DOI:10.11784/tdxbz202411024

面向多目标协同搜索的多无人船模糊满意强化学习方法

胡超芳，朱　琦
（天津大学电气自动化与信息工程学院，天津 300072）

摘　要：无人船因其高效率、低成本、强抗风险的特点，被广泛应用于各种复杂环境中执行海洋任务. 针对多无人船在未知水域内的多目标协同搜索问题，提出了一种基于模糊满意多指标优化和双经验回放池的改进强化学习方法. 首先构建了包含环境认知度和目标存在概率两个信息指标的二维栅格环境地图. 其次针对单经验回放池随机采样数据训练效率低的问题，提出使用双经验回放池分类存储数据，为提高初期训练速度和后期稳定性，按照时变比例分别调用数据改进训练. 此外，为实现对目标的快速搜索，同时保证搜索区域的覆盖度和无人船间的安全避撞，提出了目标存在概率变化量、环境搜索覆盖度和无人船分布距离 3 个奖励函数. 为满足 3 个奖励函数重要性等级要求，使用基于松弛优先级满意度的模糊多指标优化方法对奖励函数进行重新建模，从而形成了改进模糊满意 D3QN 算法. 最后，对所提算法的有效性和不同数量目标搜索任务的适用性进行仿真验证，证实了算法可以满足设计要求. 同时，考虑到无人船实际底层控制误差对上层搜索算法的影响，将所提模糊满意强化学习算法用做上层规划与下层线性自抗扰控制结合，进行了多目标协同搜索的应用仿真验证，并与其他强化学习方法进行了对比. 结果表明：使用所提算法不但可以对环境内的多个未知目标实现快速有效搜索，而且可以有效适应实际控制误差的存在，所提算法在搜索速度、环境搜索覆盖度和无人船分布性上均优于对比算法.

关键词：无人船；协同搜索；强化学习；模糊满意优化；线性自抗扰控制
中图分类号：TP181；U664.82　　**文献标志码**：A　　**文章编号**：0493-2137（2025）11-1132-13

Fuzzy Satisfactory Reinforcement Learning Method of Unmanned Surface Vessels for Multi-Target Collaborative Search

Hu Chaofang，Zhu Qi
（School of Electrical and Information Engineering，Tianjin University，Tianjin 300072，China）

Abstract：Unmanned surface vessels（USVs）were widely used in various complex environments to perform marine tasks because of their high efficiency，low cost，and strong anti risk characteristics. Addressing the multi-target collaborative search problem of USVs in unknown environments，an improved reinforcement learning method based on fuzzy satisfactory multi-objective optimization and dual-experience playback pools was proposed. First，a two-dimensional grid environment map，including the two information indicators of environmental awareness and target existence probability，was constructed. Second，aiming at low training efficiency of random sampling data in a single experience playback pool，two experience playback pools were used to classify and store data. To improve the initial training speed and lateral stability，the data were recalled according to time-varying proportions to improve the training. In addition，to realize the fast search of targets while ensuring the coverage of the search area and safe collision avoidance between USVs，the three reward functions of target existence probability variation，environmental search coverage，and USV distribution distance were proposed. To meet the importance level requirements of these three reward functions，a fuzzy multi-objective optimization method based on relaxed priority satisfaction was used to remodel the reward function and an improved fuzzy satisfaction dueling double deep Q-network（D3QN）algorithm was formed. Finally，the effective-

收稿日期：2024-11-19；**修回日期**：2025-01-27.
作者简介：胡超芳（1973—　　），男，博士，教授.
通信作者：胡超芳，cfhu@tju.edu.cn.
基金项目：天津市自然科学基金重点项目（23JCZDJC01140）.
Supported by the Natural Science Foundation of Tianjin，China（No. 23JCZDJC01140）.

ness of the proposed algorithm and its applicability to different numbers of target search tasks were verified by simulation. The algorithm was proven to meet the design requirements. Simultaneously, considering the impact of the actual bottom-level control errors of USVs on the top-level search algorithm, the proposed fuzzy satisfactory reinforcement learning algorithm was taken as the top-level planning and combined with the bottom-level linear active disturbance rejection control. The application of multi-target collaborative search was verified by simulation and compared with other reinforcement learning methods. The results show that the proposed algorithm can not only realize a fast and effective search for multiple unknown targets in the environment but also effectively adapt to the case with the presence of actual control errors. In addition, the proposed algorithm is superior to the comparison algorithm in terms of search speed, environmental search coverage, and USV distribution.

Keywords: unmanned surface vessel (USV); collaborative search; reinforcement learning; fuzzy satisfactory optimization; linear active disturbance rejection control

随着海洋经济的快速发展，大力发展海洋装备、依法维护海洋权益成了必然选择. 无人船[1]作为智能海洋系统的主要工具，具有小型化、智能化的特点，被广泛地用于海洋作业中. 由于海洋环境复杂多变，存在较大的未知性，使用无人船搜索未知水域成为执行各种海洋任务的前提，因此基于多无人船的协同搜索成为一个主要研究方向[2].

目标协同搜索是指在一定约束条件下使用多无人系统实现对区域内目标的检测与跟踪[3]. 基于该研究问题，最早采用的是以随机搜索策略[4]、遍历搜索策略[5]为代表的典型搜索算法，但是因为缺少信息引导，该类算法搜索效率较低. 之后，诸多研究学者将搜索图与智能算法或控制算法相结合，提出了基于图论引导的多种优化搜索算法，有效提高了搜索能力[6]，例如针对多无人机协同搜索路径规划问题提出的改进智能水滴算法[7]；基于预测控制思维设计的多水下机器人分布式协同搜索策略[8]. 群智能优化算法和控制算法虽然可以实现对目标的协同搜索，但其对环境的先验信息要求较高，当环境较为复杂时，算法适应性不强. 强化学习是一种不依赖于环境的方法，其不需要对环境进行建模，而是在智能体与环境的交互过程中实现对环境的学习和对任务模型的训练[9]. 因此，随着人工智能技术的不断成熟，很多研究学者将强化学习与目标协同搜索问题相结合，用于解决在搜索任务中可能存在的环境复杂度高、搜索任务复杂等问题. Wang 等[10]提出使用在线分布式强化学习方法实现无人机对目标的搜索与跟踪任务. Kim 等[11]提出了一种基于分布式高斯过程的多智能体强化学习算法处理目标搜索和跟踪问题，并通过硬件实验证明了该算法的有效性.

随着强化学习算法的日趋成熟，多种以其为依据的改进算法被广泛提出，并应用到多目标协同搜索的问题研究中. 但是很多研究都假设海洋环境为无限大，没有考虑在有界范围内的目标搜索问题，研究较为局限. 此外，协同搜索任务是一个包含多个评价指标的复杂优化问题，搜索时间最短和搜索区域分布最广等指标在该研究问题中是互斥的. 经典强化学习算法通过为每一个优化指标设置不同的权重系数实现奖励函数的设计，该种方法虽然可以放大重要指标在奖励函数中的作用，但是无法保证在指标互斥的情况下重要指标一定被优先满足，而基于松弛满意度的模糊多指标优化方法可以有效解决包含多个互相冲突指标的优化问题[12]. 该方法为每一个优化指标设置不同的优先级，并使用松弛满意度的方法对优化函数进行重新建模，确保每一个性能指标按照重要性顺序被依次满足[13]. 该方法与强化学习方法的有效结合可以在提高算法训练效果的前提下得到更理想的训练模型.

为此，本文针对有界未知水域环境内多目标协同搜索问题展开研究. 首先基于栅格法构建了包含环境认知度和目标存在概率两个信息指标的环境地图，为多无人船执行目标搜索任务建立了环境模型和无人船数学模型. 其次，本文在经典 D3QN (dueling double deep Q-network) 算法[14]上针对经验回放池和奖励函数做了改进，提出使用基于模糊满意多指标优化和双经验回放池的改进 D3QN (prioritized replay and fuzzy satisfactory D3QN, PFD3QN) 算法解决多目标协同搜索问题. 针对从单经验回放池随机均匀采样数据进行训练时，存在的有效数据利用率低、早期训练速度慢、后期训练稳定性差等问题，本文构建了双经验回放池分类存储数据，并根据时变的比例，从两个经验回放池采样数据. 同时，为了避免随机均匀采样数据造成训练数据随机性强、有效数据利用率低、连续数据相关性强的问题，引入优先经验回放方法使无人船更好地学习搜索环境. 针对多目标搜索问题，本文根据搜索区域内目标存在概率变化量最

大、环境搜索覆盖度最大、无人船分布性最广等优化性能指标要求，并提出了对应的 3 个奖励函数. 为满足 3 个奖励函数的重要性等级要求，使用基于松弛优先级满意度的模糊多指标优化方法对奖励函数进行重新建模. 通过改进的奖励函数进行模型训练，有效实现了对水域内目标的快速搜索，并满足了优化指标间的重要性差别要求，保证了对区域尽可能大的覆盖和无人船间的避碰. 对于提出的 PFD3QN 算法，本文通过仿真对基于模糊满意优化的奖励函数进行有效性分析，并针对不同数量目标的搜索进行算法适用性验证，进而将所提算法与标准 D3QN 算法、标准算法分别与改进经验回放池、改进奖励函数结合后的算法进行对比验证，证明了算法改进的优越性. 此外将所提算法与基于连续动作空间设计的其他方法进行了对比，证明了本文基于离散动作空间设计算法的可行性. 最后将 PFD3QN 算法用于上层规划，与无人船下层线性自抗扰控制相结合进行仿真验证，证实了算法的应用可行性，为实船实验提供了理论依据.

1 多无人船目标搜索问题

1.1 问题描述

以存在障碍物的有界水域环境为背景，本文针对多无人船搜索多个未知目标展开研究. 多无人船在执行搜索任务前，仅知道环境的边界信息、障碍物的分布信息和目标在环境内可能分布概率的先验信息. 但由于目标具体位置信息未知，且无人船的传感器探测概率存在误差，所以研究在未知不确定环境中展开. 针对该研究问题，做出如下假设：

(1) 无人船是同构的，具有相同且固定的航速；

(2) 目标静态且数量已知，但位置信息未知；

(3) 环境中障碍物是静态的，数量和位置信息已知.

考虑环境中的目标是静态的，所以无人船在搜索到一个目标后并不会静止在当前位置，而是继续搜索当前水域. 设计搜索规则如下：

(1) 一条无人船可以搜索多个目标；

(2) 一个目标可以被多条无人船搜索；

(3) 当区域内的所有目标被任意一条无人船搜索过一次后，所有无人船的搜索任务即完成.

由于目标搜索问题要求多无人船可以实现对环境中所有目标的快速搜索，并实现对区域尽可能大的覆盖，同时在整个搜索过程中避免发生碰撞. 为此，对无人船搜索设计如下指标要求：

(1) 搜索区域内目标存在概率变化量最大；

(2) 无人船对环境的搜索覆盖度最大；

(3) 无人船在搜索区域内分布性最广.

1.2 搜索环境建模

设定水域环境为大小为 $L_x \times L_y$ 的标准正方形，区域内随机分布着 N_{ua} 条无人船、N_t 个目标以及 N_o 个障碍物. 其中，障碍物小于无人船. 根据栅格法[15]对环境地图进行建模，将区域划分为 $a \times b$ 个栅格，栅格大小根据无人船的船体长度进行确定. 研究要求当无人船和目标处于同一个栅格内，认定该目标被搜索到，当所有目标被搜索到后，当前任务结束. 环境地图建模如图 1 所示，图中红色圆形表示待搜索的目标，黑色栅格表示当前位置存在障碍物，无人船在环境内初始位置随机分布.

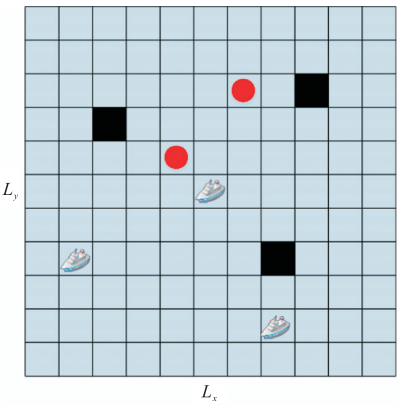


图 1 环境地图

Fig.1 Environmental map

依据图 1 栅格划分结果，为每一个栅格 (x,y) 设计环境认知度 $R(x,y)$ 和目标存在概率 $P(x,y)$ 两个环境信息指标. 其中，当栅格 (x,y) 被无人船搜索过，当前栅格的环境认知度 $R(x,y)$ 值为 1，其余情况为 0. $P(x,y)$ 表示当前栅格内目标可能存在的概率，初始值计算式为

$$P(x,y)=\frac{1}{N_t}\sum_{q=1}^{N_t}\frac{0.5}{\sqrt{\left(x_{t,q}-x\right)^2+\left(y_{t,q}-y\right)^2}+1} \tag{1}$$

式中 $\left(x_{t,q},y_{t,q}\right)$ 为第 q 个目标的位置信息. 由于环境信息已知，超出环境边界的栅格和有障碍物存在的栅格的初始目标存在概率 $P(x,y)=0$. 由于传感器探测存在误差，对无人船探测到的栅格更新目标存在概率公式[16]为

$$P_{new}(x,y)=\begin{cases}\dfrac{p_d P_{old}(x,y)}{p_f+(p_d-p_f)P_{old}(x,y)} & \text{找到目标}\\ \dfrac{(1-p_d)P_{old}(x,y)}{1-p_f+(p_f-p_d)P_{old}(x,y)} & \text{未找到目标}\end{cases} \tag{2}$$

式中：p_d 为传感器探测概率；p_f 为传感器虚警概率；

P_{old} 和 P_{new} 分别为当前栅格被探测前和探测后的目标存在概率.

1.3　无人船模型

由于无人船的六自由度数学模型控制复杂，且横摇、纵摇、垂荡对无人船运动影响较小，所以只考虑横荡、纵荡、艏摇 3 个方向的运动，采用简化的无人船的三自由度数学模型[17]，运动学模型为

$$\begin{bmatrix}\dot{x}\\ \dot{y}\\ \dot{\psi}\end{bmatrix}=\begin{bmatrix}\cos\psi & \sin\psi & 0\\ \sin\psi & -\cos\psi & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}u\\ v\\ r\end{bmatrix} \tag{3}$$

式中：$\begin{bmatrix}x & y & \psi\end{bmatrix}^{\mathrm{T}}$ 表示无人船在惯性坐标系下的位置和姿态向量；$\begin{bmatrix}u & v & r\end{bmatrix}^{\mathrm{T}}$ 表示无人船在船体坐标系下的前向速度、横漂速度和艏向角速度向量.

以由船尾左右固定螺旋桨推动的无人船作为研究对象，动力学模型为

$$\begin{cases}\dot{u}=-\dfrac{h_{11}}{m_{11}}u+\dfrac{\tau_1}{m_{11}}+\dfrac{m_{22}}{m_{11}}vr+\dfrac{h_u}{m_{11}}\\ \dot{v}=-\dfrac{h_{22}}{m_{22}}v+\dfrac{h_v}{m_{22}}-\dfrac{m_{11}}{m_{22}}ur\\ \dot{r}=-\dfrac{h_{33}}{m_{33}}r+\dfrac{\tau_2}{m_{33}}+\dfrac{h_r}{m_{33}}\end{cases} \tag{4}$$

式中：m_{11}、m_{22}、m_{33} 为惯性质量参数；h_{11}、h_{22}、h_{33} 为水动力阻尼参数；h_u、h_v、h_r 为环境扰动；τ_1、τ_2 分别为前向推力和转向力矩.

2　基于 PFD3QN 的多目标协同搜索策略

2.1　多目标协同搜索优化模型

针对 3 个优化指标的设计要求，由于无人船具有固定前向速度，通过简化无人船横漂速度和艏向角速度，其决策变量可定义为 $\{x,y,\psi\}$，由此可形成多目标协同搜索问题的优化模型为

$$\begin{cases}\max F(x,y,\psi)=\\ \quad\left[f_1^i(x_i,y_i,\psi_i),f_2^i(x_i,y_i,\psi_i),\right.\\ \quad\left.f_3^i(x_i,y_i,\psi_i)\right]\quad \forall i\in N_{\text{ua}}\\ \quad \dot{x}_i=u_i\cos\psi_i \quad \forall i\in N_{\text{ua}}\\ \quad \dot{y}_i=u_i\sin\psi_i \quad \forall i\in N_{\text{ua}}\\ \quad 0\leqslant x_i\leqslant L_x \quad \forall i\in N_{\text{ua}}\\ \quad 0\leqslant y_i\leqslant L_y \quad \forall i\in N_{\text{ua}}\\ \quad \sqrt{(x_i-x_k)^2+(y_i-y_k)^2}\leqslant d_{\text{ua}}+d_{\text{o}}\\ \qquad\qquad \forall i\in N_{\text{ua}},\forall k\in N_{\text{o}}\\ \quad \sqrt{(x_i-x_j)^2+(y_i-y_j)^2}\leqslant 2d_{\text{ua}}\\ \qquad\qquad \forall i,j\in N_{\text{ua}},i\neq j\end{cases} \tag{5}$$

式中：f_1^i、f_2^i、f_3^i 分别为第 i 条无人船的搜索区域内目标存在概率变化量、无人船对环境的搜索覆盖度和无人船的分布性；(x_k,y_k) 为第 k 条无人船的位置；u_i 和 ψ_i 分别为第 i 条无人船的前向速度和艏向角；d_{ua} 为无人船尺寸；d_{o} 为障碍物尺寸. 考虑无人船避撞与避障的安全距离约束和实际栅格建模环境，在协同搜索任务中使用栅格尺寸替代理论避撞距离 $2d_{\text{ua}}$ 和避障距离 $d_{\text{ua}}+d_{\text{o}}$.

2.2　基于双经验回放池的改进 D3QN 算法

强化学习是智能体通过与环境交互实现目标的一种方法，它是一个马尔可夫决策过程[18]. 它由 5 元组 $\{S,A,\gamma,R,P\}$ 构成，其中：S 表示状态集合；A 表示动作集合；γ 表示折扣因子；R 表示奖励函数；P 表示状态转移函数.

D3QN[14]是一种处理离散数据的强化学习方法，它针对每个智能体分别设计训练网络和目标网络，其中训练网络用于选取价值最大的动作，目标网络用于计算动作的价值实现模型训练，即

$$\begin{cases}L(\theta_{i,j})=\dfrac{1}{M}\displaystyle\sum_{m=1}^{M}\left(y_m-Q_\theta(s_m,a_m)\right)^2\\ y_m=r_m+\gamma Q'_\theta\left(s'_m,\arg\max\limits_{a'_m}Q_\theta(s'_m,a'_m)\right)\end{cases} \tag{6}$$

式中；Q_θ、Q'_θ 分别为训练网络和目标网络的价值函数；θ、θ' 为每个网络对应的参数；M 为从经验回放池采样的样本数目；s'_m、a'_m 分别为第 m 个采样数据的下一时刻状态值和该状态下的最优动作值；s_m、a_m 分别为当前智能体的状态值和动作值；L 为损失函数，用于实现网络参数的更新；r_m 为第 m 个采样数据的奖励值. 其中价值函数计算式为

$$Q(s,a)=V(s)+\left(A(s,a)-\frac{1}{|A|}\sum_{a'}A(s,a')\right) \tag{7}$$

式中：s、a 分别表示状态和动作；$V(s)$、$A(s,a')$ 分别表示状态价值函数和在该状态下不同动作获得的优势函数.

经典 D3QN 算法通过从经验回放池中随机均匀采样数据实现模型的训练，这导致有效数据在初期训练中利用率低，模型训练效果较差. 本文提出使用时变调用的共享双经验回放池对算法进行改进. 研究设计两个存储空间大小相同的经验回放池存储所有智能体的数据，每一个智能体都从共享经验池内调用信息实现独立训练. 其中经验回放池 1 存储对环境进行有效搜索的数据，经验回放池 2 存储智能体运动超出边界，或与障碍物发生碰撞的无效搜索数据. 2 个经验回放池均采用优先经验回放（prioritized ex-

perience replay, PER)[19]的方法对内部数据进行调用. 在第 j 次训练时, 经验回放池 1 和 2 被采样数据的比例为 $1-\varepsilon_j$ 和 ε_j, 当前训练轮次结束后, 可以得到第 j+1 次训练时采样数据的比例参数 ε_{j+1}, 其更新公式为

$$\varepsilon_{j+1}=\begin{cases}\varepsilon_j-\varepsilon_0 & \varepsilon_j\geqslant\varepsilon_{\mathrm{r}}\\ \varepsilon_{\mathrm{r}} & \varepsilon_j<\varepsilon_{\mathrm{r}}\end{cases} \tag{8}$$

式中: ε_0 为采样比例的每次变化量; ε_{r} 为经验回放池 2 的最低采样比例. 算法结构如图 2 所示.

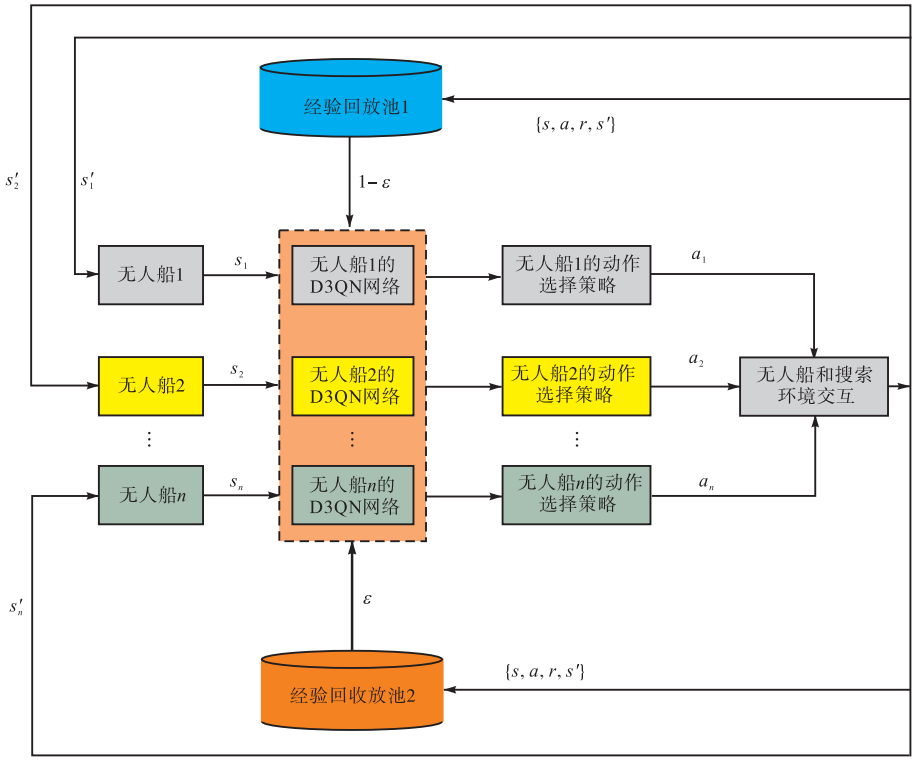


图 2 算法结构

Fig.2 Algorithm structure

图中, $\{s_1,s_2,\cdots,s_n\}$ 表示无人船的当前状态, $\{a_1,a_2,\cdots,a_n\}$ 表示无人船经过网络选择后输出的动作, $\{s_1',s_2',\cdots,s_n'\}$ 表示无人船和搜索环境交互后得到的新状态, 每一条无人船根据当前搜索的有效性, 将 $\{s,a,r,s'\}$ 存储到不同的经验回放池, 并按照不同的比例从经验回放池采样数据实现网络的训练.

2.3 状态空间和动作空间设计

设定传感器可以覆盖栅格数目 $z\times z$ 的标准正方形水域环境实现目标搜索. 每个无人船的状态空间为 $s=\{P(x,y),x_{\mathrm{ua}},y_{\mathrm{ua}}\}$ ($x=1,2,\cdots,z$; $y=1,2,\cdots,z$), 其中, $(x_{\mathrm{ua}},y_{\mathrm{ua}})$ 为当前无人船归一化处理后的位置信息.

研究针对被栅格化的环境对动作空间进行离散化处理, 设计无人船的动作为搜索方向, 动作空间为 $a=\{\psi\}$, ψ 为无人船艏向角. 由于无人船每次可以搜索一个栅格, 所以无人船的可选动作空间设计为间隔 45° 艏向角. 考虑无人船在实际航行中存在水阻力大、转向困难的问题, 设计动作空间约束为

$$\Delta\psi\leqslant\psi_0 \tag{9}$$

式中: $\Delta\psi$ 为每次无人船艏向角的变化量; ψ_0 为无人船允许的最大艏向角变化量.

基于上述算法, 由于状态空间中包含表示目标存在概率的二维地图信息, 所以使用两层卷积层和池化层对地图信息进行特征提取和训练, 并将提取结果与无人船位置信息结合, 使用全连接神经网络搭建后续网络架构. 考虑到输入数据量的大小, 将隐藏层数设置为 2 以保证训练的快速性和有效性, 并选取线性整流函数(rectified linear unit, ReLU)作为激活函数, 对于一条无人船, 完整的网络模型构建如图 3 所示, 所有无人船采用相似的网络结构进行建模.

2.4 基于模糊多指标优化的奖励函数设计

研究要求设计合理的奖励函数实现在有界水域环境中对所有目标的快速搜索, 并在此前提下实现对区域尽可能大的覆盖, 同时在整个搜索过程中避免发生碰撞. 因此针对上述多指标优化问题, 利用松弛优先级满意度对指标进行模糊优化[20]. 对于设计所需的多个指标, 研究划定了如下的 3 层优先级.

(1) 第 1 级: 搜索区域内目标存在概率变化量最大. 对于第 i 条无人船, 该性能指标函数表示为

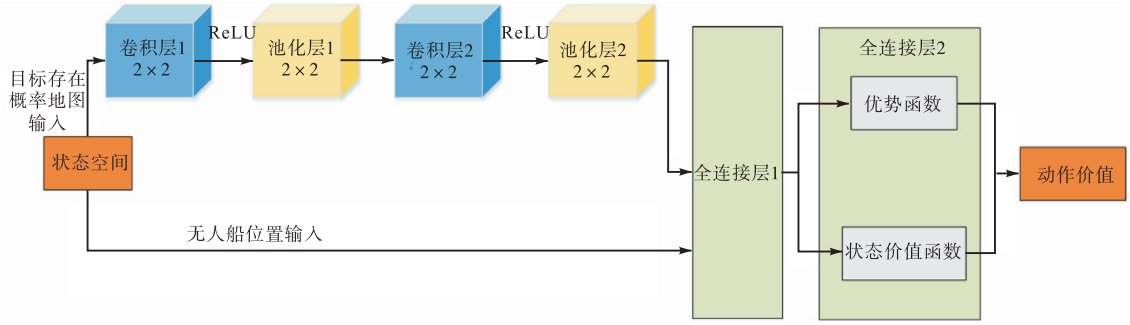


图 3　网络结构

Fig.3　Network structure

$$\max f_1^i = \frac{1}{K}\sum_{j=1}^{K} c\left(P_{\text{new}}^j - P_{\text{old}}^j\right) \tag{10}$$

式中 K 为传感器探测覆盖栅格的数量，由于搜索环境较大，但是区域内目标数量较少，为提高有效探测对指标函数值的影响，根据栅格内目标存在概率的不同变化情况选取不同的系数 c. 当 $P_{\text{new}}^j \geqslant P_{\text{old}}^j$ 时，$c = c_1$；当 $P_{\text{new}}^j < P_{\text{old}}^j$ 时，$c = c_2$. 隶属度函数表示为

$$\max \overline{f}_1^i = \mu_1^i = 1 - \frac{f_1^{i*} - f_1^i}{f_1^{i*} - f_1^{i\min}} \tag{11}$$

式中：f_1^{i*} 为搜索区域内目标存在概率变化量的期望值；$f_1^{i\min}$ 为搜索区域内目标存在概率变化量的最小值.

(2) 第 2 级：环境搜索覆盖度最大. 由于所有无人船共同执行搜索任务，所以环境搜索覆盖度为所有无人船搜索覆盖度之和，计算结果为所有无人船共享. 该性能指标函数表示为

$$\max f_2^i = \frac{1}{a \times b}\sum_{j=1}^{a \times b} R^j \tag{12}$$

隶属度函数表示为

$$\max \overline{f}_2^i = \mu_2^i = 1 - \frac{f_2^{i*} - f_2^i}{f_2^{i*} - f_2^{i\min}} \tag{13}$$

式中：f_2^{i*} 为环境覆盖度的期望值；$f_2^{i\min}$ 为环境覆盖度的最小值.

(3) 第 3 级：无人船分布性最广. 本文使用不同无人船的相对距离表示无人船的分布性，该性能指标既可以表示无人船的全局分布特征，也可以表示无人船间的避撞关系. 对于第 i 条无人船，该性能指标函数表示为

$$\max f_3^i = \frac{1}{N_{\text{u}} - 1}\sum_{j=1}^{N_{\text{u}}-1} \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \tag{14}$$

式中：(x_i, y_i) 表示第 i 条无人船的位置；(x_j, y_j) 表示其余无人船的位置. 隶属度函数表示为

$$\max \overline{f}_3^i = \mu_3^i = 1 - \frac{f_3^{i*} - f_3^i}{f_3^{i*} - f_3^{i\min}} \tag{15}$$

式中：f_3^{i*} 为无人船平均分布相对距离的期望值；$f_3^{i\min}$ 为无人船平均分布相对距离的最小值.

对于第 i 条无人船，需要使其 3 个性能指标的优先级可以满足 $\overline{f}_1^i \geqslant \overline{f}_2^i \geqslant \overline{f}_3^i$. 由于完全按照优先级规则设置目标函数容易出现忽视单个指标的寻优、最优解无法找到的情况，所以需要平衡单个指标的最优化和优先级关系被满足间的关系，因此研究采用松弛满意度方法处理优先级关系，即

$$\begin{cases} \max f^i = \dfrac{\overline{f}_1^i + \overline{f}_2^i + \overline{f}_3^i}{3} - \lambda\rho^i \\ \overline{f}_2^i - \overline{f}_1^i \leqslant \rho^i \\ \overline{f}_3^i - \overline{f}_2^i \leqslant \rho^i \\ -1 \leqslant \rho^i \leqslant 1 \end{cases} \tag{16}$$

式中：$\overline{f}_2^i - \overline{f}_1^i$ 和 $\overline{f}_3^i - \overline{f}_2^i$ 表示优先级关系；λ 为平衡每个指标最优化和优先级关系的权重系数；ρ^i 为反映优先级关系是否被满足的判断变量，$\rho^i < 0$ 表示所有优先级关系均被满足，反之则表示存在未满足的优先级关系. 使用 f^i 作为第 i 条无人船的奖励函数，由式 (16) 可计算得到无人船的综合奖励函数值，从而有效解决了传统奖励函数设计方法不能保证每一个优化指标按照期望顺序被依次满足的问题. 为了在强化学习中满足式 (16) 的约束条件，使用 $\overline{f}_2^i - \overline{f}_1^i$ 和 $\overline{f}_3^i - \overline{f}_2^i$ 的最大值代替 ρ^i 计算第 i 条无人船的奖励函数 f^i，实现算法的训练.

2.5　算法步骤

PFD3QN 算法伪代码如下.

输入：训练轮数 E，单轮训练次数 T，环境尺寸 $L_x \times L_y$，无人船数目 N_{u}，目标数目 N_{t}，障碍物数目 N_{o}，障碍物位置，无人船尺寸，经验回放池采样数据参数 ε，PFD3QN 算法参数.

输出：无人船艏向角

for $e \in [1, E]$：

初始化无人船的位置和搜索环境信息.

for $t \in [1, T]$：

根据当前网络选择每一个无人船动作 a.

执行动作，根据式(16)获得奖励值 r 和新的状态 s'.

将 $\{s,a,r,s'\}$ 储到经验回放池 1 或 2.

从经验回放池分别采样 $1-\varepsilon$ 和 ε 个数据.

根据式(6)训练网络.

根据式(8)更新 ε.

end for

end for

3 仿真验证与结果分析

3.1 仿真环境和参数设计

仿真平台为 i7-12700H CPU、RTX3070Ti GP 和 16RAM. 深度强化学习算法框架为 Pytorch. 研究假设搜索环境中存在多个静态目标、2 个静态障碍物和 3 条搜索无人船，障碍物位置固定且已知，但无人船的初始位置和目标位置随机给定，目标位置未知. 表 1 为环境信息和传感器探测信息. 表 2 为 PFD3QN 算法参数信息.

表 1 环境信息和传感器探测信息

Tab.1 Information of environment and sensor detection

水域环境/(m × m)	无人船尺寸/m	障碍物尺寸/m	障碍物位置/m
50 × 50	2	1.8	(12,44),(38,20)
网格数目	传感器探测概率 p_d	传感器虚警概率 p_f	传感器探测半径/m
25 × 25	0.9	0.1	6

表 2 PFD3QN 算法参数

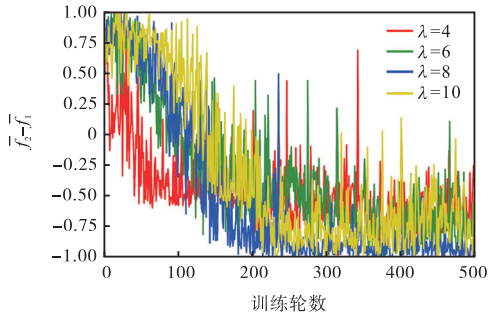
Tab.2 PFD3QN algorithm parameters

参数	参数值
训练轮数	500
单轮训练次数	500
经验回放池大小	1 000 000
批采样样本数目	128
学习率	0.000 1
折扣因子	0.99
PFD3QN 算法贪心探索参数初值	1
PFD3QN 算法贪心探索参数最小值	0.1
PFD3QN 算法贪心探索参数衰减系数	0.000 1
经验回放池调用参数初值	0.7
经验回放池调用参数衰减值	0.000 01
经验回放池调用参数最小值	0.1
第 1 级性能指标函数设计参数	$c_1=50$，$c_2=1$
优化器	Adam
优先经验回放参数	0.4

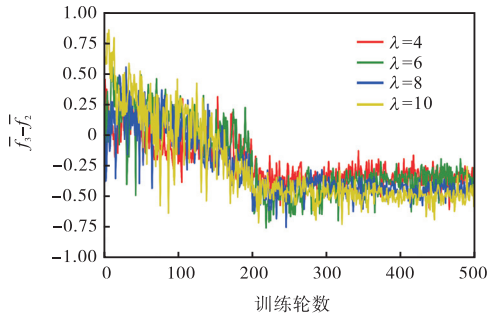
3.2 算法仿真及对比验证

3.2.1 基于优先级模糊多指标奖励函数设计验证

为验证采用基于松弛优先级模糊多指标优化方法改进奖励函数的有效性，本文在 $\lambda=4,6,8,10$ 时分别对模型进行训练，针对每一个性能指标函数将所有无人船的求和绘制得到优先级关系如图 4 所示. 图 4(a)表示 $\bar{f}_2-\bar{f}_1$ 优先级关系，图 4(b)表示 $\bar{f}_3-\bar{f}_2$ 优先级关系. 随训练轮数的增加，优先级关系曲线值逐渐递减，最后均可以稳定在小于 0 的范围内，说明通过训练 3 个优先级约束关系均可满足，且随 λ 的增加，$\bar{f}_2-\bar{f}_1$ 和 $\bar{f}_3-\bar{f}_2$ 两个优先级关系曲线值越小，符合设计要求. 综合考虑不同 λ 取值下优先级关系曲线值收敛速度和收敛值，选取 $\lambda=8$ 作为平衡单个指标最优化和优先级关系的权重系数.



(a) $\bar{f}_2-\bar{f}_1$



(b) $\bar{f}_3-\bar{f}_2$

图 4 优先级关系

Fig.4 Priority relationship

3.2.2 针对不同数量目标的搜索算法验证

为验证所提 PFD3QN 算法针对不同数量的目标均可以实现稳定搜索，本文分别对存在 $N_t=2,5,10$ 个目标的环境进行训练，奖励函数训练曲线如图 5 所示. 在给定训练周期内，无人船的奖励函数均可以实

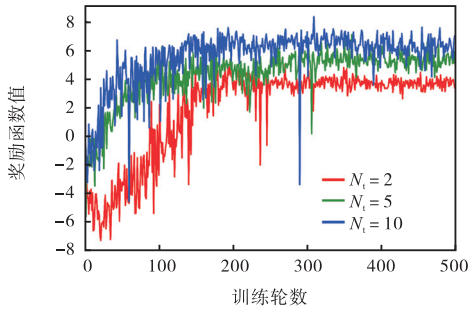


图 5 不同数量目标的奖励函数训练曲线

Fig.5 Training curves of reward function with different numbers of targets

现稳定收敛，且随着环境中目标数量的增多，搜索更容易实现. 由于状态空间只与目标存在概率地图信息和无人船位置信息有关，所以状态输入维度不随目标数量的增加而改变，训练可行性可以满足.

3.2.3　多目标协同搜索算法对比验证

为验证本文所提算法的优越性，分别选取标准 D3QN[14]算法、标准算法与改进经验回放池结合的算法（prioritized replay D3QN，PD3QN）、标准算法与改进奖励函数结合的算法（fuzzy satisfactory D3QN，FD3QN）作为对比算法进行模型训练. 此外，由于本文设计的算法基于离散动作空间实现，为验证其相对于连续动作空间算法训练的可行性，选择基于气味引导和动作偏好选择改进的强化学习算法（improved reinforce algorithm with action preference selection，IRA）[21]作为对比算法，图 6 为上述算法针对环境中存在 2 个目标的搜索任务进行训练的奖励函数曲线图. 红色、绿色、蓝色、黄色和紫色曲线分别为 PFD3QN、PD3QN、FD3QN、D3QN、IRA 算法的训练结果. 相较于对比算法，本文所提出的算法在训练前期具有更快的收敛速度，在训练后期具有更稳定的收敛效果和更大的收敛值.

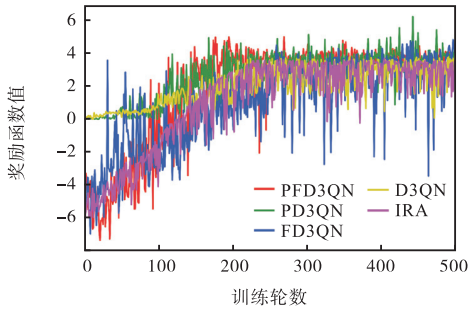


图 6　奖励函数曲线

Fig.6　Curves of reward function

3.3　结合控制系统的应用仿真验证

3.3.1　控制系统设计

考虑到无人船实际底层控制结果对上层强化学习搜索算法的影响，为验证所提算法的应用性，将上层 PFD3QN 算法搜索规划结果与下层无人船动力学控制器结合进行仿真验证. 该控制器采用了线性自抗扰控制[22]，基于式 (3) 和 (4) 设计一阶前向速度控制器和二阶艏向角控制器分别为

$$\begin{cases}\dot{\hat{u}}=\hat{g}_u+b_u\tau_1+2\omega_{o,u}\left(u-\hat{u}\right)\\ \dot{\hat{g}}_u=\omega_{o,u}^2\left(u-\hat{u}\right)\\ \tau_1=\dfrac{\omega_{c,u}\left(u_d-\hat{u}\right)-\hat{g}_u}{b_u}\end{cases}\tag{17}$$

$$\begin{cases}\dot{\hat{\psi}}=\hat{r}+3\omega_{o,\psi}\left(\psi-\hat{\psi}\right)\\ \dot{\hat{r}}=\hat{g}_\psi+b_\psi\tau_2+3\omega_{o,\psi}^2\left(\psi-\hat{\psi}\right)\\ \dot{\hat{g}}_\psi=\omega_{o,\psi}^3\left(\psi-\hat{\psi}\right)\\ \tau_2=\dfrac{\omega_{c,\psi}^2\left(\psi_d-\hat{\psi}\right)-2\omega_{c,\psi}\hat{r}-\hat{g}_\psi}{b_\psi}\end{cases}\tag{18}$$

式中：$\omega_{o,u}$ 为前向速度观测器参数；$\omega_{c,u}$ 为前向速度控制器参数；$\omega_{o,\psi}$ 为艏向角观测器参数；$\omega_{c,\psi}$ 为艏向角控制参数，期望前向速度为定值，期望艏向角为根据上层 PFD3QN 动作空间输出结果转化得到的变化值. 相关控制器参数如表 3 所示.

表 3　控制器参数

Tab.3　Control parameters

参数	参数值
前向速度控制器	$\omega_{o,u}=25$，$\omega_{c,u}=\sqrt{10}$，$b_u=76.15$
艏向角控制器	$\omega_{o,\psi}=25$，$\omega_{c,\psi}=10$，$b_\psi=20$

由于 PFD3QN 算法规划结果为针对离散栅格位置得到的离散期望艏向角，在实际控制过程中，无人船的运动轨迹为连续值，所以直接将 PFD3QN 算法输出艏向角作为控制输入存在较大误差. 因此，本文根据 PFD3QN 算法输出的艏向角和无人船当前所处栅格，得到无人船下一步的期望栅格标号，根据此栅格中心点的位置和当前实际位置计算得到期望艏向角为

$$\psi_d=\arctan\left(\frac{y_d-y_r}{x_d-x_r}\right)\tag{19}$$

式中：(x_d,y_d) 为期望栅格的中心点位置；(x_r,y_r) 为无人船的当前实际位置；ψ_d 为期望艏向角. 为避免无人船的艏向角发生较大改变，对 ψ_d 做修正，即

$$\psi_d=\begin{cases}\psi_d & -\pi\leqslant\psi_d-\psi_r\leqslant\pi\\ \psi_d+2\pi & \psi_d-\psi_r<-\pi\\ \psi_d-2\pi & \psi_d-\psi_r>\pi\end{cases}\tag{20}$$

式中 ψ_r 为当前实际艏向角.

3.3.2　PFD3QN 算法应用仿真验证

基于如上设计，本文对无人船在有界水域内执行多目标搜索任务进行测试. 其中，无人船期望前向速度为 0.6 m/s，控制周期为 5 s. 为测试无人船对环境中不同数量目标的搜索情况，研究选取目标数量小于和大于无人船数量两种情况进行仿真验证.

（1）场景 1：目标数量小于无人船数量.

针对环境中存在 2 个目标的搜索任务，图 7 为无人船搜索规划控制轨迹，L_x 和 L_y 分别为水域环境的横向和纵向尺寸，红色、绿色和蓝色曲线分别对应 3 条搜索无人船 USV1、USV2、USV3 的轨迹，圆形标

记表示无人船的出发位置,黑色矩形表示当前栅格存在障碍物,品红色矩形表示当前栅格存在目标. 根据搜索结果,无人船在步长为 21 时实现对所有目标的搜索,最后一个被搜索到的目标栅格标号为(6, 11). 其中,USV2 先实现了对标号为(13, 6)的目标栅格的搜索,由于此时环境中存在未搜索到的目标,依据研究设计的搜索规则,USV2 不会静止在当前栅格,而是继续执行搜索任务,直至 USV3 搜索到目标栅格(6, 11),此时所有目标被找到,依据搜索规则,当前所有无人船的搜索任务完成. 在上述搜索过程中,由于目标数量小于无人船数量,所以存在 USV1 未搜索到目标,但搜索任务已经结束的情况. 该无人船实现了对未知环境的探索,其搜索结果更新环境认知度信息指标,该信息指标被所有无人船共享,避免了其余无人船对无目标区域的反复搜索. 无人船在搜索过程中尽可能覆盖更多的区域,符合奖励函数的设计要求. 虽然无人船的搜索规划控制轨迹有交叉,但由于两条无人船在不同时刻到达同一地点,所以无人船之间并未发生碰撞.

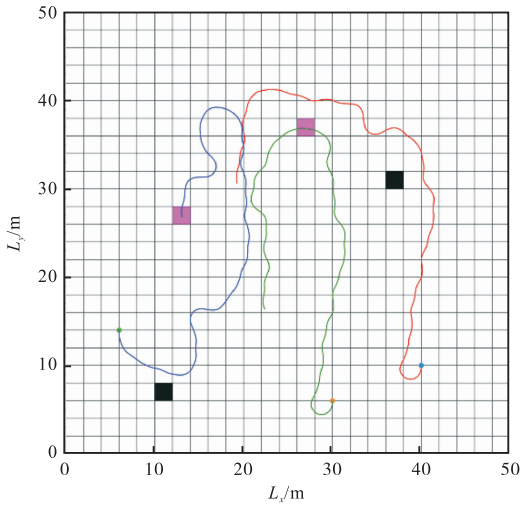


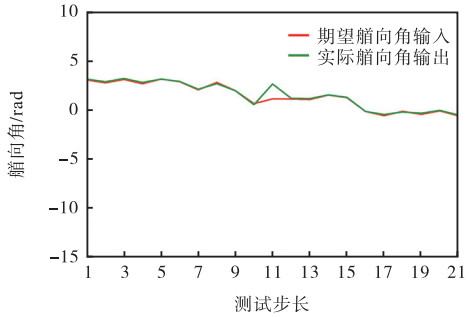
图 7 场景 1 无人船搜索规划控制轨迹($N_t=2$)

Fig.7 Search planning and control trajectory of USVs in scenario 1($N_t=2$)

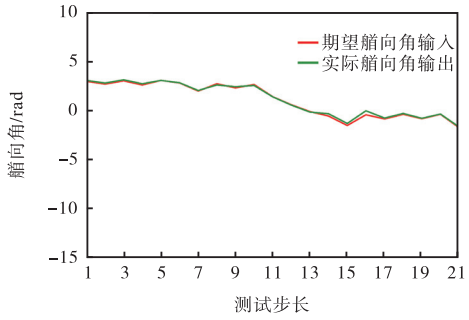
由于无人船跟踪恒定的前向速度和变化的艏向角,所以仅绘制 3 个无人船的艏向角控制器跟踪控制曲线如图 8 所示. 结果表明在搜索过程中,无人船可以实现稳定的跟踪控制.

(2) 场景 2:目标数量大于无人船数量.

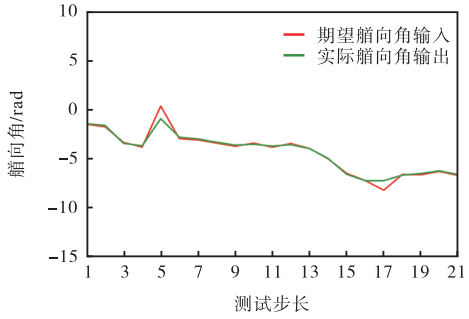
针对环境中存在 5 个目标的搜索任务,图 9 为无人船搜索规划控制轨迹,红色、绿色和蓝色曲线分别对应 3 条搜索无人船 USV1、USV2、USV3 的轨迹. 结果表明无人船在步长为 48 时实现了对所有目标的搜索,最后一个被搜索到的目标栅格标号为(7, 11).



(a) USV1



(b) USV2



(c) USV3

图 8 场景 1 艏向角控制曲线($N_t=2$)

Fig.8 Yaw angle control curves in scenario 1($N_t=2$)

其中,USV1 搜索到的目标栅格标号为(13, 6)和(7, 11),USV2 搜索到的目标栅格标号为(16, 15)、(13, 6)和(8, 16),USV3 搜索到的目标栅格标号为(8, 16)和(6, 3). 根据搜索规则,在任务执行过程中,存在一个目标被多条无人船搜索到的情况,但是直到所有目标均被搜索过一次后,当前搜索任务结束. 由于无人船会趋向于搜索目标分布多的区域,所以存在无人船在部分区域反复搜索的现象,但受奖励函数设计的影响,无人船在搜索过程中也尽可能覆盖了更多的区域,同时避免了互相发生碰撞. 根据强化学习算法的测试步长,绘制无人船艏向角控制器的跟踪控制曲线如图 10 所示. 结果表明在搜索过程中,无人船可以实现稳定的跟踪控制.

对比两种情况下搜索任务的完成情况如表 4 所示,多无人船可以适应不同数量目标的搜索任务. 受

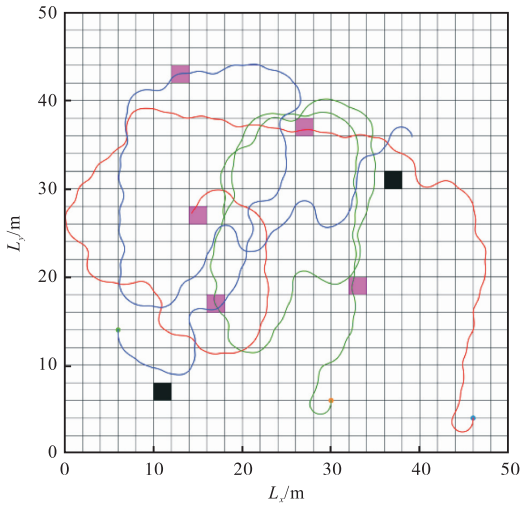
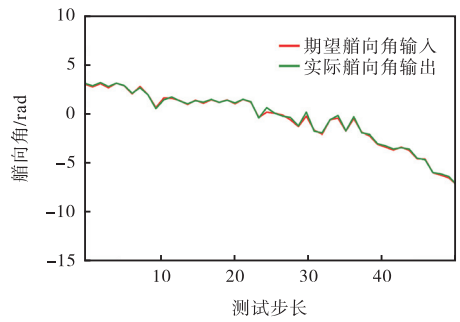
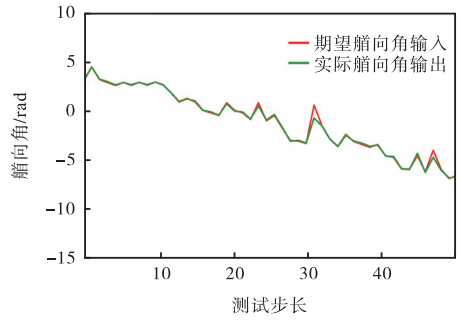


图 9　场景 2 无人船搜索规划控制轨迹（$N_t=5$）

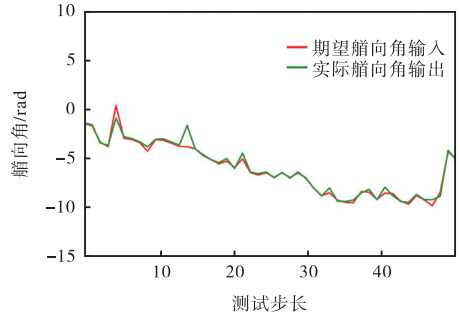
Fig.9　Search planning and control trajectory of USVs in scenario 2（$N_t=5$）



（a）USV1



（b）USV2



（c）USV3

图 10　场景 2 艏向角控制曲线（$N_t=5$）

Fig.10　Yaw angle control curves in scenario 2（$N_t=5$）

表 4　搜索任务完成情况对比

Tab.4　Comparison of search task completion

目标数	任务成功率/%	搜索步长下栅格覆盖率（实际值/期望值）/%	无人船相对距离（最小值/最大值）/m
2	100	9.8/10.1	14.1/40.3
5	100	20.0/23.0	16.1/41.2

奖励函数的影响，无人船会优先满足搜索目标存在概率变化量大的区域的搜索指标，所以存在搜索轨迹在目标分布集中的区域重复出现的问题，但受环境搜索覆盖度和无人船分布性两个指标的影响，无人船在优先满足搜索到目标的条件后，可以实现对水域尽可能大的搜索，并实现有效避碰.

3.3.3　算法应用性对比仿真验证

为验证本文提出的算法在解决多目标搜索问题上的应用性，分别再次选取 D3QN 和 IRA 作为对比算法结合控制系统进行应用仿真验证，其中无人船的初始位置、目标位置和障碍物位置设置与 PFD3QN 算法相同. 虽然 IRA 算法针对连续动作空间进行设计，但为便于算法搜索效果的对比，仍基于栅格地图绘制该算法的搜索轨迹曲线.

图 11 为使用 D3QN 算法进行目标搜索的规划控制轨迹，红色、绿色和蓝色曲线分别对应 3 条搜索无人船 USV1、USV2、USV3 的轨迹. 结果表明无人船可以在步长为 40 时实现对所有目标的搜索，最后一个被搜索到的目标栅格标号为（6，11）. 图 12 为无人船艏向角控制器的跟踪控制曲线，红色和绿色曲线分别表示期望艏向角输入和实际艏向角输出，控制器实现了对期望艏向角的稳定跟踪控制. 图 13 为使用 IRA 算法进行目标搜索的规划控制轨迹，红色、绿色

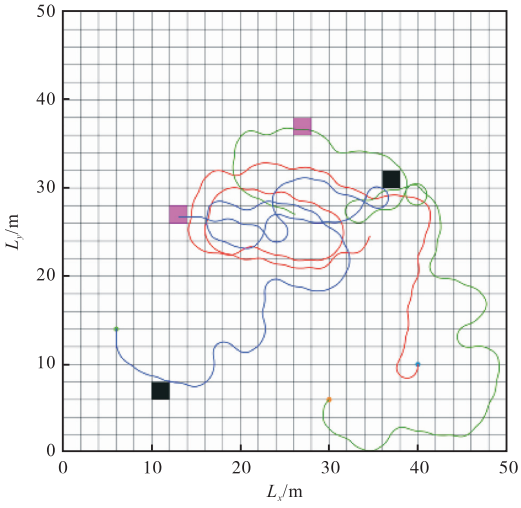
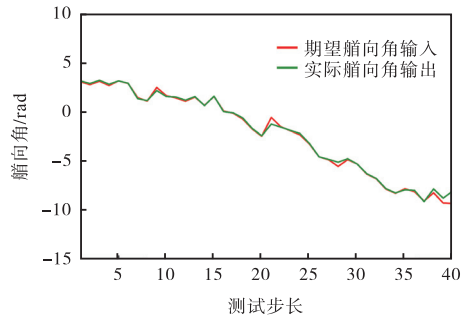
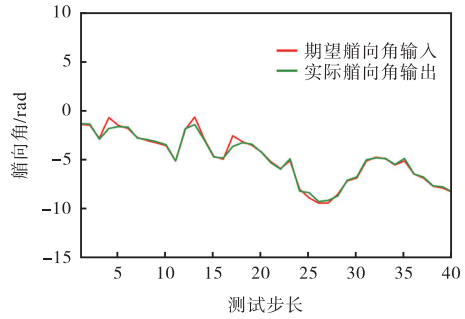


图 11　基于 D3QN 算法无人船搜索规划控制轨迹（$N_t=2$）

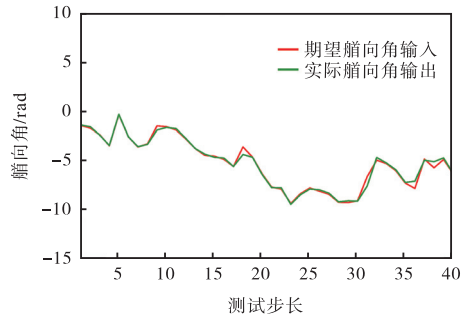
Fig.11　Search planning and control trajectory of USVs based on D3QN algorithm（$N_t=2$）



(a) USV1



(b) USV2



(c) USV3

图 12 基于 D3QN 算法艏向角控制曲线($N_t=2$)

Fig.12 Yaw angle control curve based on D3QN algorithm ($N_t=2$)

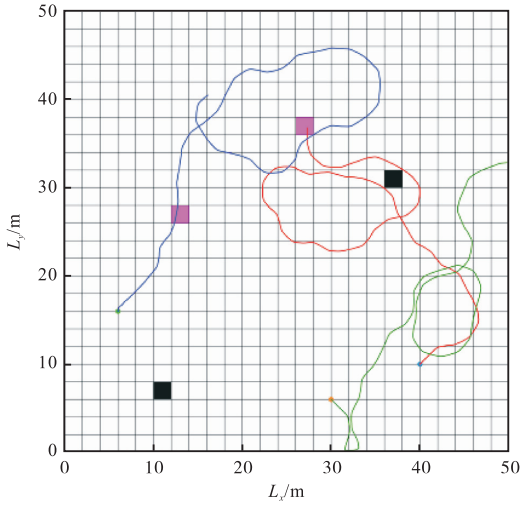
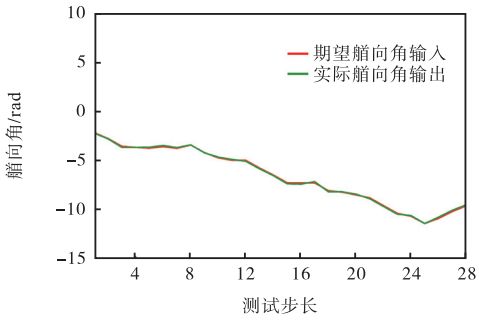


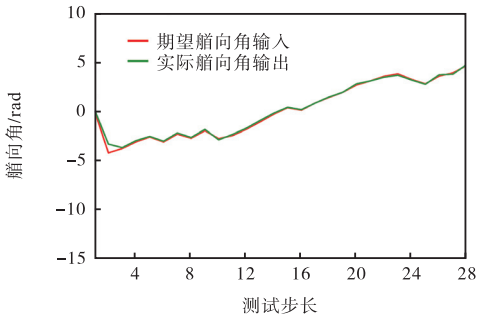
图 13 基于 IRA 算法无人船搜索规划控制轨迹($N_t=2$)

Fig.13 Search planning and control trajectory of USVs based on IRA algorithm ($N_t=2$)

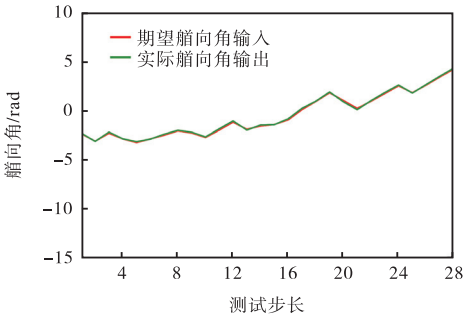
和蓝色曲线分别对应 3 条搜索无人船 USV1、USV2、USV3 的轨迹. 结果表明无人船可以在步长为 28 时实现对所有目标的搜索,最后一个被搜索到的目标栅格标号为(13,6). 图 14 为无人船艏向角控制器的跟踪控制曲线,控制器同样实现了稳定的跟踪控制.



(a) USV1



(b) USV2



(c) USV3

图 14 基于 IRA 算法艏向角控制曲线($N_t=2$)

Fig.14 Yaw angle control curve based on IRA algorithm ($N_t=2$)

对比仿真结果表明:

(1) 虽然对比算法可以实现对区域内所有目标的搜索、避障避撞以及对期望艏向角的稳定跟踪控制,然而对比算法所需步长明显大于本文所提的 PFD3QN 算法;

(2) 由于 D3QN 和 IRA 算法没有对多个优化指标进行模糊满意优化处理,仅通过不同优化指标的权重系数实现搜索要求,所以对比算法在实现快速搜索的同时容易忽略环境覆盖度这一指标,从而出现在目

标分布集中的区域反复无效搜索的情况，降低了搜索效率；

（3）相较于基于连续动作空间设计的 IRA 算法，本文基于离散动作空间设计的艏向角输出变化相对剧烈，控制器跟踪虽有误差，但仍然可以实现稳定的跟踪控制，由此可见所提算法在与控制系统结合上也是可行的.

D3QN、IRA 和 PFD3QN 算法搜索效果对比如表 5 所示，结果表明，本文所提算法在搜索速度、搜索环境覆盖度和无人船分布性上均优于两个对比算法.

表 5　D3QN、PFD3QN 和 IRA 算法对比

Tab.5　Comparison among D3QN 、 PFD3QN and IRA algorithms

算法	搜索步长	任务成功率/%	搜索步长下栅格覆盖率(实际值/期望值)/%	无人船相对距离(最小值/最大值)/m
PFD3QN	21	100	9.8/10.1	14.1/40.3
D3QN	40	100	16.9/19.2	10.1/34.4
IRA	28	100	11.7/13.4	10.8/36.1

4　结　语

为解决多无人船在未知有界水域内实现多目标协同搜索的问题，本文建立了包含环境认知度和目标存在概率两个信息指标的栅格环境模型，设计了基于模糊满意优化的 PFD3QN 算法. 使用两个经验回放池分类存储、时变调用数据改进了传统经验回放训练效率低的问题. 同时，提出使用基于松弛优先级满意度的模糊多指标优化方法设计奖励函数，得到的训练模型可以实现多无人船在水域环境内对目标的有效搜索，同时搜索覆盖尽可能大的区域，并避免发生碰撞. 最后，研究对算法性能进行仿真分析，并与线性自抗扰控制器结合进行多目标搜索的应用性仿真验证，证实了算法设计的有效性. 但是本研究只针对理想环境下的目标搜索进行仿真验证，仍有许多方面需要进一步研究，例如：对存在动态障碍物的环境展开目标协同搜索问题的研究；使用异构多无人船执行目标搜索任务.

参考文献：

[1] Liu Y，Chen C，Qu D，et al. Multi-USV system anti-disturbance cooperative searching based on the reinforcement learning method[J]. IEEE Journal of Oceanic Engineering，2023，48(4)：1019-1047.

[2] Xiao J P，Pisutsin P，Feroskhan M. Collaborative target search with a visual drone swarm：An adaptive curriculum embedded multistage reinforcement learning approach[J]. IEEE Transactions on Neural Networks and Learning Systems，2025，36(1)：317-323.

[3] Wu Y，Low K H，Lü C. Cooperative path planning for heterogeneous unmanned vehicles in a search-and-track mission aiming at an underwater target[J]. IEEE Transactions on Vehicular Technology，2020，69(6)：6782-6787.

[4] Hu X X，Liu Y H，Wang G D. Optimal search for moving targets with sensing capabilities using multiple UAVs[J]. Journal of Systems Engineering and Electronics，2017，28(3)：526-535.

[5] Miller L M，Silverman Y，Maclver M A，et al. Ergodic exploration of distributed information[J]. IEEE Transactions on Robotics，2016，32(1)：36-52.

[6] 王洪民，庄育锋，韦凌云，等. 基于信息图的多无人机三维协同搜索动目标方法[J]. 控制与决策，2023，38(12)：3534-3542.
Wang Hongmin，Zhuang Yufeng，Wei Lingyun，et al. Multi-UAV 3D collaborative searching for moving targets based on information map[J]. Control and Decision，2023，38(12)：3534-3542(in Chinese).

[7] Sun X X，Cai C，Pan S，et al. A cooperative target search method based on intelligent water drops algorithm[J]. Computers & Electrical Engineering，2019，80(4)：106494.

[8] Jia Q Y，Xu H L，Feng X S，et al. Research on cooperative area search of multiple underwater robots based on the prediction of initial target information[J]. Ocean Engineering，2019，172：660-670.

[9] Zhao C W，Li F，Hao K R，et al. A self-learning immune co-evolutionary network for multiple escaping targets search with random observable conditions[J]. IEEE Transactions on Neural Networks and Learning Systems，2020，31(10)：3853-3865.

[10] Wang T，Qin R X，Chen Y，et al. A reinforcement learning approach for UAV target searching and tracking[J]. Multimedia Tools and Applications，2019，78(4)：4347-4364.

[11] Kim J，Jang D，Kim H J. Distributed multi-agent target search and tracking with gaussian process and reinforcement learning[J]. International Journal of Control Automation and Systems，2023，21(9)：3057-3067.

[12] Qu B Y，Suganthan P N. Constrained multi-objective optimization algorithm with an ensemble of constraint handling methods[J]. Engineering Optimization，2011，43(4)：403-416.

[13] Chen L H, Tsai F C. Fuzzy goal programming with different importance and priorities[J]. European Journal of Operational Research, 2001, 133(3): 548-556.

[14] Yuan H, Ni J, Hu J B. A centralised training algorithm with D3QN for scalable regular unmanned ground vehicle formation maintenance[J]. IET Intelligent Transport Systems, 2021, 15(4): 562-572.

[15] Hou Y K, Zhao J, Zhang R Q, et al. UAV swarm cooperative target search: A multi-agent reinforcement learning approach[J]. IEEE Transactions on Intelligent Vehicles, 2024, 9(1): 568-578.

[16] 周鹤翔, 徐 扬, 罗德林. 针对动态目标的多无人机协同组合差分进化搜索方法[J]. 控制与决策, 2023, 38(11): 3128-3136.
Zhou Hexiang, Xu Yang, Luo Delin. A composite differential evolution algorithm for multi-UAV cooperative dynamic target search[J]. Control and Decision, 2023, 38(11): 3128-3136(in Chinese).

[17] 王端松, 李东禹, 梁晓玲. 干扰条件下无人艇编队有限时间同步控制[J]. 自动化学报, 2024, 50(5): 1047-1058.
Wang Duansong, Li Dongyu, Liang Xiaoling. Finite time synchronized formation control of unmanned surface vehicles with external disturbances[J]. Acta Automation Sinca, 2024, 50(5): 1047-1058(in Chinese).

[18] Alagoz O, Hsu H, Schaefer A J, et al. Markov decision processes: A tool for sequential decision making under uncertainty[J]. Medical Decision Making, 2010, 30(4): 474-483.

[19] Saglam B, Mutlu F B, Cicek D C, et al. Actor prioritized experience replay[J]. Journal of Artificial Intelligence Research, 2022, 78: 639-672.

[20] Karimi N, Feylizadeh M R, Govindan K, et al. Fuzzy multi-objective programming: A systematic literature review[J]. Expert Systems with Applications, 2022, 196: 116663.

[21] Wang X Y, Fang X. A multi-agent reinforcement learning algorithm with the action preference selection strategy for massive target cooperative search mission planning[J]. Expert Systems with Applications, 2023, 231: 120643.

[22] Cai Z H, Wang Z X, Zhao J, et al. Equivalence of LADRC and INDI controllers for improvement of LADRC in practical applications[J]. ISA Transactions, 2022, 126: 562-573.

(责任编辑:孙立华)