控制与决策
Control and Decision

基于注意力预训练自编码器的无人机集群干扰资源分配方法

张澳, 杨渡佳, 王健, 李小帅, 杨俊安, 刘辉

引用本文:
张澳， 杨渡佳， 王健， 等. 基于注意力预训练自编码器的无人机集群干扰资源分配方法[J]. 控制与决策， 2025, 40(5): 1571-1580.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2024.0814

基于注意力预训练自编码器的无人机集群干扰资源分配方法

张 澳[1]，杨渡佳[1,2†]，王 健[1]，李小帅[1,2]，杨俊安[1,2]，刘 辉[1]

(1. 国防科技大学 电子对抗学院，合肥 230037；2. 电子制约技术安徽省重点实验室，合肥 230037)

摘 要: 干扰资源分配作为认知电子战的重要环节之一, 旨在干扰资源有限的情况下, 通过合理分配干扰资源达到最大的干扰效益. 针对通信、导航受限的拒止条件下, 无人机集群协同干扰多个可移动通信目标时由于环境状态空间过大以及环境非平稳导致多智能体强化学习 (MARL) 算法决策性能较差的问题, 提出一种基于自注意力机制的预训练自编码器 (APSE), 并将其作为 MARL 算法的前置单元对环境状态进行特征提取和降维, 同时, 通过集中式训练分布式执行范式来降低环境非平稳对算法决策性能的影响. 在所建立无人机集群协同干扰仿真环境中的实验结果表明: 加入 APSE 后的 MARL 算法在平均奖励和干扰资源分配效能上提升明显. 其中: 多智能体近端策略优化算法 MAPPO-APSE 在各项指标上表现最优, 相比于 MAPPO, 其在有效干扰占空比更长的情况下干扰资源消耗量降低了 20 %.

关键词: 区域拒止；无人机集群；干扰资源分配；多智能体强化学习；注意力机制

中图分类号: TN975 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2024.0814

引用格式: 张澳, 杨渡佳, 王健, 等. 基于注意力预训练自编码器的无人机集群干扰资源分配方法 [J]. 控制与决策, 2025, 40(5): 1571-1580.

UAV swarm jamming resource allocation method based on attention-pretrained self-encoder

ZHANG Ao[1]，*YANG Du-jia*[1,2†]，*WANG Jian*[1]，*LI Xiao-shuai*[1,2]，*YANG Jun-an*[1,2]，*LIU Hui*[1]

(1. College of Electronic Countermeasure，National University of Defense Technology，Hefei 230037，China；2. Anhui Province Key Laboratory of Electronic Restriction，Hefei 230037，China)

Abstract: The allocation of jamming resources is an important aspect of cognitive electronic warfare, aimed at achieving maximum jamming effectiveness through the reasonable allocation of limited jamming resources. This paper addresses the challenges faced by multi-agent reinforcement learning(MARL) algorithms in scenarios where UAV swarms collaboratively interfere with multiple mobile communication targets under constrained communication and navigation conditions, particularly due to the expansive state space and non-stationary environment leading to suboptimal decision-making performance. We propose an attention-pretrained self-encoder(APSE) which serves as a preprocessing unit for MARL algorithms, enabling effective feature extraction and dimensionality reduction of environmental states. Additionally, we adopt a centralized training and distributed execution paradigm to mitigate the impact of environmental non-stationarity on algorithmic decision performance. The experimental results in the UAV swarm collaborative interference simulation environment established in this study demonstrate a significant improvement in average rewards and interference resource allocation efficiency with the integration of the APSE into the MARL algorithm. Among them, multi-agent proximal policy optimization APSE (MAPPO-APSE) exhibits the best performance across all metrics, reducing jamming resource consumption by 20%, while maintaining a longer effective jamming duty cycle compared to the MAPPO.

Keywords: area denial；UAV swarm；jamming resource allocation；multi-agent reinforcement learning；attention mechanism

收稿日期: 2024-07-08；**录用日期:** 2024-11-22.

责任编委: 彭木根.

†通信作者. E-mail: yangdj@nudt.edu.cn.

本文附带电子附录文件,可登录本刊官网该文“资源附件”区自行下载阅览.

0 引 言

近年来, 随着人工智能、无人技术以及信息系统的发展, 依靠无人机低成本、灵活机动、自主控制、可回收等优势, 使得无人机携带干扰载荷执行抵进干扰任务成为可能[1-2]. 然而, 受制于本身的感知和计算资源限制, 单无人机的覆盖范围有限, 且在面对复杂多变环境时其对抗和生存能力均会大打折扣. 而多个无人机组网形成的无人机集群, 通过集群间的信息共享大大扩展了任务覆盖范围, 且集群中部分无人机的故障不会导致整个系统的瘫痪, 进而提升无人机的生存对抗和整体的作战能力[3]. 但是, 随着集群规模的扩大, 其内部的合作与竞争关系愈发复杂, 且抵进干扰任务需要无人机深入电磁态势复杂的战场环境, 如在通信、导航受限的拒止环境下, 集群的决策能力将会受到很大影响. 因此, 如何实现拒止条件下的无人机集群干扰资源分配是目前亟需解决的问题.

当前用于集群智能决策的方法包括多智能体强化学习 (MARL)、基于博弈论的方法、群体智能优化算法等, 其中后两种方法在环境动态变化的场景下需要反复运算求解, 泛化性较差, 无法适应拒止条件下的智能决策[4]. 得益于深度学习和强化学习的发展[5], 近年来, MARL 已被证明是解决现实多智能体系统中非凸和高复杂性问题的有力工具[6], 在干扰资源分配上也有不少研究者使用 MARL 取得了不错的效果. 文献 [7] 提出了一种融合噪声网络的深度强化学习模型, 实现了通信对抗场景下的高效干扰资源分配; 文献 [8] 提出了一种基于最大策略熵强化学习的通信干扰资源分配方法, 通过策略熵来增强策略探索性, 从而加速收敛至全局最优; 文献 [9-10] 结合了分层学习思想和 MARL 算法, 分别实现了无人机网络中联合信道和功率分配, 以及通信干扰资源不足条件下的干扰资源分配; 文献 [11] 设计的全并行深度Q网络大大提高了大规模决策空间情况下智能干扰系统的学习速度; 文献 [12] 提出了一种基于改进近端策略优化算法的干扰资源分配方法, 实现了 3 架无人机对多个静态目标的干扰; 文献 [13] 针对无人机集群对多个异构电磁目标进行协同干扰时面临的决策问题, 提出了基于动态联盟的无人机集群协同干扰方法.

在通信干扰资源分配任务中, 文献 [7-11] 的研究场景为地面干扰站对敌方通信系统的干扰, 属于传统的通信对抗场景, 与本文研究的无人机集群抵进干扰场景不同. 实际上, 随着无人机集群在认知电子战领域的快速发展, 分布式、小功率、协同干扰所需的系统优化方法已成为现实需求[13]. 文献 [12-13] 虽然使用了无人机集群作为移动干扰站进行抵进干扰, 但是实验环境设计较为理想, 未考虑通信、导航受限对于无人机集群的影响, 且干扰目标默认为静态目标.

上述使用 MARL 进行干扰资源分配的研究大多集中于算法改进以及网络结构的优化, 忽略了 MARL 中由于智能体数量增多, 状态空间维度过大对算法决策性能的影响. 对此, 本文借鉴文献 [14] 的思路, 提出一种基于注意力机制的预训练自编码器 (APSE), 在环境状态输入算法前对其进行特征提取和降维, 以降低过大的状态空间对后续算法求解带来的影响. 同时, 采用集中式训练分布式执行 (CTDE) 范式, 以减小环境非平稳对算法训练的影响. APSE 不针对特定的 MARL 算法, 其可与绝大部分 MARL 算法相结合使用, 适用性较强. 本文主要内容如下.

1) 设计拒止条件下包含通信、干扰以及运动模型在内的无人机集群协同干扰对抗仿真环境. 同时, 将干扰资源分配过程建模为局部可观测马尔可夫决策过程 (POMDP), 将无人机映射为智能体, 建立多智能体系统下的干扰资源分配模型.

2) 提出一种基于注意力机制的预训练自编码器 (APSE), 结合多智能体近端策略优化算法 (MAPPO), 并采用 CTDE 范式进行训练, 高效求解无人机集群协同干扰资源分配的优化问题. 实验结果表明, APSE 能够缓解状态空间维度过大和环境非平稳对 MARL 算法的影响, 加入 APSE 的算法相比于原始算法, 在平均奖励和资源分配效能上提升明显. 其中: MAPPO-APSE 在 3 项指标上均取得了最优的表现; 相比于 MAPPO, MAPPO-APSE 在有效干扰占空比更长的情况下干扰资源消耗量降低了 20 %.

1 系统模型

本节首先建立包含我方无人机集群和敌方电台的系统模型; 然后, 根据无人机干扰功率大小、无人机与敌方电台间的干扰链路, 以及敌方电台通信链路的信道状态等条件建立干扰效果评估模型; 最后, 根据优化目标和约束条件, 建立本文所考虑的干扰资源分配优化问题的数学模型.

1.1 无人机模型

本系统拟采用四旋翼无人机集群执行对抗干扰任务, 无人机具有飞行速度、干扰功率和电池容量限制. 此外, 我方无人机集群具备通信能力, 能够与以

自身为圆心的一定半径内的无人机进行通信, 最大通信半径[15] 可通过下式进行计算:

$$R_{\max}=\frac{c}{4\pi f_c}10^{\frac{P_T^{\max}-P_{I,N}-T_h}{20}}. \tag{1}$$

其中: $P_T^{\max}$为无人机最大通信功率; $P_{I,N}$为无人机受到的干扰与噪声的和; T_h为接收信噪比门限, 当无人机i接收无人机k的信号的信噪比大于该门限时可正常通信.

为了模拟拒止条件下无人机通信和导航受限的情况, 本文对每台无人机能够进行信息交换的其他无人机的数量进行了限制. 同时, 将导航受限情况下由惯性导航导致的无人机自定位误差建模为均值为0、方差为 1 的高斯噪声向量[16-17], 此时, 无人机i的自定位误差ΔS_i可表示为

$$\Delta S_i=\Delta S_{i,0}+\sum_{t'=1}^{t}\Delta S_{i,t'}. \tag{2}$$

其中: $\Delta S_{i,0}$为无人机在 0 时刻的自定位误差, $\Delta S_{i,t'}$为t'时刻无人机i的累积自定位误差. 在集群状态下, 每台无人机可获得周围通信可达无人机处的目标定位信息, 通过综合估计得到最终的目标估计位置信息$\bar{S}_j$, 可表示为

$$\bar{S}_j=\dot{S}_j+\frac{\sum_{k\in\mathbb{N}_i}(\lambda_k d_{kj}+\Delta S_k)}{|\mathbb{N}_i|}. \tag{3}$$

这里: $\dot{S}_j$为目标j的真实位置信息, d_{kj}为无人机k到敌方目标j的距离, λ_k为无人机k的测向误差系数, $|\mathbb{N}_i|$为与无人机i直接相连的无人机个数.

1.2　敌方电台模型

与大多数研究中只有一类目标不同, 在本系统中敌方目标分为两类: 普通通信电台目标和时敏目标, 其中普通通信电台目标具备运动能力, 可在区域内随机移动. 在未受干扰的情况下, 每个电台目标能够与其他$N-1$个电台进行周期性通信.

时敏目标为存续时间较短的通信目标, 不具备运动能力, 在系统运行初期处于静默状态, T_{tct}时刻后出现, 随后维持T'_{tct}时间后消失 ($T'_{\text{tct}}<T'_{\text{radio}}$). 时敏目标出现的时间短, 更随机, 威胁性也更大. 因此, 在时敏目标出现的时间范围内, 需要对其赋予更大的干扰优先级, 干扰成功也会获得更大的奖励. 本文通过对两类目标设置不同的奖励值, 可以测试资源分配算法在不同情况下资源分配的智能性.

本系统中假设通信电台为组网目标, 使用地地通信信道模型[15] 对敌方通信电台的通信信道进行建模, 电台j_1与j_2间的路径损耗模型可表示为

$$\text{PL}_{j_1j_2}=\text{PL}_0+10n\log_{10}d_{j_1j_2}+X_\sigma. \tag{4}$$

其中: PL_0为参考路径损耗, n为路径损耗指数, $d_{j_1j_2}$为j_1与j_2间的距离, X_σ为服从正态分布的阴影衰落.

1.3　无人机集群协同干扰系统模型

本文所研究的系统考虑在5 km × 5 km的区域内执行N架无人机对M个敌方电台的干扰任务. 无人机集群协同干扰组网电台系统如图 1 所示. 我方无人机固定在$h=1$ km的高空对敌方电台实施干扰. 无人机集合为$n=\{1,2,\dots,i,\dots,N\}$, 敌方电台集合为$m=\{1,2,\dots,j,\dots,M\}$. 初始时刻, 我方无人机和敌方电台目标在固定区域生成, 仿真过程中无人机根据基准任务分配和航迹规划算法向任务目标飞行, 敌方电台在区域内随机移动. 在整个任务过程中无人机通过干扰资源分配算法自动控制其干扰功率.

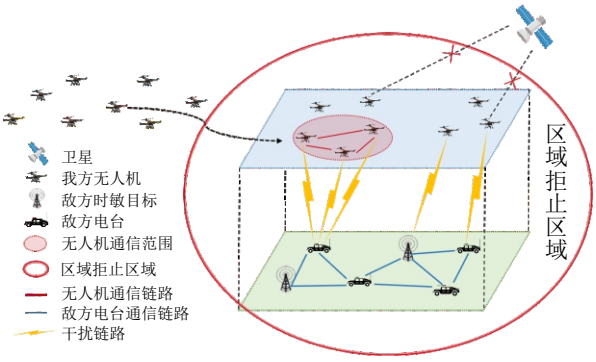


图1　无人机集群协同干扰组网电台系统示意图

本文中我方无人机使用垂直向下的定向天线进行干扰, 可通过提高发射天线功率和增益的方式来提升干扰效果. 假设无人机i对敌方电台j实施干扰, 无人机发送干扰信号功率为P_{ij}^t, 敌方电台接收到的干扰信号功率为P_{ij}^r, 功率单位均为dB · m. 利用弗里斯传输方程[18] 可构建P_{ij}^t与P_{ij}^r间的关系为

$$P_{ij}^r-P_{ij}^t=G_i^t+G_j^r+\text{PL}_{ij}. \tag{5}$$

其中: G_i^t为无人机i天线增益, G_j^r为敌方电台天线增益, PL_{ij}为传播路径损耗. G_i^t和G_j^r的大小与天线方向以及双方的相对位置有关, 其计算方式将在第 1.4 节中介绍. 在自由空间下, 信号路径损耗与无人机和目标间的欧氏距离成正比. 考虑敌方目标周围环境比较复杂, 信号传播存在多径效应, 信号的路径衰减可推广为与距离的$\tau\in[3,5]$次方成正比[15], 此时, 干扰信号的路径损耗可建模为

$$\text{PL}_{ij}=10\tau\log_{10}d_{ij}+K_{\text{pl}}. \tag{6}$$

这里: K_{pl}为路径损耗常数, $d_{ij}=\|s_i-s_j\|$为无人机

i与敌方目标j的距离, $\|\cdot\|$表示对向量求 2 范数.

1.4 无人机集群干扰效果评估模型

本文所构建的系统中, 敌方通信电台不仅受到我方无人机的干扰, 还受到其他单位非目标信号以及环境噪声的影响, 因此, 采用信干噪比作为干扰效果的评估指标更加合理. 假设目标电台接收到的多个干扰信号符合线性叠加条件, 故目标电台j_2所接收到的信号信干噪比为

$$\mathrm{SINR}_{j_2}(P^I_{ij_2})=10\lg\frac{P^r_{j_1j_2}\times G^t_{j_1}\times G^r_{j_2}}{\mathrm{PL}_{j_1j_2}\times\left(\sum\limits_{i=1}^{N}P^I_{ij_2}+N_{\mathrm{noise}}\right)}. \tag{7}$$

其中: $P^r_{j_1j_2}$为目标电台j_2接收的通信信号功率; $G^t_{j_1}$为电台j_1的发射天线增益; $G^r_{j_2}$为目标电台j_2的接收天线增益; $P^I_{ij_2}$为目标电台j_2处接收到的无人机i发出的干扰功率; N_{noise}表示均值为 0, 方差为σ^2的高斯白噪声. 此处的功率单位为 W. $P^I_{ij_2}$的大小与无人机i发出的干扰功率以及无人机对电台j_2的天线增益$G^t_{ij_2}$大小有关, $G^t_{ij_2}$的大小可通过辐射方向图来描述, 垂直方向下时最大, 且随着无人机i与电台j_2间的夹角θ_{ij_2}的增大而减小, 可通过下式进行计算:

$$\begin{cases}G^t_{ij_2}=10^{\frac{-30\times\theta^2_{ij_2}}{10}},\ \theta_{ij_2}\leqslant 60^\circ;\\ G^t_{ij_2}=0,\ \theta_{ij_2}>60^\circ.\end{cases} \tag{8}$$

综合考虑干扰波束的天线增益和信号传输损耗, 可得到$P^I_{ij_2}$的计算方式为

$$P^I_{ij_2}=\frac{P^t_{ij_2}\times G^t_{ij_2}}{\mathrm{PL}_{ij_2}}. \tag{9}$$

为了保证无人机干扰的有效性, 经我方估测, 当敌方电台j的某个通信链路接收信号的信干噪比$\mathrm{SINR}_j\leqslant H$时, 认定干扰有效, 该链路将会被切断, 其中$H$为实施有效干扰的信干噪比阈值.

1.5 无人机集群干资源分配优化模型

本文研究拒止条件下, 通过无人机集群协同干扰资源分配, 以实现干扰效益最大化的优化问题. 干扰效益最大化是指在实现有效干扰的同时, 最小化使用的干扰功率. 为此, 所建立干扰资源分配优化模型如下所示:

$$\mathbb{F}[\mathrm{SINR}_j(P^I_{ij}),\omega_i]=\sum_{j=1}^{M}\omega_i\left(\lambda_1\times\mathrm{sgn}(H-\mathrm{SINR}_j)-\lambda_2\times\sum_{i=1}^{N}P^I_{ij}\right). \tag{10}$$

其中: λ_1和λ_2为权重常数, 且λ_1、$\lambda_2\in[0,1]$, 用于表示干扰有效性以及资源消耗量之间的相对重要程度; ω_i为通信链路威胁系数.

综上所述, 拒止条件下的干扰资源分配问题可建模为带约束条件的组合优化问题, 具体可表示为

$$\begin{aligned}&\max_{P^t_{ij}}\ \mathbb{F}[\mathrm{SINR}_j(P^I_{ij}),\omega_i].\\ &\text{s.t. } C1: 0\leqslant P^t_{ij}\leqslant P_{\max},\ i\in\mathbb{N},\ j\in\mathbb{M};\\ &\qquad C2:\sum_{j=1}^{M}P^t_{ij}=P^t_{ik},\ i\in\mathbb{N},\ k\in\mathbb{M};\\ &\qquad C3:\frac{L_i}{N-1}\leqslant\mathrm{ratio},\ i\in\mathbb{N}.\end{aligned} \tag{11}$$

其中: $\mathbb{F}[\cdot]$为资源分配函数, 与敌方电台j的信干噪比有关, $\mathbb{F}[\cdot]$越高干扰效益越大; $C1$为无人机发射的干扰功率大小约束; $C2$为干扰目标数量约束, 即每个无人机只能同时干扰一个目标; $C3$为无人机集群全局信息交互约束, 即能够与无人机i通信的无人机数量不能超过L_i.

2 基于 APSE 的无人机集群干扰资源分配方法

所建立系统模型中环境状态随着无人机集群干扰动作的改变而改变, 下一时刻的状态只取决于当前的状态和干扰动作, 且各无人机的决策只依赖于各自的局部观测, 符合部分可观测马尔可夫决策过程的定义[19]. 为此, 本节首先建立无人机集群协作干扰的部分可观测马尔可夫过程模型, 然后详细介绍所提出 APSE 原理和结构.

2.1 无人机集群协作干扰的局部可观测马尔可夫过程模型

一个局部可观测马尔可夫过程可表示为七元组形式$\Gamma=(\mathcal{S},\mathcal{N},\mathcal{A},\varepsilon,P,\mathcal{R},\gamma)$, 其中:

1) $\mathcal{S}$为状态空间集, 且$\forall i\in\mathcal{N}, S_i\in\mathcal{S}$. 本文中智能体的观测状态由 3 个部分组成, 分别为t时刻无人机i估计的敌方电台的位置信息$M_i(t)$、无人机间的相对位置和通信情况$G_i(t)$以及上一次无人机的干扰策略$E_i(t)$. 对于无人机i通信不可达目标其对应部分的值设置为 0. 按照以上设置, 无人机的局部观测为$S_i=[M_i(t),E_i(t),G_i(t)]^{\mathrm{T}}$, 全局的环境状态为$S=[S_1,S_2,\ldots,S_n]$.

2) $\mathcal{N}$为智能体集合, 此处定义为无人机集群集合, $n=\{1,2,\ldots,N\}$.

3) $\mathcal{A}$为动作空间集, $\forall i\in\mathcal{N}, a_i\in\mathcal{A}_i$定义为智能体$i$采取的动作, $\mathcal{A}_i$为智能体i可以采取的动作集合. $\mathcal{A}=\otimes\mathcal{A}_i(\forall i\in\mathcal{N})$, $\otimes$为笛卡尔积. 在本文中无人机动作为量化后的干扰功率, 分为 5 个等级, 等级越大, 干扰功率越高.

4) P为状态转移概率, $P(S_{t+1}|S_t, A_t): \mathcal{A}\times\mathcal{S}\to\mathbb{R}$. 表示无人机集群在状态$S_t$采取联合动作$A_t=[a_1^t, a_2^t, \ldots, a_n^t]$, 转移至下一状态$S_{t+1}$的概率.

5) $\mathcal{R}$为奖励函数集合, 智能体i在t时刻的奖励可表示为$r_i^t(s,a): \mathcal{A}_i\times\mathcal{S}\to\mathbb{R}$. 本文中无人机的奖励值与其消耗的干扰功率、产生的干扰效果以及干扰目标类型有关, 按照以上原则设计如下奖励函数:

$$\mathcal{R}=\sum_j^M\Big[4T_j\omega_j\Big(\lambda_1\times\mathrm{sgn}(H-\mathrm{SINR}_j)-\lambda_2\times\sum_{i=1}^N P_{ij}^t\Big)\Big/\mathrm{sum}_j\Big]. \tag{12}$$

其中: T_j为敌方目标j的类型, 普通电台目标取 1, 时敏目标取 2; sum_j为参与干扰敌方目标j的无人机个数; $\mathrm{sgn}(\cdot)$为符号函数, 当$H-\mathrm{SINR}_j>0$时, 表示无人机集群成功干扰敌方目标j. 在成功干扰的情况下, 无人机消耗的功率越小, 获得的奖励越大.

6) $\gamma\in[0,1]$, 为奖励折扣因子, 用于权衡过去采取的动作对当前奖励值的影响程度.

2.2　基于注意力机制的预训练自编码器 APSE

考虑到拒止条件下无人机集群的干扰任务场景较为复杂, 环境状态维度大, 对于无人机的信息处理和特征理解区分能力要求更高, 所提出 APSE 的主要目的如下: 1) 利用自编码器对环境状态完成特征降维; 2) 利用注意力机制计算环境状态中各子状态间的关系, 提高自编码器的特征提取能力; 3) 利用预训练完成对自编码器的初始化. 本节首先介绍 APSE 的整体架构, 最后介绍 APSE 的预训练方法.

2.2.1　APSE

所提出 APSE 是一种基于注意力机制的预训练自编码器, 用于 MARL 中不同智能体局部观测的信息聚合和特征提取, 由编码器和解码器两个部分组成, 基本结构如图 2 所示. 其中: 编码器负责将输入的高维环境状态S编码为低维的特征H, 解码器对H进行重构得到输出S'. APSE 的最终目的是使得输出S'尽可能地接近S. 在训练过程中会迫使编码器提取出S中信息量较大的本质特征, 以便解码器能够更好地恢复原始输入.

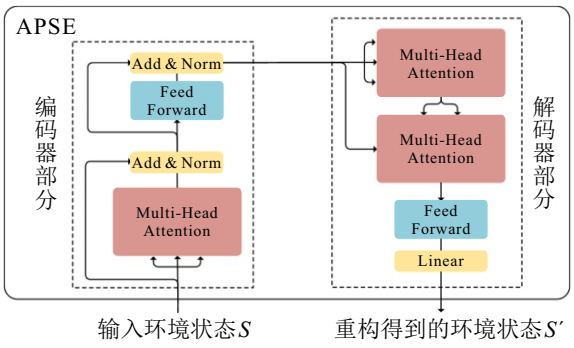


图2　APSE 基本结构

APSE 编码器部分采用与 Transformer[20] 中编码器部分相似的结构, 不同之处在于省略了词嵌入操作直接将环境状态作为编码器输入. 图 2 中: Multi-Head Attention 为多头注意力机制, 通过计算输入环境状态间的注意力得分来得到状态间的关系, 并区分出较为重要的特征; Add & Norm 为残差连接和层归一化操作, 用于缓解深层网络中的梯度消失和网络退化问题; Feed Forward 为多层前馈神经网络, 主要用于特征维度的放缩.

在本文研究的无人机集群通信对抗场景中, APSE 的输入为环境状态, 通过智能体与环境的交互便可得到, 能够节省大量的数据收集、标注和处理时间. 对于环境状态矩阵$S=[s_1, s_2, \ldots, s_n]$, 通过注意力机制计算各子状态间的关系, 可为算法提取出更优的状态表征. S中的每个子状态$s_i\in S$分别与 3 个系数矩阵相乘得到其Q、K、V值的初始表示, 即

$$\begin{aligned} Q&=W^qX,\\ K&=W^kX,\\ V&=W^vX. \end{aligned} \tag{13}$$

其中: Q(query) 为查询向量, 用于与其他子状态进行匹配; K(key) 被用于与Q进行匹配, 可理解为子状态的关键字; V(value) 为子状态的重要信息或特征. 状态矩阵S中任意子状态s_i与其他子状态间的关联程度可通过Q和K计算得到, 计算方式如下所示:

$$\mathrm{score}=\frac{Q\cdot K}{\sqrt{d}}. \tag{14}$$

这里: score为计算得到的自注意力得分, $\cdot$为点乘操作, d为Q和K的矩阵维度. 在自注意力机制中这两个矩阵维度一致. 除以$\sqrt{d}$的目的是防止点乘结果过大. 随后, 通过softmax函数将注意力得分归一化为注意力权重. 最后, 将注意力权重与V相乘得到自注意力机制的输出, 有

$$\mathrm{Attention}(Q,K,V)=\mathrm{softmax}\Big(\frac{QK}{\sqrt{d_k}}\Big)V. \tag{15}$$

在采用 Actor-Critic 结构的 MARL 算法中, Actor 网络接收智能体所观测到的局部状态, 而 Critic 网络为中心化网络, 接收环境的全局状态, 全局状态维度更大, 且相对更加复杂. 因此, 本文针对网络处理的数据维度的不同, 设计不同的 APSE, 用于 Critic 网络的 APSE 注意力头的数量越多, 输出的特征维度越大. 需要指出的是, APSE 在正式用于解决干扰资

源分配问题时只使用预训练后的编码器部分.

2.2.2 APSE 的预训练

预训练是所提出 APSE 中不可缺少的一步. 一方面, 直接将随机初始化的 APSE 加入 MARL 算法中会因为特征提取得不正确而导致训练不稳定; 另一方面, APSE 位于整个神经网络的最前端, 反向传播时的梯度小, 参数更新慢, 导致直接训练 APSE 的效果不佳.

针对上述问题, 所提出 APSE 首先通过预训练学习环境状态中的共性特点, 然后将经过预训练的 APSE 编码器部分纳入 MARL 算法中, 通过智能体与环境交互产生的奖励对编码器进行微调, 以实现特性学习的目的, 从而缓解随机初始化 APSE 所带来的训练不稳定问题.

在预训练过程中, 自编码器的主要目标是提取出最有利于重构原始输入的特征, 因此, APSE 的优化目标为

$$L(\varphi) = \min \mathbb{E}|S - S'|^2. \tag{16}$$

其中: S为环境输出的状态, S'为 APSE 重构的环境状态, 优化函数的目标是尽可能地使得两者的差距最小. APSE 的预训练伪代码如算法 1 所示.

算法 1 APSE 预训练算法.

输入: 网络仿真参数, 仿真环境返回的环境状态 S;

输出: 重构的环境状态S'.

1. 初始化环境, APSE 网络参数ϕ.
2. for $i = 1$ to episodes:
3. for $i = 1$ to episode_length:
4. 智能体随机执行动作a, 环境返回状态 $s_t = \text{env.step}(a)$
5. 保存数据至缓存池
6. end
7. for j in range (episode_length/batch_size):
8. 从缓存池数据中加载一批训练数据 data
9. 将训练数据送入 APSE 计算得到S'
10. 由式 (16) 计算损失
11. 利用 Adam 优化器更新优化目标函数 $L(\phi)$中的ϕ
12. end
13. end

APSE 在预训练过程中用于训练的环境状态数据可以是随机动作产生的, 也可以是智能体与环境交互产生的. 这保证了后续与 MARL 算法结合时输入输出的一致性, 同时还易于实现.

2.3 基于 APSE 的 MAPPO 算法框架

APSE 通过 MARL 进行端到端的训练, 可用于所有 MARL 算法. 在本文中使用 MAPPO[21] 算法作为样例, 实现基于 APSE 的无人机集群协同干扰资源分配.

MAPPO 是当前多智能体强化学习中应用最广泛的算法之一, 采用 Actor-Critic 架构和 CTDE 范式, 通过训练为每个智能体寻找最优策略π_θ和值函数 $V_\varphi(s)$. 在训练过程中, MAPPO 利用 Critic 网络来减少$V_\varphi(s)$的方差, 同时, 通过该网络整合的全局信息使得各单独的智能体能够相互配合, 缓解多智能体系统中的环境非平稳性.

MAPPO-APSE 算法的基本框架如图 3 所示. APSE 作为编码器, 对环境状态进行特征提取和压缩后输入 Critic 和 Actor 网络. 整个算法分为数据收集和算法更新两个阶段. 其中: 数据收集阶段用蓝色虚线描述, 算法更新阶段用红色实线描述. 数据收集阶段通过智能体与环境交互产生奖励、新的环境状态等数据并保存于数据缓存池. 当缓存池中的数据量到达阈值时进入算法更新阶段, 该阶段利用之前采集的数据分别计算 Critic 和 Actor 网络的损失, 同时进行网络参数的更新.

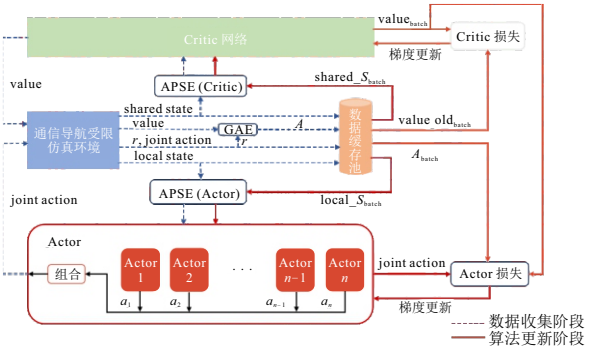


图3 MAPPO-APSE 基本框架

3 仿真实验和对比分析

本节对比 MAPPO、IQL[22]、COMA[23] 算法以及分别加入 APSE 后的 3 种算法在仿真环境中的实验结果以验证 APSE 对于 MARL 算法的决策性能提升的有效性以及拒止条件下采用 CTDE 范式相对于 DTDE 范式的优势. 在测试阶段, 从平均资源消耗量以及有效干扰占空比两个方面对算法决策性能进行评估, 并对算法的复杂度以及实时性进行分析和测试. 最后, 通过消融实验来验证自注意力机制、预训练等机制对于 APSE 性能的影响.

3.1 仿真参数设置

实验过程中设定的仿真环境参数如表 1 所示. 在该场景中, 我方 32 架无人机组成集群执行对 10

个敌方电台目标的干扰任务, 其中敌方目标包含 8 个普通电台目标和 2 个时敏目标. 此外, 初始阶段我方无人机和敌方目标均在5 km × 5 km矩形区域的固定位置处生成, 我方无人机的通信半径为 800 m, 敌方电台目标的运动加速度服从均值为 0、方差为 0.03 的高斯分布. 仿真过程中用到的计算机硬件为 Intel i5-12400F CPU, 32 GB RAM, NVIDIA GeForce RTX 4060 Ti GPU.

表1　仿真环境参数设置

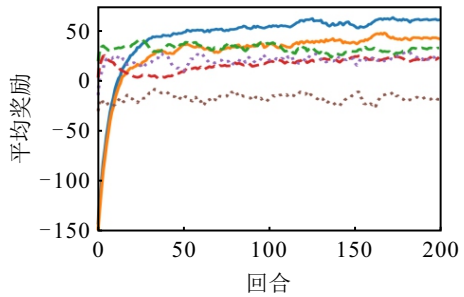
参数	数值
信干噪比阈值H	−3 dB
干扰有效性指标的相对重要程度λ_1	1
资源消耗量指标的相对重要程度λ_2	0.3
我方无人机最大干扰功率P_i^t	10 dB
我方无人机最大飞行速度$v_{\max}^i$	30 m/s
敌方收发天线增益G_j	3 dB
敌方电台最大移动速度$v_{\max}^j$	10 m/s

3.2　实验分析

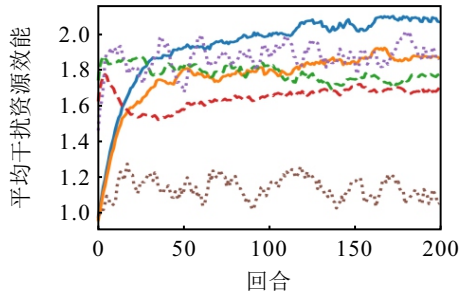
为了保证实验的可靠性, 本文进行的所有仿真实验均在 3 个不同的随机数种子下分别进行 3 次, 取 3 次实验的平均值为最终实验结果.

3.2.1　模型训练结果与分析

训练结果如图 4 所示. 其中: 干扰资源效能为无人机获得的奖励与其消耗的干扰资源的比值, 该指标越大, 无人机的干扰资源利用率越高, 与第 1.5 节所建立的优化目标一致.



(a)　训练阶段平均奖励对比



(b)　训练阶段平均干扰资源效能对比

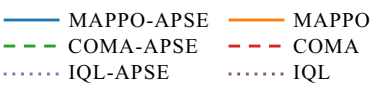


图4　训练过程中的奖励值与干扰效能比较

由图 4 可见, 加入 APSE 后的 MARL 算法相比于原始算法在奖励值和干扰资源效能上提升明显. 从网络结构上分析, 传统的全连接网络在处理输入的环境状态时并未对其提取的特征做区分, 而 APSE 中自注意机制的特征加权方式使得各智能体能够在复杂的环境状态中关注到对于自己而言更加重要的部分, 从而做出更合理的决策, 这也验证了 APSE 对于 MARL 算法性能的提升是有效的.

对比采用 DTDE 范式的算法 (IQL 类) 与采用 CTDE 的算法 (MAPPO 类、COMA 类) 可以发现, CTDE 范式下的算法奖励值远高于 DTDE 范式. 这是由于在本文所设置的仿真环境中, 智能体数量较多, 环境状态原本带有误差, 环境不稳定性较强, 而采用 CTDE 范式的算法能够凭借 Critic 网络在训练阶段得到全局状态信息, 并为 Actor 网络的决策提供指导, 一定程度上缓解了环境的不稳定问题. 另外, 在图 4(b) 中可以注意到, IQL-APSE 在干扰资源分配效能上甚至超过了 MAPPO 算法, 进一步表明了 APSE 对于智能体提取环境状态中的关键信息有很大帮助.

3.2.2　模型测试结果与分析

为了进一步验证 APSE 对于 MARL 算法决策的性能提升效果, 本文在测试阶段对比了各算法在平均资源消耗量以及有效干扰占空比两个指标上的测试结果. 其中: 有效干扰占空比为每轮测试中, 无人机干扰导致敌方电台全体通信失效的时间占整个任务时长比值的平均值, 占空比越大, 干扰任务完成度越好, 是评判干扰资源分配算法决策性能好坏的第 1 标准; 平均资源消耗量为每轮测试中无人机的干扰资源消耗情况, 结合前一指标可以反映算法在干扰资源分配上的效能. 有效干扰占空比越长, 消耗的干扰资源越少, 算法决策性能越好. 测试轮次为 200 轮, 每轮测试步长为 700. 表 2 为各算法在这两个指标上的测试结果.

由表 2 的测试结果可见, 在加入 APSE 后, 各算法在有效干扰占空比上均有所提升. 其中: MAPPO-

表2　算法测试结果

算法	有效干扰占空比/%	平均干扰资源消耗量
MAPPO-APSE	11.8	37.09
COMA-APSE	12.4	42.05
IQL-APSE	4.47	20.78
MAPPO	11	46.40
COMA	8.15	44.98
IQL	0	4.58

APSE 在有效干扰占空比比 MAPPO 更长的情况下干扰资源消耗量降低了 20 %, 符合第 1.5 节中建立的干扰资源分配优化模型的目标; COMA-APSE 相比原算法在有效干扰占空比上提升了 4.24 %, 同时干扰资源消耗量降低了约 6 %.

尽管 IQL 的平均资源消耗量很少, 但是, 无法有效干扰敌方目标带来的低能耗是没有意义的, 而在加入 APSE 后其干扰性能提升明显, 但是仍然与采用 CTDE 范式的算法有较大差距, 进一步表明了 CTDE 在拒止条件下的优越性.

综合训练和测试结果来看, MAPPO-APSE 在本文所设置的仿真环境中取得了最佳表现, 其在平均奖励、干扰资源分配效能以及平均干扰资源消耗量 3 个指标上最优, 有效干扰占空比稍逊于 COMA-APSE.

3.3 算法复杂度与实时性分析

所提出 APSE 为基础的框架创新, 并未对 MARL 算法本身进行改进, 加入 APSE 前后算法的计算复杂度的改变只与网络参数量有关. 因此, 本文借鉴文献 [7] 的思路, 采用网络参数量级来衡量算法的计算复杂度. 假设一个神经网络参数的计算复杂度为ℓ, 输入状态维度为$x_{\rm in}$, 输出维度为$x_{\rm out}$, 隐藏层维度为$x_{\rm h}$, 则 APSE 的计算复杂度为

$$
\begin{aligned}
&O(\text{APSE}) = \\
&O(\text{FC}_{\rm in}) + O(\text{MHA}) + O(\text{FFNN}) + O(\text{FC}_{\rm out}) = \\
&O(\ell(x_{\rm in}\cdot x_{\rm h})) + O(\ell(12x_{\rm h}^2 + 2x_{\rm h}))+ \\
&O(\ell(8x_{\rm h}^2 + 2x_{\rm h})) + O(\ell(x_{\rm h}\cdot x_{\rm out})) = \\
&O(\ell(x_{\rm in}\cdot x_{\rm h} + 20x_{\rm h}^2 + 4x_{\rm h} + x_{\rm h}\cdot x_{\rm out})).
\end{aligned}
$$

同样的算法, 但是使用全连接层处理输入数据时的计算复杂度为

$$
\begin{aligned}
O(\text{MLP}) = O(\text{FC}_{\rm in}) + O(\text{FC}_{\rm h}) + O(\text{FC}_{\rm out}) = \\
O(\ell(x_{\rm in}\cdot x_{\rm h} + 3x_{\rm h}^2 + x_{\rm h}\cdot x_{\rm out})).
\end{aligned}
$$

可见在引入 APSE 后, 算法的计算代价相对更高. 为了进一步测试 APSE 计算复杂度增加带来的时间开销, 同时衡量算法的实时性能, 本文测试了各算法从输入环境状态到输出最终决策结果所用时间, 并将该测试指标定义为单次决策用时, 测试结果如表 3 所示.

表3 算法测试运行用时

算法	用时/ms	算法	用时/ms
MAPPO	21.8	MAPPO-APSE	24.9
COMA	17.3	COMA-APSE	22.3
IQL	17.5	IQL-APSE	21.2

由表 3 可见, 加入 APSE 后的算法单次决策用时有所增加, 这与预期相符, 单次决策耗时增加在 3 ~ 5 ms 左右. 结合第 3.2 节中的算法测试结果来看, APSE 以较小的时间代价得到了平均奖励和资源分配效能的较大提升.

表4 消融实验算法设置

算法	预训练机制	Attention机制
MAPPO-APSE	√	√
MAPPO-Attention	×	√
MAPPO	×	×

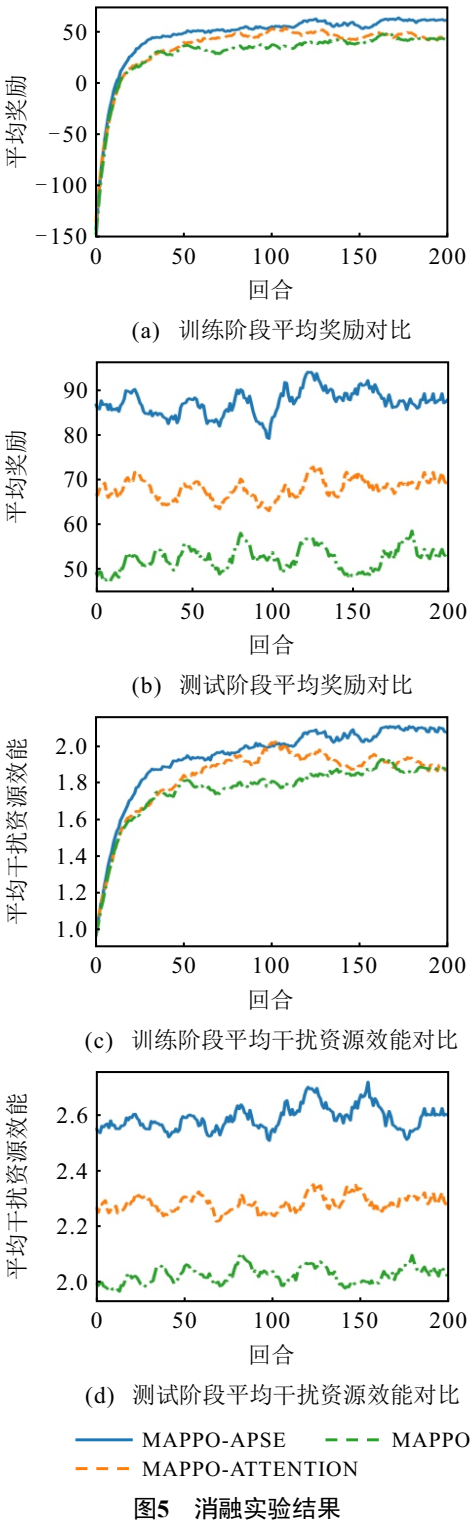


图5 消融实验结果

3.4 消融实验

本文在 APSE 的基础上删除了预训练机制产生的 MAPPO-Attention 算法, 通过比较不同算法的性能来评估各机制的作用. 消融实验中的 3 种算法具体设置如表 4 所示. 其中: √表示包含该机制, ×表示不包含该机制.

图 5 为各消融实验算法在平均奖励和平均资源效能两个指标上的训练结果和测试结果. 从结果上看, MAPPO-Attention 算法相较于 MAPPO 在性能上有所提高, 而加入了预训练后的 MAPPO-APSE 的性能相比于未进行预训练的算法有进一步地提高, 表明注意力机制和预训练机制对于 APSE的性能均有所影响. MAPPO-Attention 采用注意力机制代替了原算法中的全连接网络, 通过注意力加权的方式在压缩环境状态的同时, 能够提取出状态中不同智能体应关注的本质特征, 提升了网络的特征提取能力, 这对于后续算法的计算是十分重要的. 消融实验结果也表明引入注意力机制对于提升 APSE 的特征提取能力是有效的.

加入预训练机制后的 MAPPO-APSE 算法性能在 MAPPO-Attention 的基础上有进一步提升, 这是由于随机初始化的 APSE 会因为特征提取的不正确导致后续算法的性能受到影响. 而 APSE 通过预训练, 解决了随机初始化导致的特征提取的不正确、训练不稳定问题, 进而提升了 MAPPO-APSE 的决策性能.

4 结　论

本文针对拒止条件下无人机集群协同干扰资源分配任务中, 由于环境状态空间维度过大、特征提取困难、环境非平稳导致 MARL 算法决策性能差的问题, 提出了一种基于注意力机制的预训练自编码器(APSE), 并设计了拒止条件下的无人机集群协同干扰资源分配仿真环境. 首先, 通过仿真实验验证了 APSE 对于 MARL 算法决策性能提升的有效性. 其中: MAPPO-APSE 在有效干扰占空比比原算法更长的情况下干扰资源消耗量降低了 20 %, IQL-APSE 将有效干扰占空比从原来的 0 % 提升至 4.47 %. 然后, 结合实验数据分析了 CTDE 范式相比于 DTDE 范式更适用于拒止条件下的干扰资源分配任务的原因. 最后, 通过消融实验进一步验证了 APSE 中引入的预训练机制和自注意力机制可以有效提升其特征提取能力以及训练稳定性. 总之, 所提出 APSE 为基础的框架创新, 适用于大多数 MARL 算法, 可以为现实的 POMDP 和未来大规模多智能体场景理论研究和实际应用提供参考.

参考文献 (References)

[1] Arjoune Y, Faruque S. Smart jamming attacks in 5G new radio: A review[C]. Proceedings of the 10th Annual Computing and Communication Workshop and Conference. Las Vegas, 2020: 1010-1015.

[2] 韩子硕, 范喜全, 郝齐. 国内外无人机系统研究进展及应用[J]. 无线电工程, 2024, 54(5): 1236-1246.
(Han Z S, Fan X Q, Hao Q. Research progress and applications of UAV systems at home and abroad[J]. Radio Engineering, 2024, 54(5): 1236-1246.)

[3] 王健, 杨渡佳, 黄科举, 等. 认知电子战发展趋势: 从单体智能到群体智能[J]. 信息对抗技术, 2023, 2(4): 151-170.
(Wang J, Yang D J, Huang K J, et al. Developing trend of cognitive electronic warfare: From single-agent intelligence to multi-agent intelligence[J]. Information Countermeasure Technology, 2023, 2(4): 151-170.)

[4] Sun W F, Tang M, Zhang L J, et al. A survey of using swarm intelligence algorithms in IoT[J]. Sensors, 2020, 20(5): 1420.

[5] Arulkumaran K, Deisenroth M P, Brundage M, et al. Deep reinforcement learning: A brief survey[J]. IEEE Signal Processing Magazine, 2017, 34(6): 26-38.

[6] 闫超, 相晓嘉, 徐昕, 等. 多智能体深度强化学习及其可扩展性与可迁移性研究综述[J]. 控制与决策, 2022, 37(12): 3083-3102.
(Yan C, Xiang X J, Xu X, et al. A survey on scalability and transferability of multi-agent deep reinforcement learning[J]. Control and Decision, 2022, 37(12): 3083-3102.)

[7] 彭翔, 许华, 蒋磊, 等. 一种融合噪声网络的深度强化学习通信干扰资源分配算法[J]. 电子与信息学报, 2023, 45(3): 1043-1054.
(Peng X, Xu H, Jiang L, et al. A deep reinforcement learning communication jamming resource allocation algorithm fused with noise network[J]. Journal of Electronics & Information Technology, 2023, 45(3): 1043-1054.)

[8] 饶宁, 许华, 齐子森, 等. 基于最大策略熵深度强化学习的通信干扰资源分配方法[J]. 西北工业大学学报, 2021, 39(5): 1077-1086.
(Rao N, Xu H, Qi Z S, et al. Allocation method of communication interference resource based on deep reinforcement learning of maximum policy entropy[J]. Journal of Northwestern Polytechnical University, 2021, 39(5): 1077-1086.)

[9] Liu S Y, Xu Y H, Li G X, et al. Multidimensional resource management for distributed MEC networks in jamming environment: A hierarchical DRL approach[J]. IEEE Internet of Things Journal, 2024, 11(9): 16859-16872.

[10] 许华, 宋佰霖, 蒋磊, 等. 一种通信对抗干扰资源分配智能决策算法[J]. 电子与信息学报, 2021, 43(11): 3086-3095.
(Xu H, Song B L, Jiang L, et al. An intelligent decision-

making algorithm for communication countermeasure jamming resource allocation[J]. Journal of Electronics & Information Technology, 2021, 43(11): 3086-3095.)

[11] 陆永安, 陈杰豪, 张琪露, 等. 基于全并行深度Q网络的通信干扰资源快速分配算法[J]. 现代电子技术, 2024, 47(13): 47-54.
(Lu Y A, Chen J H, Zhang Q L, et al. Communication jamming resource fast allocation algorithm based on fully parallel deep Q-network[J]. Modern Electronics Technique, 2024, 47(13): 47-54.)

[12] 刘旖菲, 李小帅, 杨俊安, 等. 基于 SANER-PPO 算法的无人机集群干扰资源分配方法[J]. 控制与决策, 2024, 39(12): 3937-3945.
(Liu Y F, Li X S, Yang J A, et al. SANER-PPO algorithm-based jamming resource allocation for UAV swarm[J]. Control and Decision, 2024, 39(12): 3937-3945.)

[13] 姚昌华, 万中妨, 张建照, 等. 基于动态联盟的无人机集群协同干扰方法[J]. 电讯技术, 2024, 64(9): 1353-1360.
(Yao C H, Wan Z F, Zhang J Z, et al. Collaborative jamming based on dynamic alliances in UAV cluster[J]. Telecommunication Engineering, 2024, 64(9): 1353-1360.)

[14] Liu X Y, Tan Y. Attentive relational state representation in decentralized multiagent reinforcement learning[J]. IEEE Transactions on Cybernetics, 2022, 52(1): 252-264.

[15] Goldsmith A. Path loss and shadowing[M]. Cambridge: Cambridge University Press, 2005.

[16] 史殿习, 刘聪, 佘馥江, 等. GPS 拒止环境下基于定位置信度的多无人机协同定位方法[J]. 计算机科学, 2022, 49(4): 302-311.
(Shi D X, Liu C, She F J, et al. Cooperation localization method based on location confidence of multi-UAV in GPS-denied environment[J]. Computer Science, 2022, 49(4): 302-311.)

[17] 贾耀. GNSS 拒止未知环境无人机自主导航技术[D]. 西安: 西安电子科技大学, 2021: 61-66.
(Jia Y. GNSS denies unmanned aerial vehicle navigation technology in unknown environments[D]. Xi'an: Xidian University, 2021: 61-66.)

[18] Friis H T. A note on a simple transmission formula[J]. Proceedings of the IRE, 1946, 34(5): 254-256.

[19] François-Lavet V, Henderson P, Islam R, et al. An introduction to deep reinforcement learning[J]. Foundations and Trends® in Machine Learning, 2018, 11(3/4): 219-354.

[20] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J/OL]. 2017, arXiv: 1706.03762.

[21] Yu C, Velu A, Vinitsky E, et al. The surprising effectiveness of PPO in cooperative, multi-agent games[J/OL]. 2021, arXiv: 2103.01955.

[22] Tampuu A, Matiisen T, Kodelja D, et al. Multiagent cooperation and competition with deep reinforcement learning[J]. PLoS One, 2017, 12(4): e0172395.

[23] Foerster J, Farquhar G, Afouras T, et al. Counterfactual multi-agent policy gradients[C]. Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans, 2018: 2974-2980.

作者简介

张澳 (2001−), 男, 博士生, 主要研究方向为群体智能、强化学习, E-mail: zhangao23@nudt.edu.cn;

杨渡佳 (1991−), 男, 讲师, 博士, 主要研究方向为群体智能、人工智能, E-mail: yangdj@nudt.edu.cn;

王健 (1991−), 男, 讲师, 博士, 主要研究方向为认知电子战、群体智能, E-mail: wangjiannudt@nudt.edu.cn;

李小帅 (1989−), 女, 副教授, 博士, 主要研究方向为无人机集群、群体智能、资源分配, E-mail: xiaoshuai.li@nudt.edu.cn;

杨俊安 (1965−), 男, 教授, 博士, 博士生导师, 主要研究方向为认知电子战、信号处理、 智能计算, E-mail: yangjunan@ustc.edu;

刘辉 (1983−), 男, 副教授, 博士, 主要研究方向为智能信息处理、认知电子战, E-mail: christ592604@163.com.