doi:10.6041/j.issn.1000-1298.2025.10.006

基于改进 QMIX 算法的远洋捕捞多无人艇全覆盖路径规划策略研究

吴清云　王　东　陶　军　李志坚　殷奕杰

(上海海洋大学工程学院, 上海 201306)

摘要: 在远洋捕捞任务中,需要在特定水域内进行多无人艇全覆盖巡航以探测鱼群分布情况,但传统多智能体强化学习路径规划方法缺乏考虑自身与相邻智能体状态的能力,且反馈机制不够明确,导致路径覆盖效率较低、重复率过高。本文提出了一种基于改进 QMIX 算法(LH-QMIX)的远洋捕捞多无人艇全覆盖路径规划策略。由一个混合网络和多个智能体网络组成多智能体强化学习结构,通过混合网络将每个智能体网络的局部 Q 值融合成全局 Q 值,以指导各智能体行动。考虑到在远洋捕捞环境中无人艇通信和感知范围受限,为每个智能体网络引入一个局部损失函数,提供更明确的反馈机制,同时,引入混合注意力机制以加强无人艇之间的协作能力。在简单障碍物环境和复杂障碍物环境中,将提出的 LH-QMIX 算法与 IQL 算法、QMIX 算法进行对比仿真。结果表明,LH-QMIX 算法在简单障碍物环境下覆盖效率分别提升 14.2%、6.9%,在复杂障碍物环境下覆盖效率分别提升 22.3%、10.6%,奖励曲线在收敛后也更加稳定。研究结果为多无人艇远洋捕捞全覆盖探测任务提供了一个高效可行的解决方案,能够提升远洋捕捞效率。

关键词: 多无人艇路径规划; LH-QMIX 算法; 多智能体强化学习; 覆盖效率; 模型稳定性

中图分类号: U66　　**文献标识码**: A　　**文章编号**: 1000-1298(2025)10-0063-08　　**OSID**:

Complete Coverage Path Planning Strategy for Offshore Fishing Multi-USV Based on Improved QMIX Algorithm

WU Qingyun　WANG Dong　TAO Jun　LI Zhijian　YIN Yijie

(*College of Engineering*, *Shanghai Ocean University*, *Shanghai* 201306, *China*)

Abstract: In offshore fishing missions, multiple unmanned surface vessels (USVs) need to perform comprehensive coverage path planning within specific waters to detect fish distribution. However, traditional multi-agent reinforcement learning lacks the capability to simultaneously consider both its own and neighboring agents' states, coupled with an unclear feedback mechanism, leading to low efficiency and high redundancy in coverage tasks. To address these issues, a comprehensive coverage path planning strategy for offshore fishing was proposed by using multiple USVs based on the improved QMIX algorithm (LH-QMIX). QMIX was a multi-agent reinforcement learning method, consisting of a mixing network and multiple agent networks, which integrated the local Q-values from each agent network into a global Q-value through the mixing network to guide agent actions. Considering that communication and perception ranges were typically limited in offshore environment, a local loss function was introduced for each agent network to provide a clearer feedback mechanism. Additionally, a hybrid attention mechanism was incorporated to enhance collaboration among USVs. The proposed LH-QMIX algorithm was compared with the independent Q-learning (IQL) algorithm and the original QMIX algorithm through simulations in both simple and complex obstacle environment. Simulation result showed that compared with the traditional QMIX algorithm, the LH-QMIX achieved improvements of 6.9% and 10.6% in coverage efficiency in simple and complex obstacle environments, respectively, with more stable reward curves

收稿日期: 2025-07-25　修回日期: 2025-08-11

基金项目: 国家重点研发计划项目(2024YFD2400103)和上海市水产动物良种创制与绿色养殖协同创新中心项目(2021 科技 02-12)

作者简介: 吴清云(1986—),男,工程师,博士,主要从事无人艇智能控制研究,E-mail: qywu@shou.edu.cn

通信作者: 李志坚(1983—),男,讲师,博士,主要从事海洋工程装备研究,E-mail: zj-li@shou.edu.cn

after convergence. The research provided an effective solution for multiple USVs to efficiently achieve comprehensive coverage in offshore fishing missions, thereby enhancing the efficiency of offshore fishing operations.

Key words: multi-USV path planning; LH - QMIX algorithm; multi-agent reinforcement learning; coverage efficiency; model stability

0 引言

当前中国远洋渔业正从规模扩张向质量效益转型,通过智能化装备升级推动产业发展[1]。将无人艇应用于远洋捕捞中,可有效提升作业的智能化与可持续性水平,减少船员安全作业风险与运营成本[2-5]。每艘无人艇作为一个独立的智能体,具备自主导航、环境感知与决策能力。通过先进的通信技术和协同算法,这些无人艇能够实时交换信息,从而实现对目标区域的高效协作探索与作业。无人艇在执行任务时,路径规划的智能程度直接影响其作业效率与任务完成质量[6-8]。而多无人艇通过协同规划路径,能够更好地应对复杂环境,提高作业覆盖效率,保证任务完成的可靠性[9-11]。因此,多智能体协同技术成为了目前的研究热点之一[12-14]。

针对多智能体协同路径规划问题,SONG 等[15]在 IQL 算法[16](Independent Q-learning)的基础上引入了动态奖励函数,有效提升了规划效率,但奖励函数的设计较为复杂且泛化能力较差。NANTOGMA 等[17]提出一种结合专家知识引导的混合多智能体强化学习框架,通过引入启发式机制提升复杂环境下多智能体的学习效率和协同性能,但该方法过于依赖预设的环境知识,在未知环境中的泛化能力较弱。WAGNER 等[18]提出一个子维度扩展框架,先为每个智能体单独规划路径,再在必要时协调运动,通过局部增加搜索空间维度解决智能体间碰撞的问题,从而减少计算成本,但此方法在复杂环境下效率不高,性能不如预期。YU 等[19]提出一种结合虚拟力与 RRT 的改进算法,综合考虑洋流和障碍物对路径规划的影响并提升了路径质量与任务效率,但该算法在大规模动态环境中的实时性与计算开销未充分验证。HUANG 等[20]通过机器学习选择代理集来优化路径规划,该算法能够快速找到初始解决方案,并随着时间推移不断改进解决方案,直至接近最优,但算法性能取决于训练数据的质量和多样性,且同样存在适应性不足的问题。

相较于传统多智能体路径规划算法,多智能体强化学习算法能够综合考虑路径长度、时间成本、能耗等多个因素,学习到更优的路径策略,从而更好地适应动态环境下的多智能体协同与对抗问题[21-23]。然而,现有的多智能体强化学习覆盖算法在极端海洋环境下,仍存在问题:①协同规划能力较弱,覆盖策略缺乏同时考虑智能体自身和相邻智能体状态进行协同规划的能力,导致算法的覆盖效率较低[24]。②反馈机制不够明确,传统的多智能体强化学习算法通常依赖于全局奖励信号来指导个体的学习过程,这种全局奖励难以精确反映每个智能体的具体贡献[25]。③远洋捕捞这类真实环境中,通信范围往往会受到诸多限制,算法的性能也会明显降低。

针对上述问题,本文提出一种基于改进 QMIX 算法(LH - QMIX)的多无人艇全覆盖路径规划策略。考虑到通信范围的限制,为每个智能体网络引入一个局部损失函数,以提供更直接、更明确的反馈信号;同时加入混合注意力机制(Hybird attention mechanism, HAM),自适应地聚合相邻智能体和障碍物的重要特征信息,在智能体感知、通信受限的情况下,以期加强各智能体间的协作能力。

1 任务建模

1.1 环境建模

鉴于远洋捕捞环境下多无人艇全覆盖探测任务的复杂性,将环境建模为二维栅格地图,如图 1 所示。其中黑色栅格为障碍区域,白色栅格为可行区域,每个栅格尺寸为 1 m × 1 m。

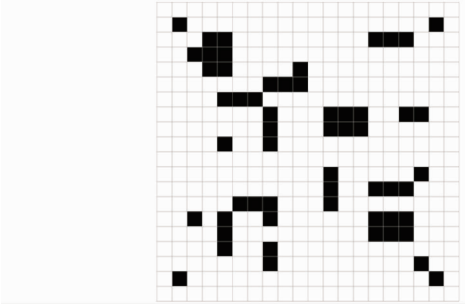

图 1 二维栅格地图

Fig. 1 2D grid map

同时将栅格地图转换为 $h \times w$ 的状态矩阵,即

$$\boldsymbol{S} = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,w} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,w} \\ \vdots & \vdots & & \vdots \\ p_{h,1} & p_{h,2} & \cdots & p_{h,w} \end{bmatrix} \tag{1}$$

其中

$$p=\begin{cases}0 & (\text{未覆盖区域})\\ 1 & (\text{障碍物区域})\\ 2 & (\text{已覆盖区域})\\ 3 & (\text{智能体位置})\end{cases} \tag{2}$$

式中　$\boldsymbol{S}$——栅格地图状态矩阵

h——栅格地图高度

w——栅格地图宽度

p——栅格地图上相应位置环境状态

在无人艇进行全覆盖任务的同时，状态矩阵也会同步更新，当前状态 S_t 更新到下一个状态 S_{t+1} 的过程表示为

$$S_{t+1}=\begin{cases}S_t(p_{i-1,j}=3,p_{i,j}=2) & (a_t=1)\\ S_t(p_{i,j-1}=3,p_{i,j}=2) & (a_t=2)\\ S_t(p_{i+1,j}=3,p_{i,j}=2) & (a_t=3)\\ S_t(p_{i,j+1}=3,p_{i,j}=2) & (a_t=4)\end{cases} \tag{3}$$

式中　a_t——无人艇上左下右操作

$p_{i,j}$——t 时刻无人艇所在位置

$p_{i\pm1,j\pm1}$——$t+1$ 时刻无人艇所在位置

S_t——t 时刻环境全局状态空间

1.2　去中心化部分可观马尔可夫决策模型

去中心化部分可观马尔可夫决策模型（Markov decision process，MDP）是多智能体系统中的一种决策模型，用于描述多个智能体在部分可观测环境中进行决策的问题。MDP 模型为

$$G=(S,A,P(s_t|s_{t+1},a),R(s,a),Z,O(s,i),n,\gamma) \tag{4}$$

式中　S——环境的全局状态空间

A——智能体共享动作集合

$P(s_t|s_{t+1},a)$——在状态 s_t 下执行动作 a 转移到状态 s_{t+1} 的概率

$R(s,a)$——在状态 s 下执行联合动作 a 获得的奖励

Z——观测空间，智能体可能接收到的观测

$O(s,i)$——在状态 s 下智能体 i 接收到的观测

n——智能体数量

γ——折扣因子

MDP 涉及多个智能体，每个智能体只能观察到部分环境信息，并且只能基于自己的观测来选择动作，但所有智能体又共享一个全局奖励信号，这有利于促进智能体之间的协作。共享的全局奖励信号使得智能体在决策时不仅要考虑自身局部利益，还要兼顾系统整体利益，从而在复杂环境中实现更高效的协同工作。

2　LH－QMIX 算法

2.1　QMIX 算法

QMIX 算法[26]由混合网络（Mixing network）和智能体网络（Agent networks）组成，算法框架如图 2 所示。每个智能体拥有一个独立的 DRQN 网络，循环输入当前时刻 t 的观测值 o_t 以及上一时刻 $t-1$ 的动作 a_{t-1} 来计算其局部 Q 值。混合网络是一个前馈神经网络，将智能体网络的输出 $Q(\tau,a_t)$ 作为输入进行单调混合，从而计算出全局 Q 值 $Q_{tot}(\tau,a)$。

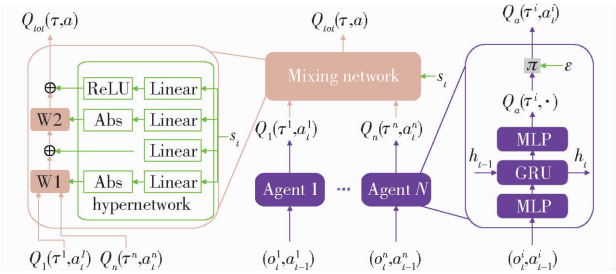

图 2　QMIX 算法框架图

Fig. 2　QMIX algorithm framework diagram

QMIX 算法的损失函数为

$$\begin{cases}L_{global}=\sum_{i=1}^{b}(y_{tot}^i-Q_{tot}(\tau,a,s;\theta_{tot}))^2\\ y_{tot}=r+\gamma\max_{a'}Q_{tot}(\tau',a',s';\theta_{tot}^-)\end{cases} \tag{5}$$

式中　b——每次从经验回放缓冲区中随机采样批次大小

L_{global}——全局损失函数，用于训练模型时最小化预测全局 Q 值与目标 Q 值之间的均方误差

r——即时奖励

τ'——下一个状态轨迹

a'——下一个状态联合动作

s'——下一个状态

θ_{tot}^-——目标混合网络参数

θ_{tot}——当前混合网络参数

τ——当前轨迹

a——当前联合动作

s——当前状态

y_{tot}——全局目标 Q 值

2.2　添加局部损失

个体智能体特定反馈对于优化其行为至关重要，尤其是在多无人艇全覆盖路径规划这类复杂的多智能体协作任务中，每个智能体行为都会影响整体性能。QMIX 算法训练目标是最小化全局 TD 误差。然而，全局 TD 误差是所有智能体联合动作价值的误差，难以反映单个智能体的动作对全局价值的贡献，对智能体网络的反馈不明确，使得智能体在

某些情况下无法获得足够的反馈,从而影响学习效率和性能。为此,将为每个智能体网络引入一个局部损失,从而提供更直接、更明确的反馈信号。局部损失函数为

$$\begin{cases} L_{local} = \sum_{i=1}^{b} (y^i - Q(\tau, a; \theta))^2 \\ y = r + \gamma \max_{a'} Q(\tau', a'; \theta^-) \end{cases} \tag{6}$$

式中 L_{local}——局部损失函数

$Q(\tau, a; \theta)$——当前时刻局部 Q 值

$Q(\tau', a'; \theta^-)$——下一时刻局部 Q 值

y——局部目标 Q 值

加入局部损失后的算法框架图如图 3 所示。在原始的 QMIX 算法中,全局损失 L_{global} 通过混合网络反向传播到智能体网络,因此智能体网络接收到的反馈信号并不清晰。而局部损失 L_{local} 直接反向传播到智能体网络,为每个智能体提供更加清晰的反馈,使得每个智能体可独立地优化自己的 Q 值,提高系统性能。

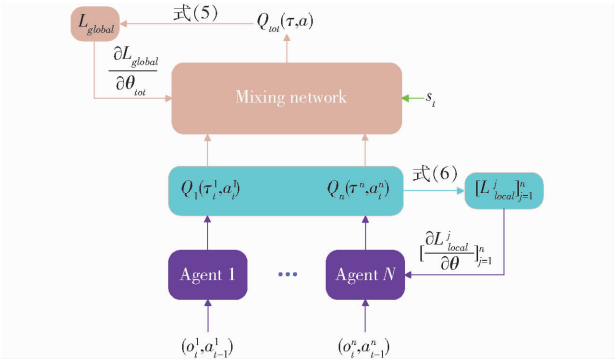

图 3 加入局部损失的算法框架图

Fig. 3 Algorithm framework diagram with local loss

2.3 混合注意力机制

在远洋捕捞环境中,智能体通信和感知范围有限,这会使智能体网络的输入维度不确定。且传统多智能体强化学习本身的协同规划能力较弱,覆盖策略缺乏同时考虑智能体自身和相邻智能体状态进行协同规划的能力。因此,引入混合注意力机制,使智体能够自适应地提取相邻智能体和障碍物的必要特征信息,加强智能体间协作能力的同时也能够排除冗余信息。

将智能体 i 自身位置信息记为 $o_{t,self}^i$,智能体 i 观测到的相邻智能体位置信息记为 $o_{t,others}^i$,智能体 i 观测到的相邻障碍物位置信息记为 $o_{t,obst}^i$,因此智能体 i 观测信息可表示为 $o_t^i = (o_{t,others}^i, o_{t,self}^i, o_{t,obst}^i)$。

智能体 i 自身位置信息 $o_{t,self}^i$ 表达式为

$$\begin{cases} o_{t,self}^i = p_i \\ p_i = \left(\frac{f_i}{h}, \frac{g_i}{w}\right) \end{cases} \tag{7}$$

式中 (f_i, g_i)——智能体 i 在栅格地图上的实际坐标

p_i——智能体 i 相对位置

智能体 i 观测到的相邻智能体位置信息 $o_{t,others}^i$ 表达式为

$$o_{t,others}^i = (p_1^i, \cdots, p_j^i, \cdots, p_c^i) \tag{8}$$

其中 $$p_j^i = \left(\frac{f_j - f_i}{r_c}, \frac{g_j - g_i}{r_c}\right)$$

式中 p_j^i——智能体 j 与智能体 i 相对位置

c——智能体 i 通讯范围内智能体数量

(f_j, g_j)——在栅格地图上智能体 j 实际坐标

r_c——智能体通信半径

智能体 i 观测到的相邻障碍物的位置信息 $o_{t,obst}^i$ 表达式为

$$\begin{cases} o_{t,obst}^i = (p_{1,obst}^i, \cdots, p_{k,obst}^i, \cdots, p_{d,obst}^i) \\ p_{k,obst}^i = \left(\frac{f_k - f_i}{r_{\det}}, \frac{g_k - g_i}{r_{\det}}\right) \end{cases} \tag{9}$$

式中 d——智能体 i 观测范围内的障碍物数量

$p_{k,obst}^i$——第 k 个障碍物与智能体 i 相对位置

(f_k, g_k)——智能体 i 观测到的第 k 个障碍物坐标位置

$r_{\det}$——智能体感知半径

将提取智能体 i 与相邻智能体间的位置关系特征模块记为 Atten1,使用参数矩阵 $\boldsymbol{W}_{q1}^i$ 将 $o_{t,self}^i$ 映射为查询向量 $\boldsymbol{Q}_1$,使用参数矩阵 $\boldsymbol{W}_{k1}^i$、$\boldsymbol{W}_{v1}^i$ 将 $o_{t,others}^i$ 映射为键值对 $\boldsymbol{K}_1$、$\boldsymbol{V}_1$,算式为

$$\begin{cases} \boldsymbol{Q}_1 = o_{t,self}^i \boldsymbol{W}_{q1}^i \\ \boldsymbol{K}_1 = o_{t,others}^i \boldsymbol{W}_{k1}^i \\ \boldsymbol{V}_1 = o_{t,others}^i \boldsymbol{W}_{v1}^i \end{cases} \tag{10}$$

同样地,将提取智能体 i 与相邻障碍物间的位置关系特征的模块记为 Atten2,使用参数矩阵 $\boldsymbol{W}_{q2}^i$ 将 $o_{t,self}^i$ 映射为查询向量 $\boldsymbol{Q}_2$,使用参数矩阵 $\boldsymbol{W}_{k2}^i$、$\boldsymbol{W}_{v2}^i$ 将 $o_{t,obst}^i$ 映射为键值对 $\boldsymbol{K}_2$、$\boldsymbol{V}_2$,算式为

$$\begin{cases} \boldsymbol{Q}_2 = o_{t,self}^i \boldsymbol{W}_{q2}^i \\ \boldsymbol{K}_2 = o_{t,obst}^i \boldsymbol{W}_{k2}^i \\ \boldsymbol{V}_2 = o_{t,obst}^i \boldsymbol{W}_{v2}^i \end{cases} \tag{11}$$

随后计算注意力系数 $\boldsymbol{h}_1$、$\boldsymbol{h}_2$,计算公式为

$$\begin{cases} \boldsymbol{h}_1 = \text{softmax}\left(\frac{\boldsymbol{Q}_1 \boldsymbol{K}_1^{\mathrm{T}}}{\sqrt{d_h}}\right) \boldsymbol{V}_1 \\ \boldsymbol{h}_2 = \text{softmax}\left(\frac{\boldsymbol{Q}_2 \boldsymbol{K}_2^{\mathrm{T}}}{\sqrt{d_h}}\right) \boldsymbol{V}_2 \end{cases} \tag{12}$$

式中 $\boldsymbol{h}_1$——智能体 i 与相邻智能体间聚合特征向量

softmax——将注意力得分归一化为概率分布函数

$\boldsymbol{h}_2$——智能体 i 与相邻障碍物间聚合特征向量

$\sqrt{d_h}$——缩放因子，可稳定梯度

最后将注意力系数 $\boldsymbol{h}_1$、$\boldsymbol{h}_2$ 以及经 MLP 处理过的动作 a^i_{t-1} 拼接成一个特征向量作为特征提取网络的输出。加入混合注意力机制的 QMIX 算法框架如图 4 所示。

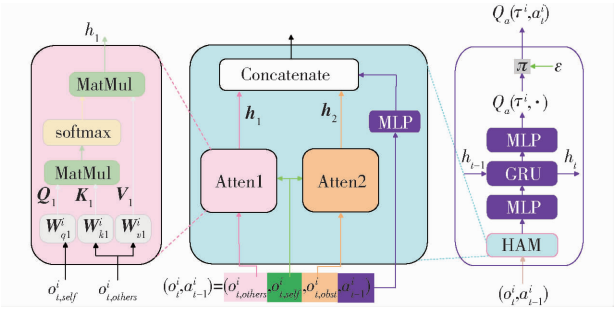

图 4　加入混合注意力机制的 QMIX 算法框架

Fig. 4　QMIX algorithm framework with hybrid attention mechanism

2.4　奖励函数

在远洋捕捞全覆盖路径规划任务中，安全是首要前提，因此将安全性设置为主要奖励函数，当出现安全事故时，立刻停止规划作业并进入下一次迭代，即

$$r_{\text{safe}}=\begin{cases}-50 & \text{（出现安全事故）}\\ 0 & \text{（未出现安全事故）}\end{cases} \tag{13}$$

除了安全性问题外，还需要设计一些辅助性奖励函数来提高算法性能。辅助性奖励函数为

$$r_{\text{len}}=-0.05\sum_{i=1}^{n}N^i_{\text{step}} \tag{14}$$

$$r_{\text{cov}}=0.3N_{\text{cov}}+\mu\frac{N_{\text{cov}}}{N_{\text{all}}} \tag{15}$$

$$r_{\text{rep}}=-0.1N_{\text{rep}}-\lambda\frac{N_{\text{rep}}}{N_{\text{all}}} \tag{16}$$

式中　r_{len}——路径总长度设计的奖励函数

N^i_{step}——智能体 i 所走的时间步

r_{cov}——覆盖率设计的奖励函数

N_{cov}——被覆盖栅格数量

N_{all}——可通行栅格总数

μ——覆盖率折扣因子

r_{rep}——路径重复率设计的奖励函数

N_{rep}——被重复覆盖的栅格数量

λ——路径重复率折扣因子

因此总奖励函数 r_{total} 为

$$r_{\text{total}}=r_{\text{safe}}+r_{\text{len}}+r_{\text{cov}}+r_{\text{rep}} \tag{17}$$

3　仿真试验

3.1　试验设计

为了验证本文算法有效性，将 LH－QMIX 算法与 IQL 算法[16]、QMIX 算法[26]分别在尺寸为 20 m × 20 m 的简单障碍物环境和复杂障碍物环境中进行仿真，并以覆盖率、覆盖效率、全局奖励等作为主要指标来评估全覆盖算法性能。算法参数设置如表 1 所示。

表 1　算法参数设置

Tab. 1　Algorithm parameters setting

参数	数值
栅格地图高度 h/m	20
栅格地图宽度 w/m	20
智能体感知半径 r_{det}/m	3
智能体通信半径 r_c/m	8
智能体数量 n	4
Agent 网络学习率 l_{r_a}	0.001
Mixing 网络学习率 l_{r_m}	0.001
最大训练步数 $T_{\max}$	160
网络更新间隔 T_{update}	40
全局 Q 值折扣因子 γ	0.95
覆盖率折扣因子 μ	0.9
路径重复率折扣因子 λ	0.5
采样批次大小 b	32

3.2　试验结果与分析

3.2.1　简单障碍物环境

3 种算法在简单障碍物环境下的路径如图 5 所示，试验设置 15 个随机障碍物，4 个无人艇从地图中心出发执行全覆盖任务，能够看出 3 种算法都能够在保持任务安全性的情况下完成全部覆盖率。但也能够发现 IQL 算法的路径存在较多绕行和局部重复覆盖，尤其在障碍物密集区域，多个无人艇的路径交叉，导致冗余动作较多。QMIX 算法通过值分解策略优化了个体与全局目标的一致性，路径规划更为紧凑，但仍存在部分智能体在障碍物边缘区域重复扫描。LH－QMIX 算法路径规划表现出显著的协同特性：通过混合注意力机制，智能体能够动态聚合相邻无人艇和障碍物的关键信息（如位置、运动方向、障碍物密度），从而在覆盖过程中优先选择无冲突且高效的路径。

3 种算法在简单障碍物环境中覆盖效率如图 6 所示，覆盖效率通过覆盖率与规划步数的关系来衡量。从图 6 可以看出，3 种算法都能在有限的规划步数中实现对整个目标区域的完全覆盖。在简单障碍物环境中，IQL 算法使用 126 个时间步完成全覆盖任务，QMIX 算法使用 116 个时间步完成全覆盖任务，本文提出的 LH－QMIX 算法仅用 108 个时间步完成了任务，相比于传统的 IQL 算法与 QMIX 算法，覆盖效率分别提升 14.2% 和 6.9%，原因为引入的局部损失函数为每个智能体提供了更直接的反馈

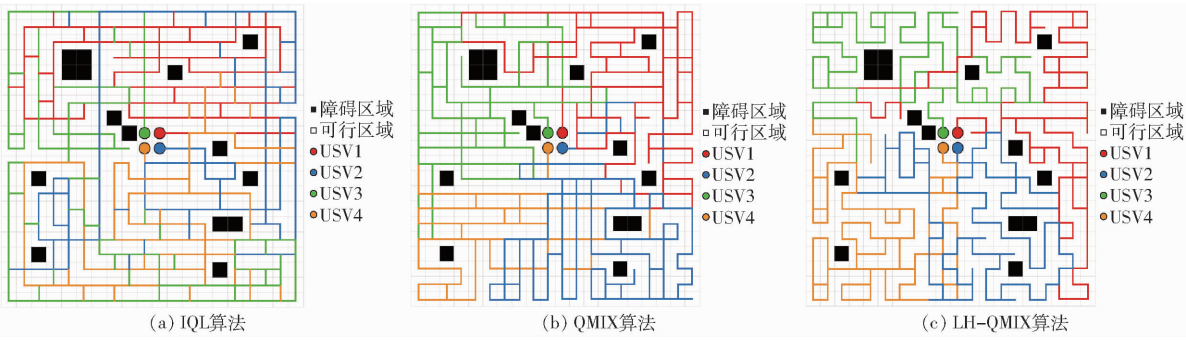

图 5 简单障碍物环境下各算法路径

Fig. 5 Path planning trajectories of algorithms in simple obstacle environment

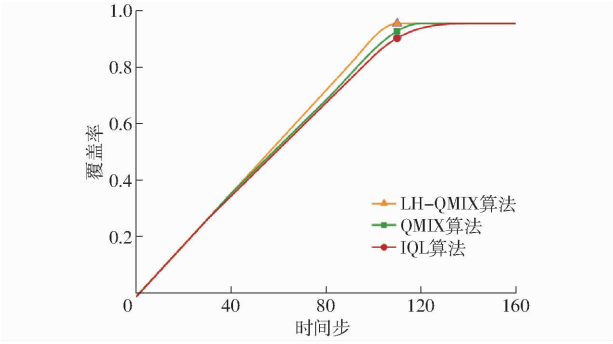

图 6 简单障碍物环境下覆盖效率

Fig. 6 Coverage efficiency in simple obstacle environment

信号,使其在训练过程中能够更快调整路径策略,减少无效动作。同时注意力机制通过自适应加权聚合相邻智能体和障碍物的信息,解决了通信受限环境下信息不完整问题。例如,在障碍物稀疏区域,智能体优先关注远处的障碍物和未覆盖区域,而在障碍物密集区域,则聚焦于邻近智能体的实时状态,从而动态优化路径规划。这种机制显著降低了重复覆盖率。

在简单障碍物环境中,3 种算法在经过 5 000 个回合训练后的回合平均奖励对比曲线如图 7 所示。由图 7 可以看出 3 种算法回合平均奖励都取得了收敛结果。IQL 算法奖励曲线在早期快速上升,但后期波动剧烈,其收敛性受局部最优策略影响较大,导致模型稳定性不足。

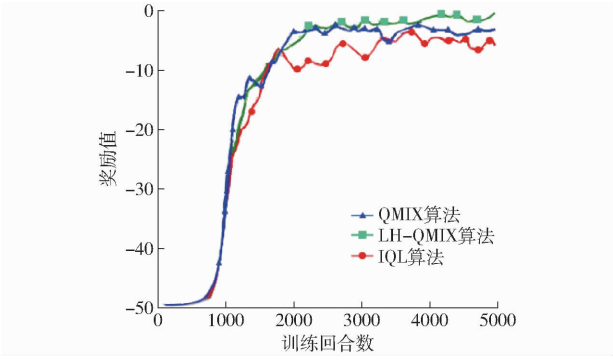

图 7 简单障碍物环境下回合平均奖励

Fig. 7 Episode-averaged rewards in simple obstacle environment

QMIX 算法奖励曲线呈平滑上升趋势,其值分解机制有效平衡了个体与全局目标,但在复杂环境中仍存在策略探索不足的问题。本文提出的 LH - QMIX 算法在初始阶段相较于其他 2 个算法收敛较慢,但在第 2 000 个回合左右超过其他算法,并在最后的收敛过程中保持了较低的方差,原因为混合注意力机制引入新的动态特性,使得智能体在训练初期需要更长时间来学习如何有效地利用注意力机制来聚焦关键信息。这种动态特性的变化会影响训练的稳定性,导致收敛速度变慢。但是,当训练达到后期时,也能发现本文提出的 LH - QMIX 算法奖励值更高,收敛效果也更加稳定。

3.2.2 复杂障碍物环境

3 种算法在复杂障碍物环境下的路径如图 8 所示。试验设置 50 个随机障碍物,能够看出 3 种算法在复杂障碍物环境下也能够完成全部覆盖率。但从图中发现,IQL 算法在障碍物密集区域,多个无人艇因缺乏全局信息感知而频繁绕行,导致路径交叉和冗余动作激增。例如,在左上角区域,2 个无人艇因未及时共享障碍物信息,分别沿不同方向绕行,导致覆盖效率严重下降。QMIX 算法路径规划相对紧凑,但在障碍物高度密集区域仍存在局部重复扫描问题。例如,在中央障碍物群中,3 个无人艇因通信延迟未能同步避障策略,导致部分区域被重复覆盖。LH - QMIX 算法的路径规划表现出更强的适应性和协同能力:通过混合注意力机制,智能体能够动态聚合相邻无人艇和障碍物的实时信息(如位置、运动方向、障碍物密度)。例如,在右下角障碍物密集区,LH - QMIX 算法的智能体通过共享感知信息,快速调整路径以避免碰撞,同时优先覆盖未扫描区域,显著减少冗余动作。

同时,在复杂障碍物环境中进行了仿真试验,3 种算法在复杂障碍物环境下的覆盖效率如图 9 所示。从图 9 可以看出,在复杂障碍物环境下,IQL 算法与 QMIX 算法分别在第 130 个时间步与第 113 个时间步完成了任务,本文提出的 LH - QMIX 算法仅

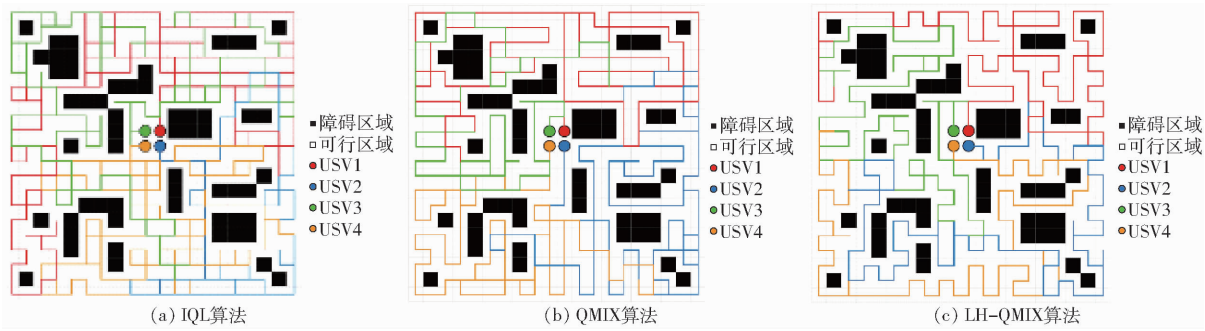

图 8　复杂障碍物环境下各算法路径

Fig. 8　Path planning trajectories of algorithms in complex obstacle environment

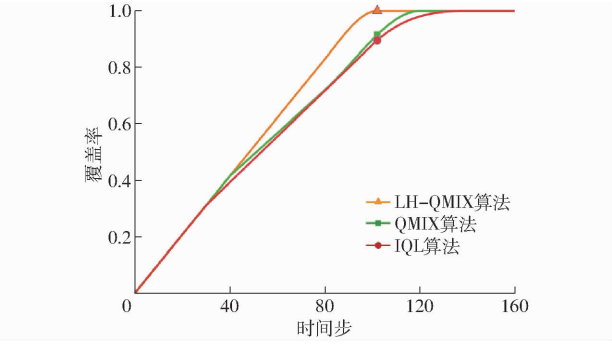

图 9　复杂障碍物环境下覆盖效率

Fig. 9　Coverage efficiency in complex obstacle environment

用 101 步完成了任务，相比 IQL 算法与 QMIX 算法，覆盖效率分别提升 22.3% 与 10.6%。这表明 LH－QMIX 算法在障碍物密集的区域内执行任务更有优势，原因为 LH－QMIX 算法的智能体通过局部损失可以快速识别冗余动作，从而优化路径。通过注意力机制，智能体能够在障碍物稀疏区域优先关注远处的障碍物和未覆盖区域，而在障碍物密集区域，则聚焦于邻近智能体的实时状态，从而动态优化路径规划。

图 10 为在复杂障碍物环境下，3 种算法在经过

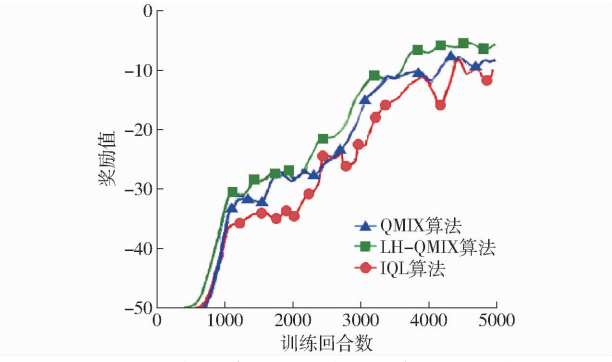

图 10　复杂障碍物环境下回合平均奖励

Fig. 10　Episode-averaged rewards in complex obstacle environment

5 000 个回合训练后回合平均奖励对比曲线。可以看出 3 种算法回合平均奖励也同样取得了收敛的结果，但 IQL 算法收敛效果波动较大，收敛不稳定。与传统的 QMIX 算法相比，LH－QMIX 算法依旧收敛较慢，但与图 7 所示的简单障碍物收敛曲线不同，LH－QMIX 算法在训练初期得到的回合平均奖励就高于其他两个算法，并在训练中后期也保持领先，在奖励收敛后也一直保持着稳定状态。这同样说明本文提出的 LH－QMIX 算法在处理多无人艇全覆盖路径规划时更具优势。

4　结论

(1)提出了一种基于 LH－QMIX 算法的全覆盖路径规划策略。针对传统 QMIX 算法反馈机制不够明确的问题，给每个智能体网络引入一个局部损失，从而提供更明确的反馈信号。针对传统 QMIX 算法协同规划能力较弱的问题，在其中引入了混合注意力机制，使得无人艇能够自适应提取相邻无人艇与障碍物的特征，从而加强群体协作能力。

(2)模拟仿真试验结果表明，LH－QMIX 算法具备执行 100% 全覆盖规划任务的能力。相较于 IQL 算法和 QMIX 算法，LH－QMIX 算法在简单障碍物环境下覆盖效率分别提升 14.2%、6.9%，在复杂障碍物环境下覆盖效率分别提升 22.3%、10.6%。所提出的算法虽然在训练前期收敛较慢，但在训练后期收敛过程更为稳定，有效改善了传统 QMIX 算法反馈机制不明确的缺点。在通信受限的现场情况下，LH－QMIX 算法可有效增强多无人艇协同规划能力，提升覆盖率。

(3)提出的 LH－QMIX 算法计算参数可以根据现场海洋环境进行实时调节，实现更智能、高效的远洋捕捞作业。

参　考　文　献

[1]　中华人民共和国农业农村部. 全国远洋渔业发展规划(2025—2030)[Z]. 北京：农业农村部，2025.

[2]　MATSUSHITA Y, ONUMA A, TAKESHITA C, et al. Unmanned surface vehicle (USV) with a fish attraction lamp to assist the

purse seine operations[J]. Fisheries Science, 2024, 90(3): 357 - 367.
[3] HANDEGARD N O, DE ROBERTIS A, HOLMIN A J, et al. Uncrewed surface vehicles (USVs) as platforms for fisheries and plankton acoustics[J]. ICES Journal of Marine Science, 2024, 81(9): 1712 - 1723.
[4] SOTELO-TORRES F, ALVAREZ L V, ROBERTS R C. An unmanned surface vehicle (USV): development of an autonomous boat with a sensor integration system for bathymetric surveys[J]. Sensors, 2023, 23(9): 4420.
[5] EVANS T M, RUDSTAM L G, SETHI S A, et al. Fish avoidance of ships during acoustic surveys tested with quiet uncrewed surface vessels[J]. Fisheries Research, 2023, 267: 106817.
[6] 刘天湖,赖嘉上,孙伟龙,等.基于跳点优化蚁群算法的菠萝田间导航路径规划[J].农业机械学报,2025,56(4):387 - 396.
LIU Tianhu, LAI Jiashang, SUN Weilong, et al. Pineapple field navigation path planning based on jump point optimization ant colony algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(4): 387 - 396. (in Chinese)
[7] XING B, WANG X, YANG L, et al. An algorithm of complete coverage path planning for unmanned surface vehicle based on reinforcement learning[J]. Journal of Marine Science and Engineering, 2023, 11(3): 645.
[8] 袁杰,张迎港,加尔肯别克,等.基于 AFD 融合算法的运输机器人路径规划方法[J].农业机械学报,2025,56(6):594 - 607.
YUAN Jie, ZHANG Yinggang, JIA Erkenbieke, et al. Path planning method for transport robots based on AFD fusion algorithm [J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(6): 594 - 607. (in Chinese)
[9] 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 - 18567.
[10] CHEN Z, ZHAO Z, XU J, et al. A cooperative hunting method for multi-USV based on the A* algorithm in an environment with obstacles[J]. Sensors, 2023, 23(16): 7058.
[11] 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.
[12] CHEN C, LIANG X, ZHANG Z, et al. Cooperative game method of heterogeneous unmanned surface vehicles based on distributed decision-making framework[J]. Ocean Engineering, 2025, 338: 122048.
[13] JEONG S K, KIM M K, PARK H Y, et al. Study on fault diagnosis technology for efficient swarm control operation of unmanned surface vehicles[J]. Applied Sciences, 2024, 14(10): 4210.
[14] YUAN P, ZHANG Z, LI Y, et al. Leader-follower control and APF for multi-USV coordination and obstacle avoidance[J]. Ocean Engineering, 2024, 313: 119487.
[15] SONG Z, ZHENG R, ZHANG S, et al. Cooperative reward shaping for multi-agent path finding[J]. arXiv Preprint, arXiv: 2407.10403, 2024.
[16] KOSTRIKOV I, NAIR A, LEVINE S. Offline reinforcement learning with implicit Q-learning[J]. arXiv Preprint, arXiv: 2110.06169, 2021.
[17] NANTOGMA S, ZHANG S, YU X, et al. Multi-USV dynamic navigation and target capture: a guided multi-agent reinforcement learning approach[J]. Electronics, 2023, 12(7): 1523.
[18] WAGNER G, CHOSET H. Subdimensional expansion for multirobot path planning[J]. Artificial Intelligence, 2015, 219: 1 - 24.
[19] 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.
[20] HUANG T, LI J, KOENIG S, et al. Anytime multi-agent path finding via machine learning-guided large neighborhood search [C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(9): 9368 - 9376.
[21] RAMEZANI M, AMIRI ATASHGAH M A, REZAEE A. A fault-tolerant multi-agent reinforcement learning framework for unmanned aerial vehicles-unmanned ground vehicle coverage path planning[J]. Drones, 2024, 8(10): 537.
[22] MA Z, YOU J, ZHANG Y, et al. Reinforcement learning-based dynamic coverage control of multi-rotor UAVs with safety priority[J]. IEEE Transactions on Automation Science and Engineering, 2024, 22: 17474 - 17485.
[23] CHEN D, QI Q, FU Q, et al. Transformer-based reinforcement learning for scalable multi-UAV area coverage[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(8): 10062 - 10077.
[24] YUAN G, XIAO J, HE J, et al. Multi-agent cooperative area coverage: a two-stage planning approach based on reinforcement learning[J]. Information Sciences, 2024, 678: 121025.
[25] CHOI H B, KIM J B, HAN Y H, et al. MARL-based cooperative multi-AGV control in warehouse systems[J]. IEEE Access, 2022, 10: 100478 - 100488.
[26] RASHID T, SAMVELYAN M, DE WITT C S, et al. Monotonic value function factorisation for deep multi-agent reinforcement learning[J]. Journal of Machine Learning Research, 2020, 21: 1 - 51.