挑战。

1) 复杂动态环境适应性不足: 现有 DRL 算法在强扰动和不确定性高的海洋环境中的决策能力有限。多数研究通过简化环境模型进行仿真训练, 导致策略模型在实际部署时鲁棒性不佳。

2) 多目标协同约束耦合: USV 集群协同路径规划需满足多维约束, 容易使设计的 DRL 奖励函数陷入多目标冲突, 引发策略震荡。

3) 数据依赖性强: DRL 依赖大量数据训练, 但海洋场景数据获取成本高, 且仿真/现实差异易导致策略失效。数据差异小也容易使策略网络过拟合, 导致算法泛化与扩展性差。

4.2　未来方向

为突破上述挑战, 可从以下方向展开研究。

1) 环境感知与决策融合增强: 开发多模态传感器感知与路径决策的融合框架, 结合传感器数据与物理模型构建 USV 与环境的交互关系, 形成面向动态复杂环境的路径决策策略。

2) 多目标奖励函数优化: 设计动态奖励分配方法, 通过自适应权重平衡各约束指标。开发基于元强化学习的自适应奖励调节框架, 通过元策略动态调整多目标权重系数。探索分层强化学习架构, 将高层策略用于约束优先级决策, 底层策略专注局部路径优化等。

3) 虚实结合训练与验证平台: 建立开放训练数据集, 开发高保真的海洋环境数字孪生系统, 集成海洋环境动力学模型与 DRL 训练测试接口, 实现仿真-测试闭环优化。

5　结束语

USV 集群协同路径规划是当前 USV 集群领域的前沿研究课题。文中从技术发展背景出发, 梳理了协同路径规划所面临的主要难题及其技术演进脉络。随后, 对基于 DRL 的技术框架进行了分类阐述, 介绍了 3 种典型算法的基本原理及其相关应用, 揭示了 DRL 在解决 USV 集群协同路径优化问题中的巨大潜力。最后, 总结了现有研究中存在的挑战, 并指出了未来亟待突破的关键方向, 以期推动 USV 集群协同路径规划技术向更高水平的自主化与智能化发展。

参考文献:

[1] 孙峰. 一种基于海空无人集群的自杀式无人艇防御策略[J]. 水下无人系统学报, 2024, 32(2): 267-274, 319.
SUN F. Defense strategy for suicide unmanned surface vessels based on sea and air unmanned clusters[J]. Journal of Unmanned Undersea Systems, 2024, 32(2): 267-274, 319.

[2] 翁磊, 杨扬, 钟雨轩. 多无人艇协同遍历路径规划算法[J]. 水下无人系统学报, 2020, 28(6): 634-641.
WENG L, YANG Y, ZHONG Y X. Collaborative traversal path planning algorithm of for multiple unmanned survey vessels[J]. Journal of Unmanned Undersea Systems, 2020, 28(6): 634-641.

[3] 王宁, 刘永金, 高颖. 未知扰动下的无人艇编队优化轨迹跟踪控制[J]. 中国舰船研究, 2024, 19(1): 178-190.
WANG N, LIU Y J, GAO Y. Optimal trajectory tracking control of unmanned surface vehicle formation under unknown disturbances[J]. Chinese Journal of Ship Research, 2024, 19(1): 178-190.

[4] 王秀玲, 尹勇, 赵延杰, 等. 无人艇海上搜救路径规划技术综述[J]. 船舶工程, 2023, 45(4): 50-57.
WANG X L, YIN Y, ZHAO Y J, et al. Overview of USV maritime search and rescue path planning technology[J]. Ship Engineering, 2023, 45(4): 50-57.

[5] 焦宇航, 王宁. 欠驱动无人船集群有限时间跟踪控制[J]. 中国舰船研究, 2023, 18(6): 76-87.
JIAO Y H, WANG N. Finite-time trajectory tracking control of underactuated surface vehicles swarm[J]. Chinese Journal of Ship Research, 2023, 18(6): 76-87.

[6] WANG N, HE H, HOU Y, et al. Model-free visual servo swarming of manned-unmanned surface vehicles with visibility maintenance and collision avoidance[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(1): 697-709.

[7] WANG N, LIU Y, LIU J, et al. Reinforcement learning swarm of self-organizing unmanned surface vehicles with unavailable dynamics[J]. Ocean Engineering, 2023, 289: 116313.

[8] NIU Y, MU Y, ZHANG K, et al. Path planning and search effectiveness of USV based on underwater target scattering model[J]. Journal of Physics: Conference Series, 2023, 2478(10): 102035.

[9] MA Y, ZHAO Y, LI Z, et al. CCIBA*: An improved BA* based collaborative coverage path planning method for multiple unmanned surface mapping vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(10): 19578-88.

[10] XUE K, HUANG Z, WANG P, et al. An exact algorithm for task allocation of multiple unmanned surface vehicles with minimum task time[J]. Journal of Marine Science and Engineering, 2021, 9(8): 907.

[11] 刘祥, 叶晓明, 王泉斌, 等. 无人水面艇局部路径规划算法研究综述[J]. 中国舰船研究, 2021, 16(z1): 1-10.
LIU X, YE X M, WANG Q B, et al. Review on the research of local path planning algorithms for unmanned surface vehicles[J]. Chinese Journal of Ship Research, 2021, 16(z1): 1-10.

[12] LIN X, LIU Y. Research on multi-USV cooperative search method[C]//2019 IEEE International Conference on Mechatronics and Automation. Tianjin, China: IEEE, 2019.

[13] 徐善文, 曾庆化, 李方东, 等. 无人集群系统协同导航资源及算法综述[J]. 导航与控制, 2024, 23(5): 25-37.
XU S W, ZENG Q H, LI F D, et al. A review of cooperative navigation resources and algorithms for unmanned swarm systems[J]. Navigation and Control, 2024, 23(5): 25-37.

[14] WANG H, FU Z, ZHOU J, et al. Cooperative collision avoidance for unmanned surface vehicles based on improved genetic algorithm[J]. Ocean Engineering, 2021, 222: 108612.

[15] ZHAO L, BAI Y, PAIK J K. Global path planning and waypoint following for heterogeneous unmanned surface vehicles assisting inland water monitoring[J]. Journal of Ocean Engineering and Science, 2023, 10(1): 88-108.

[16] MENG X, SUN B, ZHU D. Harbour protection: Moving invasion target interception for multi-AUV based on prediction planning interception method[J]. Ocean Engineering, 2021, 219: 108268.

[17] GAN W, QU X, SONG D, et al. Multi-USV cooperative chasing strategy based on obstacles assistance and deep reinforcement learning[J]. IEEE Transactions on Automation Science and Engineering, 2023, 21(4): 5895-910.

[18] YAN X, JIANG D, MIAO R, et al. Formation control and obstacle avoidance algorithm of a multi-USV system based on virtual structure and artificial potential field[J]. Journal of Marine Science and Engineering, 2021, 9(2): 161.

[19] 欧阳子路, 王鸿东, 黄一, 等. 基于改进 RRT 算法的无人艇编队路径规划技术[J]. 中国舰船研究, 2020, 15(3): 18-24.
OUYANG Z L, WANG H D, HUANG Y, et al. Path planning technologies for USV formation based on improved RRT[J]. Chinese Journal of Ship Research, 2020, 15(3): 18-24.

[20] LI Y, ZHANG J, LI Y, et al. Research on the frame of formation of multi-USV[C]//2022 5th World Conference on Mechanical Engineering and Intelligent Manufacturing(WCMEIM). Ma'anshan, China: IEEE, 2022: 746-749.

[21] SANG T, XIAO J, XIONG J, et al. Path planning method of unmanned surface vehicles formation based on improved A* algorithm[J]. Journal of Marine Science and Engineering, 2023, 11(1): 176.

[22] 宋利飞, 徐凯凯, 史晓骞, 等. 多无人艇协同围捕智能逃跑目标方法研究[J]. 中国舰船研究, 2023, 18(1): 52-59.
SONG L F, XU K K, SHI X Q, et al. Multiple USV cooperative algorithm method for hunting intelligent escaped targets[J]. Chinese Journal of Ship Research, 2023, 18(1): 52-59.

[23] SANG H, YOU Y, SUN X, et al. The hybrid path planning algorithm based on improved A* and artificial potential field for unmanned surface vehicle formations[J]. Ocean Engineering, 2021, 223: 108709.

[24] YU J, CHEN Z, ZHAO Z, et al. A traversal multi-target path planning method for multi-unmanned surface vessels in space-varying ocean current[J]. Ocean Engineering, 2023, 278: 114423.

[25] SHARMA A, SHOVAL S, SHARMA A, et al. Path planning for multiple targets interception by the swarm of UAVs based on swarm intelligence algorithms: A review [J]. IETE Technical Review, 2022, 39(3): 675-697.

[26] NAZARAHARI M, KHANMIRZA E, DOOSTIE S. Multi-objective multi-robot path planning in continuous environment using an enhanced genetic algorithm[J]. Expert Systems with Applications, 2019, 115: 106-120.

[27] LUO Q, YAN X, WU D, et al. Unmanned surface vehicle cooperative task assignment based on genetic algorithm [C]//2022 Global Reliability and Prognostics and Health Management. Yantai, China: IEEE, 2022: 1-5.

[28] YAO P, WU K, LOU Y. Path planning for multiple unmanned surface vehicles using Glasius bio-inspired neural network with Hungarian algorithm[J]. IEEE Systems Journal, 2022, 17(3): 3906-17.

[29] TANG F. Coverage path planning of unmanned surface vehicle based on improved biological inspired neural network[J]. Ocean Engineering, 2023, 278: 114354.

[30] ZHAI H, WANG W, ZHANG W, et al. Path planning algorithms for USVs via deep reinforcement learning [C]//2021 China Automation Congress. Beijing, China: IEEE, 2021: 4281-86.

[31] YANG C, ZHAO Y, CAI X, et al. Path planning algorithm for unmanned surface vessel based on multi-objective reinforcement learning[J]. Computational Intelligence and Neuroscience, 2023, 2023(1): 2146314.

[32] CHEN C, CHEN X Q, MA F, et al. A knowledge-free path planning approach for smart ships based on reinforcement learning[J]. Ocean Engineering, 2019, 189: 106299.

[33] ZHAO Y, MA Y, HU S. USV formation and path-following control via deep reinforcement learning with random braking[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(12): 5468-78.

[34] LUIS S Y, REINA D G, MARÍN S L T. A multiagent deep reinforcement learning approach for path planning in autonomous surface vehicles: The Ypacaraí lake patrolling case[J]. IEEE Access, 2021, 9: 17084-99.

[35] 彭周华, 吴文涛, 王丹, 等. 多无人艇集群协同控制研究进展与未来趋势[J]. 中国舰船研究, 2021, 16(1): 51-64.
PENG Z H, WU W T, WANG D, et al. Coordinated control of multiple unmanned surface vehicles: Recent advances and future trends[J]. Chinese Journal of Ship Research, 2021, 16(1): 51-64.

[36] LIU Y, CHEN C, QU D, et al. Multi-USV system antidisturbance cooperative searching based on the reinforcement learning method[J]. IEEE Journal of Oceanic Engineering, 2023, 48(4): 1019-47.

[37] ZHANG J, REN J, CUI Y, et al. Multi-USV task planning method based on improved deep reinforcement learning[J]. IEEE Internet of Things Journal, 2024, 11(10): 18549-67.

[38] LI Y, LI X, WEI X, et al. Sim-real joint experimental verification for an unmanned surface vehicle formation strategy based on multi-agent deterministic policy gradient and line of sight guidance[J]. Ocean Engineering, 2023, 270: 113661.

[39] WANG C C, WANG Y L, HAN Q L, et al. Multi-USV cooperative formation control via deep reinforcement

learning with deceleration[EB/OL]. [2024-12-06]. https://ieeexplore.ieee.org/document/10621696.

[40] WANG C, WANG Y, SHI P, et al. Scalable-MADDPG-based cooperative target invasion for a multi-USV system[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 35(12): 17867-77.

[41] WEI X, WANG H, TANG Y. Deep hierarchical reinforcement learning based formation planning for multiple unmanned surface vehicles with experimental results[J]. Ocean Engineering, 2023, 286: 115577.

[42] JIN K, WANG J, WANG H, et al. Soft formation control for unmanned surface vehicles under environmental disturbance using multi-task reinforcement learning[J]. Ocean Engineering, 2022, 260: 112035.

[43] 任璐, 柯亚男, 柳文章, 等. 基于优势函数输入扰动的多无人艇协同策略优化方法[J]. 自动化学报, 2024, 51(4): 1-11.
REN L, KE Y N, LIU W Z, et al. Multi-USVs cooperative policy optimization method based on disturbed input of advantage function[J]. Acta Automatica Sinica, 2025, 51(4): 1-11.

[44] YAO P, LOU Y, WU K. Cooperative path planning for USVs assembly task[C]//2023 38th Youth Academic Annual Conference of Chinese Association of Automation (YAC). Hefei, China: IEEE, 2023: 526-531.

[45] 于长东, 刘新阳, 陈聪, 等. 基于多智能体深度强化学习的无人艇集群博弈对抗研究[J]. 水下无人系统学报, 2024, 32(1): 79-86.
YU C D, LIU X Y, CHEN C, et al. Research on game confrontation of unmanned surface vehicles swarm based on multi-agent deep reinforcement learning[J]. Journal of Unmanned Undersea Systems, 2024, 32(1): 79-86.

[46] LI F, YIN M, WANG T, et al. Distributed pursuit-evasion game of limited perception USV swarm based on multiagent proximal policy optimization[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2024, 54(10): 6435-46.

[47] XIA J, LUO Y, LIU Z, et al. Cooperative multi-target hunting by unmanned surface vehicles based on multi-agent reinforcement learning[J]. Defence Technology, 2023, 29: 80-94.

[48] QU X, GAN W, SONG D, et al. Pursuit-evasion game strategy of USV based on deep reinforcement learning in complex multi-obstacle environment[J]. Ocean Engineering, 2023, 273: 114016.

[49] LOWE R, WU Y I, TAMAR A, et al. Multi-agent actor-critic for mixed cooperative-competitive environments [C]//NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: ACM, 2017: 6383-93.

[50] REYNOLDS C W. Flocks, herds and schools: A distributed behavioral model[C]//Proceedings of the 14th annual conference on Computer graphics and interactive techniques. [S.l.]: Publication History, 1987: 25-34.

[51] WANG Z, JIN X, ZHANG T, et al. Expert system-based multiagent deep deterministic policy gradient for swarm robot decision making[J]. IEEE Transactions on Cybernetics, 2022, 54(3): 1614-24.

[52] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimization algorithms[EB/OL]. [2025-02-20]. https://arxiv.org/abs/1707.06347.

[53] XUE D, WU D, YAMASHITA A S, et al. Proximal policy optimization with reciprocal velocity obstacle based collision avoidance path planning for multi-unmanned surface vehicles[J]. Ocean Engineering, 2023, 273: 114005.

(责任编辑: 许　妍)

《水下无人系统学报》相关文献导航

1. 李健翔, 张文乐, 黎明. 带有输入时延的多无人艇系统固定时间编队控制. 2025, 33(1).
2. 郑兵, 董超, 刘涵, 等. 无人艇–机协同定位起降关键技术与验证. 2024, 32(2).
3. 梁霄, 陈聪, 刘殿勇, 等. 面向自杀式无人机饱和攻击的海空跨域无人协同反制策略. 2024, 32(2).
4. 孙峰. 一种基于海空无人集群的自杀式无人艇防御策略. 2024, 32(2).
5. 于长东, 刘新阳, 陈聪, 等. 基于多智能体深度强化学习的无人艇集群博弈对抗研究. 2024, 32(1).
6. 张海胜, 董早鹏, 杨莲, 等. 基于改进 WLSSVM 的无人艇操纵性参数辨识. 2023, 31(5).
7. 宋吉广, 李德隆, 冯亮, 等. 基于感知信息的 USV 目标环绕跟踪方法. 2023, 31(5).
8. 陈明志, 刘兰军, 陈家林, 等. 基于 HCOPSO 算法的 USV 舵向 PID 控制参数整定方法. 2023, 31(3).
9. 郭苗, 徐琰锋, 陈铢蕾. 基于博弈论的无人艇探查策略研究. 2022, 30(6).
10. 翁磊, 杨扬, 钟雨轩. 多无人艇协同遍历路径规划算法. 2020, 28(6).
11. 谢少荣, 刘坚坚, 张丹. 复杂海况无人艇集群控制技术研究现状与发展. 2020, 28(6).