自动化学报
ACTA AUTOMATICA SINICA

图注意力驱动的跨场景协同拦截强化学习方法

吴子鹏 马麒超 秦家虎 詹晨光 张金鹏

Graph Attention-driven Reinforcement Learning for Cross-scenario Cooperative Interception

WU Zi-Peng, MA Qi-Chao, QIN Jia-Hu, ZHAN Chen-Guang, ZHANG Jin-Peng

在线阅读 View online: https://doi.org/10.16383/j.aas.c250382

图注意力驱动的跨场景协同拦截强化学习方法

吴子鹏 [1] 马麒超 [1] 秦家虎 [1] 詹晨光 [2] 张金鹏 [2]

摘 要 针对复杂动态场景下大规模无人集群拦截任务, 提出一种基于图注意力机制与动态分组的集群协同拦截框架. 现有基于规则或优化的方法在实时性、泛化性与目标分配效能方面存在局限, 而多智能体强化学习 (MARL) 在复杂动态场景下面临维度爆炸、策略泛化性不足等挑战. 为提升复杂动态场景下集群拦截策略学习效率以及跨场景泛化性, 创新性地设计了动态分组模块、图注意力模块与改进多智能体强化学习模块, 并融合成一套闭环算法框架: 1) 动态分组模块通过周期性聚类将敌方集群分解为低维战术小组, 敌方小组信息作为节点传输给图注意力模块, 降低状态–动作维度; 2) 图注意力模块利用敌方小组节点信息, 基于图注意力网络进行特征融合并构建己方智能体–敌方小组间相对关系, 生成目标重要性权重以引导差异化奖励函数设计, 提升了策略目标分配与泛化能力; 3) MARL 模块结合差异化奖励函数与融合特征, 基于 SAC 算法与对抗性训练机制, 以进一步增强策略泛化性. 仿真实验表明该框架显著提升了复杂动态场景下集群拦截效率以及跨场景泛化性.

关键词 协同拦截; 多智能体强化学习; 图注意力机制; 动态分组

引用格式 吴子鹏, 马麒超, 秦家虎, 詹晨光, 张金鹏. 图注意力驱动的跨场景协同拦截强化学习方法. 自动化学报, 2026, **52**(8): 1677−1689

DOI 10.16383/j.aas.c250382 **CSTR** 32138.14.j.aas.c250382

Graph Attention-driven Reinforcement Learning for Cross-scenario Cooperative Interception

WU Zi-Peng[1] MA Qi-Chao[1] QIN Jia-Hu[1] ZHAN Chen-Guang[2] ZHANG Jin-Peng[2]

Abstract For large-scale unmanned swarm interception in complex dynamic scenarios, this paper proposes a coordinated interception framework based on graph attention mechanisms and dynamic grouping. Existing rule-based or optimization-based methods suffer from poor real-time performance, weak generalization, and low target allocation efficiency, while MARL (multi-agent reinforcement learning) faces dimensionality explosion and limited policy generalization in such scenarios. To improve learning efficiency of swarm interception strategies and cross-scenario generalization in complex dynamic scenarios, this work innovatively designs three modules——A dynamic grouping module, a graph attention module, and an improved MARL module——And integrates them into a closed-loop algorithm framework. The dynamic grouping module periodically clusters enemy units into low-dimensional tactical groups and passes the enemy group information as nodes to the graph attention module, reducing state-action dimensionality. The graph attention module uses enemy group-node data, applies graph attention networks for feature fusion, builds relative relations between friendly agents and enemy groups, and outputs target-importance weights to guide differentiated reward function design, improving target allocation and generalization. The MARL module combines differentiated reward function with fused features, adopting the soft actor-critic (SAC) algorithm and adversarial training to further enhance policy generalization. Simulation results show that the framework significantly improves swarm interception efficiency and cross-scenario generalization in complex dynamic scenarios.

Keywords coordinated interception; multi-agent reinforcement learning; graph attention mechanism; dynamic grouping

Citation Wu Zi-Peng, Ma Qi-Chao, Qin Jia-Hu, Zhan Chen-Guang, Zhang Jin-Peng. Graph attention-driven reinforcement learning for cross-scenario cooperative interception. *Acta Automatica Sinica*, 2026, **52**(8): 1677−1689

近年来, 无人集群作战技术发展迅速, 已广泛应用于军事突袭、战场侦察和领空防御等多样化场景[1−2]. 该技术在成本效益与任务适应性等方面具备显著优势, 不仅能大幅提升作战效率, 还能显著降

收稿日期 2025-08-11 录用日期 2025-11-14
Manuscript received August 11, 2025; accepted November 14, 2025
国家重点研发计划 (2022ZD0120002), 空基信息感知与融合全国重点实验室开放课题 (202413) 资助
Supported by the National Key Research and Development Program of China (2022ZD0120002) and Open Project of National Key Laboratory of Air-based Information Perception and Fusion (202413)
本文责任编委 杨涛
Recommended by Associate Editor YANG Tao

1. 中国科学技术大学自动化系 合肥 230027 2. 空基信息感知与融合全国重点实验室 洛阳 471009
1. Department of Automation, University of Science and Technology of China, Hefei 230027 2. National Key Laboratory of Air-based Information Perception and Fusion, Luoyang 471009

低作战成本和人员损失. 目前, 无人集群协同攻击等体系化作战形态已成为现代防空体系的核心威胁之一, 受到广泛关注[3−4]. 国内外已先后启动众多研究项目, 例如美国 DARPA 的“灰山鹑”、“低成本无人机技术蜂群”和“小精灵”项目[5] 以及中国电子科技集团的大规模集群飞行试验. 由于无人集群具有规模大、机动性强、动态性高等特点, 能快速突破传统静态拦截体系, 因此发展具备跨场景适应能力的智能动态拦截方法是当前研究重点之一[6]. 本文围绕兼具复杂性 (集群对抗规模庞大、动力学复杂) 与动态性 (对抗规模与策略可变) 的集群对抗场景, 旨在提出一种可跨场景泛化的多智能体强化学习框架, 实现有效的协同目标拦截. 该框架不仅能提升大规模场景、强机动目标情形下的拦截策略学习效率, 还能有效增强拦截策略在动态变化场景中的泛化能力.

传统协同拦截方法主要包括基于规则 (rule-based) 和基于优化 (optimization-based) 的两类方法. 基于规则的方法依赖预设的专家规则进行决策[7]. 这类方法能有效融合专家经验, 但过度依赖规则设定导致策略僵化, 难以应对环境动态变化. 基于优化的方法将问题建立为数学模型并进行求解[8], 给出的拦截策略通常具有明确的理论性能保证. 然而当对抗规模增大或动态性增强 (如敌方策略转变) 时, 由于计算复杂度急剧增长, 策略求解将变得难以处理. 此外, 基于优化的方法在动态跨场景下也容易失效.

近年来, 基于多智能体强化学习的方法在完成协同任务方面展现出巨大潜力[9−12], 该方法通过智能体与环境不断进行交互, 自主探索学习最优协同策略, 有望克服基于规则方法的僵化性并突破基于优化方法的计算瓶颈. 文献 [13] 在具有动力学模型的无人机集群空战场景中, 通过特定奖励设计, 采用 MAPPO[14] 训练策略. 文献 [15] 在集群空战对抗环境中利用 MADDPG[16], 使无人机集群学习到更优策略与任务表现. 文献 [17] 则引入分层强化学习思想[18], 基于预定义无人机动作库, 提出分层 MADDPG 框架用于集群对抗. 文献 [19] 进一步提出 EA-MADDPG 算法, 通过特殊奖励设计与动作增强解决多导弹拦截单目标问题. 然而上述方法的训练场景相对单一, 任务复杂度不高, 且均未考虑跨场景泛化这一需求. 针对上述不足, 文献 [20] 提出 DRGRL 算法, 利用图神经网络[21] 提取战斗关系并实现多无人机协作. 文献 [22] 则基于 MAPPO 提出两阶段注意力机制框架, 捕获关键交互关系以提升集群对抗性能. 以上两种方法 [20, 22] 通过提取交互关系 (如图结构或注意力权重) 辅助策略学习, 在一定程度上提升了集群对抗策略的跨场景泛化性, 但所考虑的对抗规模较小且交互关系难以解释.

综上所述, 在兼具环境复杂性与动态性的大规模无人集群对抗场景中, 现有多智能体强化学习方法仍面临如下核心挑战:

1) 维度爆炸与训练收敛难题. 无人集群状态−动作空间随规模指数级增长, 在大规模情形下训练收敛难、学习效率低[23].

2) 动态场景泛化能力缺陷. 场景动态性要求策略具备强跨场景泛化性[24], 而现有方法难以有效应对敌方策略及对抗规模的灵活演变.

3) 大规模动态目标分配失准. 拦截目标威胁等级不同、空间分布多样[25], 现有方法难以精准感知并表征此类相对关系, 导致目标分配失准.

针对上述挑战, 本文提出基于动态分组与图注意力机制的集群协同拦截框架, 以完成复杂动态场景下协同拦截任务并满足跨场景泛化需求. 该框架包含三个核心模块: 首先, 动态分组模块通过周期性聚类, 将大规模敌方目标按关联性划分为小组, 从而降低状态−动作空间维度以提升学习效率; 其次, 图注意力模块以敌方小组为节点构建图神经网络[26], 融合敌我特征并抽象出己方智能体−敌方小组间相对关系, 从而驱动差异化奖励函数设计, 增强目标分配合理性与跨场景泛化性; 最后, 多智能体强化学习模块基于 SAC (soft actor-critic)[27] 算法, 结合上述融合特征与差异化奖励函数, 引入阶段性存档 (checkpoints)[28] 与对抗性训练机制[29], 以提升拦截策略多样性. 本文主要贡献如下:

1) 动态分组模块通过聚类大规模敌方集群, 降低状态−动作空间维度以提升训练效率. 该模块周期性地触发 K-means 算法[30], 根据敌方集群的空间分布将其动态分解为多个低维战术小组, 并提取多重小组信息 (如中心位置、速度向量和小组规模等), 从而大大压缩状态−动作空间, 显著缓解了维度爆炸带来的挑战. 同时, 该模块还将分组特征输出至图注意力模块以进一步捕获敌我相对关系.

2) 图注意力模块利用敌我状态特征, 通过交叉注意力机制计算融合特征与相对关系, 自适应生成敌方小组重要性权重. 这些权重能够灵活缩放不同敌方目标的相关奖励, 实现目标奖励差异化, 实时引导拦截高价值目标. 此外, 图注意力模块通过特征相关性提取和差异化奖励机制, 实现了拦截策略跨场景迁移, 能有效应对不同敌方策略和对抗规模. 并且进一步采用多头注意力机制[31] 过滤战场噪声, 增强了策略的跨场景稳定性.

总体而言, 动态分组缓解了状态–动作维度爆炸带来的挑战, 图注意力机制确保敌我关系的关键特征被捕获, 而差异化奖励则将敌我关系反馈至策略更新中, “动态分组–关系构建–差异化奖励”闭环设计实现了大规模无人集群的高效策略学习.

本文后续内容如下: 第 1 节介绍集群拦截环境, 包含动力学模型以及任务规则; 第 2 节进行集群拦截任务的问题描述与数学建模; 第 3 节介绍动态分组图注意力集群拦截算法, 详细描述动态分组模块、图注意力模块以及多智能体强化学习 (multi-agent reinforcement learning, MARL) 模块的工作原理; 第 4 节设计集群拦截任务的仿真实验, 验证了算法的有效性; 第 5 节总结本文内容并展望未来工作.

1　集群拦截环境

在集群拦截环境中, 本文任务设定为敌方目标采用可变策略入侵地图的目标区域, 要求控制己方导弹高效拦截所有敌方目标. 本节首先介绍环境中敌方目标与己方导弹的动力学模型和控制输入, 随后介绍集群拦截任务关键规则的设定.

1.1　敌方目标模型

敌方目标采用无人机模型, 其控制输入为加速度 a_u 和滚转角速度 ω_u [17], 二者通过影响无人机的线速度和偏航角, 从而控制无人机的运动. 其速度与滚转角更新方式如下

$$\begin{cases} v_{t+1} = \mathrm{Clip}(v_t + a_u \cdot a_{\max} \cdot \Delta t,\ v_{\min},\ v_{\max}) \\ \phi_{t+1} = \mathrm{Clip}(\phi_t + \omega_u \cdot \omega_{\max} \cdot \Delta t,\ \phi_{\min},\ \phi_{\max}) \end{cases} \tag{1}$$

其中, v_t 和 v_{t+1} 分别代表 t 时刻和 $t+1$ 时刻的线速度; $a_{\max}$ 代表最大加速度, 并且限制速度范围在 $[v_{\min},\ v_{\max}]$ 之间; ϕ_t, ϕ_{t+1} 代表 t 时刻和 $t+1$ 时刻的滚转角; $\omega_{\max}$ 为最大滚转角速度, 并且限制滚转角度范围在 $[\phi_{\min},\ \phi_{\max}]$ 之间; Δt 为控制时间步长.

得到速度 v_t 和滚转角 ϕ_t 之后, 根据设定好的转弯模型进行偏航角更新

$$\psi_{t+1} = \arctan(\sin(\psi_t + \Delta\psi),\ \cos(\psi_t + \Delta\psi)) \tag{2}$$

其中, 偏航角增量 $\Delta\psi = -(c_{turn}/v_t) \cdot \tan(\phi_t) \cdot \Delta t$, ψ 为偏航角, $c_{turn} = 0.004$ 为转弯系数, 与无人机气动特性相关. 从而得到无人机的速度向量 $\boldsymbol{v}_t = v_t \cdot [\cos\psi_t,\ \sin\psi_t]^{\mathrm{T}}$. 根据速度向量在每一个时间步进行位置更新, $\boldsymbol{p}_{t+1} = \mathrm{Clip}(\boldsymbol{p}_t + \boldsymbol{v}_t \cdot \Delta t,\ -B,\ B)$. 其中 $\boldsymbol{p}$ 代表无人机位置向量, 包含了 x, y 坐标; B 为拦截任务边界, 设定 $x,\ y \in [-B,\ B]$ 以避免越界. 该无人机模型以真实动力学模型为基础, 忽略了空气阻力等因素, 同时考虑了关键的速度、偏航角与滚转角之间的复杂耦合关系.

1.2　己方导弹模型

与敌方目标模型类似, 己方导弹的控制输入为轴向加速度 a_A 与侧向加速度 a_L, 状态更新遵循以下规则

$$\begin{cases} v_{t+1}^M = \mathrm{Clip}(v_t^M + a_A \cdot \Delta t,\ v_{\min}^M,\ v_{\max}^M) \\ \psi_{t+1}^M = \arctan\left(\sin\left(\psi_t^M + \dfrac{a_L}{v_t^M} \cdot \Delta t\right)\cos\left(\psi_t^M + \dfrac{a_L}{v_t^M} \cdot \Delta t\right)\right) \\ \boldsymbol{p}_{t+1}^M = \mathrm{Clip}(\boldsymbol{p}_t^M + \boldsymbol{v}_t^M \cdot \Delta t,\ -B,\ B) \end{cases} \tag{3}$$

其中 v_t^M 为导弹线速度; $\boldsymbol{v}_t^M = v_t^M \cdot [\cos\psi_t^M,\ \sin\psi_t^M]^{\mathrm{T}}$ 为导弹速度向量; ψ^M 为导弹偏航角; $\boldsymbol{p}_t^M$ 为导弹位置; B 为环境边界, 与无人机模型中设置相同.

1.3　拦截任务规则

在初始状态中, 敌方目标集群与己方导弹集群分别位于交战区域的两侧且各自呈直线排列, 初始朝向为两组集群均朝向对方. 由于仿真环境设定为二维平面, 己方导弹和敌方目标的高度被限制在同一平面上. 同时, 如第 1.1 节和第 1.2 节所述, 目标区域存在空间限制, 任意敌方目标的坐标 $(x_j,\ y_j)$ 均需要满足 $-4\,000\ \mathrm{m} \le x_j,\ y_j \le 4\,000\ \mathrm{m}$, 即 $B = 4\,000\ \mathrm{m}$, 约束训练区域有助于提升算法训练效率. 在任务执行过程中, 如图 1 所示, 导弹攻击生效需同时满足距离与角度条件[17], 仅当导弹 i 与目标 j 之间的距离小于有效伤害半径, 且攻击方向夹角小于一定的角度阈值时, 导弹 i 才有一定概率摧毁敌方目标. 攻击的成功取决于空间临近性与攻击姿态对准性的联合满足. 敌方目标的任务为入侵地图目标区域, 己方导弹的任务为拦截所有敌方目标. 因此当所有敌方目标都被成功击毁时, 己方任务成功

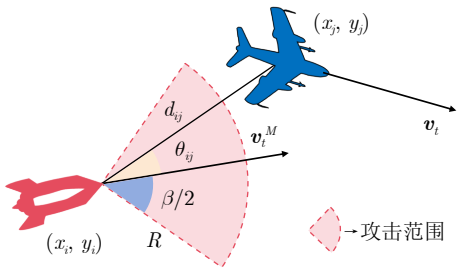


图 1　导弹拦截无人机示意图

Fig. 1　Schematic diagram of missile intercepting unmanned aerial vehicle

完成, 该回合结束; 或者存在敌方目标逃脱己方导弹攻击, 并成功入侵目标区域, 即敌方完成任务, 该回合结束; 如果双方均没有完成任务, 则在到达预先设定的最大时间步时回合结束.

2 问题描述与数学建模

本文考虑的场景是己方导弹集群协作拦截敌方目标集群. 敌方目标集群需躲避导弹拦截, 并入侵地图的目标区域, 己方导弹则需在目标入侵过程中相互协作拦截所有敌方目标. 其中己方导弹集群由本文提出的框架训练所得策略控制, 敌方目标由基于 SAC 的多智能体强化学习基线算法训练所得策略集合控制. 本文将控制己方导弹拦截敌方目标的过程建立为马尔科夫博弈 (Markov game) 模型.

马尔科夫博弈模型可以表示为一个元组 $(N, S, A, O, P, r, \gamma, \omega_e)$, 其中 $N = \{1, \cdots, m\}$ 为己方导弹集合; S 代表状态空间; $A = A_1 \times A_2 \times \cdots \times A_m$ 代表联合动作空间, 且 A_i 代表己方第 i 个导弹的动作空间; $o_i = O(s; i)$ 是己方导弹 i 在全局状态 s 时的局部观测; $P: S \times A \times S' \mapsto [0, 1]$ 代表给定联合动作 $\boldsymbol{a} = \{a_1, a_2, \cdots, a_m\}$ 时, 从状态 S 到状态 S' 的状态转移概率; $r: S \times A \mapsto \mathbf{R}$ 代表共享奖励函数; $\gamma \in [0, 1)$ 代表折扣因子; ω_e 编码了敌方目标小组的重要性信息. 每个己方导弹具有一个 θ_i 参数化的策略网络 $\pi_{\theta_i}(a_i|o_i, \omega_e)$, 根据当前的局部观测 o_i 和敌方目标重要性信息 ω_e 生成动作 a_i. 随后己方导弹会得到一个奖励 $r_i(s, \boldsymbol{a})$ 并且优化自身的累计奖励 $R_i = \sum_{t=0}^{T} \gamma^t r_i(s_t, \boldsymbol{a}_t)$. 其中 t 是当前步数, T 是轨迹最大步数, s_t 为环境状态向量, $\boldsymbol{a}_t$ 代表联合动作向量.

3 动态分组图注意力集群拦截算法

基于集群对抗拦截任务的多维度需求, 本文提出一种集成动态分组、图注意力网络与 MARL 的协同决策框架, 如图 2 所示. 该框架通过三个模块相互作用, 提升算法学习效率与拦截策略跨场景泛化性: 1) 动态分组模块进行周期性动态分组降低任务复杂度, 基于敌方位置聚类生成敌方小组简化战场结构. 2) 图注意力模块将敌方分组结果作为节点输入, 以实现敌我关系构建及策略自适应调整; 融合节点特征并解析战场动态关系以生成敌方小组重要性权重; 驱动差异化奖励函数设计, 通过强化攻击重要性高的敌方小组, 将战术意图嵌入学习过程. 3) MARL 模块基于 SAC 算法, 利用融合特征、差异化奖励与对抗性训练提供的多样化敌方策略模型来优化集群策略.

3.1 动态分组模块

在动态分组模块中, 敌方目标的空间聚类过程基于实时的 K-means 算法实现, 并以固定时间间隔

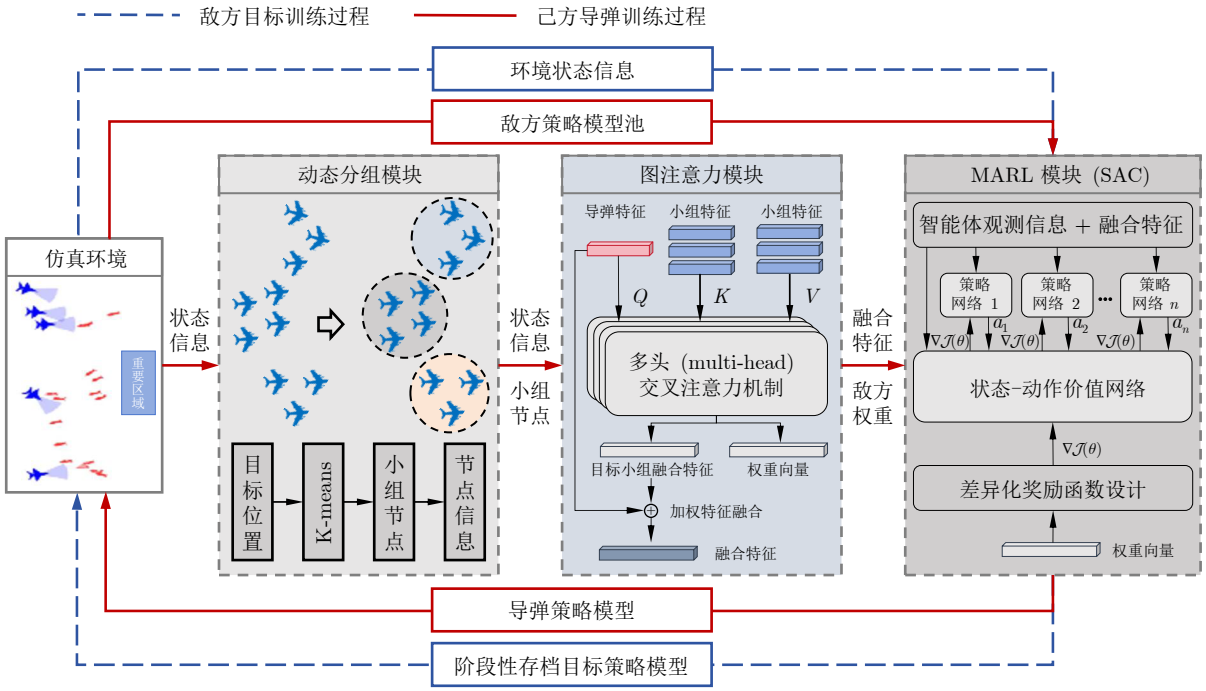


图 2 动态分组图注意力集群拦截算法框架

Fig. 2 Dynamic grouping graph attention swarm interception algorithm framework

触发. 其核心在于将高维战场中离散分布的敌方单位依据动态位置信息聚合为低维战术单元, 即各个敌方小组. 各阶段具体执行流程如下:

步骤 1. 根据战场规模、敌方目标分布与分组拦截复杂度, 预设聚类数 K, 然后随机选取 K 个敌方目标的坐标作为初始聚类中心 $\{c_k^{(0)}\}_{k=1}^{K}$. 后续聚类中心的选择需满足空间分布多样性约束, 以确保覆盖整个战场区域:

$$\min_{\{c_k\}} \sum_{k=1}^{K} \left\| c_k^{(0)} - c_{k'}^{(0)} \right\|_2^2, \quad \forall k' \neq k \tag{4}$$

随后的迭代优化阶段包含两个步骤 (步骤 2、步骤 3) 循环.

步骤 2. 进行归属分配, 即计算每个敌方位置 x_j 到各个聚类中心 $\{c_k^{(t)}\}_{k=1}^{K}$ 的欧氏距离

$$dist\left(x_j,\ c_k^{(t)}\right) = \left\| x_j - c_k^{(t)} \right\|_2 \tag{5}$$

将 x_j 分配至距离最近的聚类

$$\boldsymbol{C}_k^{(t)} = \left\{ x_j \mid k = \arg\min_{k'} dist\left(x_j,\ c_{k'}^{(t)}\right) \right\} \tag{6}$$

步骤 3. 进行聚类中心更新, 即按照分配结果重新计算聚类中心

$$c_k^{(t+1)} = \frac{1}{\left|\boldsymbol{C}_k^{(t)}\right|} \sum_{x \in \boldsymbol{C}_k^{(t)}} x \tag{7}$$

步骤 2 和步骤 3 循环至满足聚类收敛条件: 达到最大迭代次数 $T_{\max}$ 或中心偏移量收敛到极小值 ε

$$\left\| c_k^{(t+1)} - c_k^{(t)} \right\|_2 < \varepsilon \tag{8}$$

步骤 4. 动态分组模块根据战场动态特性, 每隔固定时间步长对敌方目标进行重新聚类, 然后将聚类结果映射为敌方战术小组 $\{u_i\}_{i=1}^{n}$ 并提取多维战术特征向量 u_i:

$$u_i = \left[n_i,\ c_i,\ D_i,\ \frac{1}{n_i} \sum_{i=1}^{n_i} v_i \right] \tag{9}$$

每个战术小组具有成员数 n_i、中心坐标 c_i、边界半径 $D_i = \max_{x \in \boldsymbol{C}_i} \|x - c_i\|_2$ 以及小组平均速度等属性. 依据对抗规模与敌方分布, 人为事先设定聚类数 K, 本文取 $K = 3$. 该模块通过空间压缩缓解了大规模集群对抗训练过程中的维度爆炸问题. 聚类结果作为敌方小组节点输入图注意力模块, 使其专注于解析小组级战术关系而非个体细节, 降低决策难度, 提升后续决策效率.

3.2　图注意力模块

为有效捕捉复杂对抗场景中己方导弹与敌方目标小组之间动态变化的战术交互关系, 本文引入图注意力网络 (graph attention network, GAT) 并构建图注意力模块, 其核心架构如图 3 所示. 该模块构建了一个异构战术交互图作为其底层表示, 该图结构定义了一个节点集合 V_{unit}, 且 $V_{unit} = \{u_i\}_{i=1}^{n} \cup \{b_i\}_{i=1}^{m}$, 其中 u_i 为敌方小组节点, 如式 (9) 所示, 编码了每个小组的个体数量、中心位置、边界半径以及平均速度等关键战术属性; b_i 为己方导弹节点, 包含当前位置、速度、偏航等状态信息; n 和 m 分别为敌方目标小组和己方导弹的实时数量.

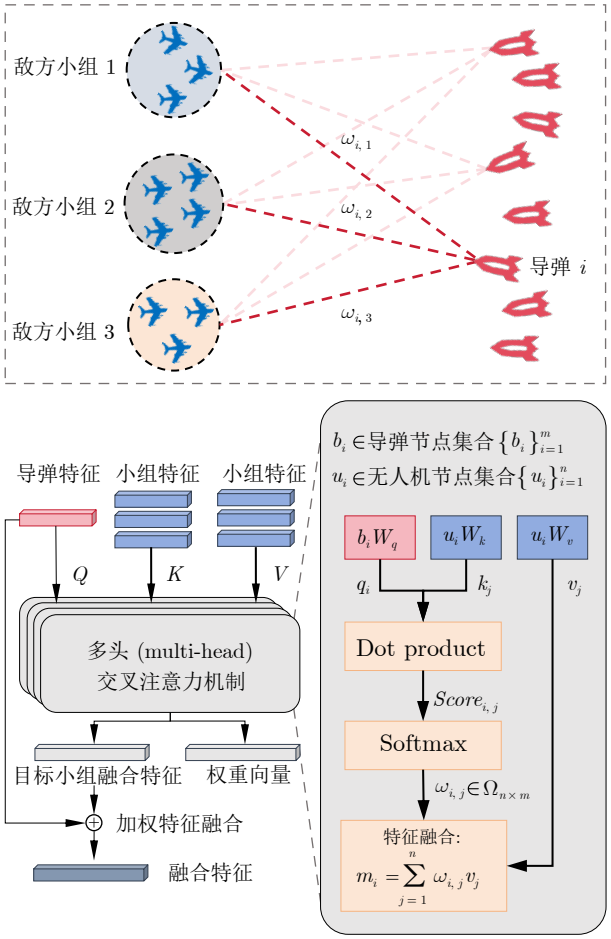


图 3　图注意力机制流程图

Fig. 3　Graph attention mechanism flow chart

该模块具有两个核心功能: 1) 动态地捕捉和分析上述节点之间的相对关系, 并生成敌我融合特征. 节点间的关系主要通过节点间的边 (edges) 来体现, 本文将边设计为连接导弹节点 b_i 与各敌方小组节点 u_i. 值得注意的是, 这些连接关系并非静态预设, 而是依据战场态势 (如相对位置、威胁程度等) 动态生成, 并给不同敌方小组分配差异化关注程度. 2) 将差异化关注程度与 MARL 算法的训练过程相结合, 从而将其形式化为不同的权重系数. 然后基

于该权重系数设计差异化奖励函数, 从而生成与各敌方小组重要程度对应的奖励值, 提升策略在大规模敌方目标场景下的泛化性.

3.2.1 交叉注意力机制

动态相对关系捕获的核心在于注意力机制 (attention mechanism). 对于图中每一个节点 (如一个导弹节点 b_i), 注意力机制能自适应学习并计算其所有相邻节点 (如它所关注的若干敌方小组节点 u_i) 对其当前决策的重要性权重, 该过程无需依赖任何预先定义的关联规则. 注意力权重的计算基于节点特征, 这意味着系统能够捕捉导弹–敌方小组之间复杂且非线性的相互关系, 从而实时识别对当前导弹节点 b_i 决策最关键的敌方小组节点 u_i. 随着战场态势变化, 图结构和节点特征相应发生改变, 注意力机制会自动重新评估节点间关系的重要性, 促使策略进行自适应调整. 利用图结构使得算法更加专注于建模导弹与目标组之间的战术关系, 以此增强导弹的拦截决策能力.

本模块采用目标导向的交叉注意力机制 (cross attention), 使用导弹节点 $b_i \in \mathbf{R}^{1\times d_b}$ 构建查询向量 (query) q_i, 敌方小组节点 $u_i \in \mathbf{R}^{1\times d_u}$ 构建关键向量 (key) k_j 和价值向量 (value) v_j. 其中, d_b 为导弹节点特征向量长度, d_u 为敌方小组节点向量长度. 在具体实现上, 查询向量、关键向量以及价值向量分别通过 $q_i = f_b(b_i)W_q$, $k_j = f_u(u_i)W_k$, $v_j = f_u(u_i)W_v$ 编码获得. 其中 f 是一个多层感知机 (multilayer perceptron), 用来将导弹与敌方小组特征 b_i, u_i 编码成维度为 d 的特征向量; W_q, W_k, $W_v \in \mathbf{R}^{d\times d_k}$ 是维度为 $d\times d_k$ 的可学习权重矩阵; $q_i \in Q = \{q_1, q_2, \cdots, q_m\}$, $k_j \in K = \{k_1, k_2, \cdots, k_n\}$, $v_j \in V = \{v_1, v_2, \cdots, v_n\}$

对于每一个导弹节点特征对应的查询向量 q_i, 需要计算其与其他敌方小组节点特征之间的相对关系, 即计算对应的注意力权重. 计算过程如下:

$$Score_{i,\,j} = q_i^{\mathrm{T}} k_j \tag{10}$$

其中, $Score_{i,\,j}$ 代表己方导弹 i 与第 j 个敌方小组之间的战术关系, 随后使用 Softmax 函数对注意力 $Score_{i,\,j}$ 进行归一化处理, 得到注意力权重

$$\omega_{i,\,j} = \mathrm{Softmax}_j(Score_{i,\,j}) = \frac{\exp\left(Score_{i,\,j}/\sqrt{d_k}\right)}{\sum\limits_{j=1}^{n}\exp\left(Score_{i,\,j}/\sqrt{d_k}\right)} \tag{11}$$

其中 $\omega_{i,\,j}$ 是矩阵 $\Omega_{n\times m}$ 中的元素, 代表归一化后的注意力权重, 即敌方小组 j 对于导弹 i 的威胁程度与重要性, 导弹将优先拦截权重大的敌方小组; $\Omega_{n\times m}$ 是权重矩阵, 用于编码己方导弹与敌方目标小组 $(b_i,\, u_i)$ 的相关性; $\sqrt{d_k}$ 是缩放因子, 用来防止 $Score_{i,\,j}$ 结果过大.

在得到注意力权重后, 进一步计算对应己方导弹的敌方融合特征向量, 最终通过加权融合输出导弹的态势感知特征:

$$m_i = \sum_{j=1}^{n}\omega_{i,\,j}v_j \tag{12}$$

$$\boldsymbol{z}_u = \mathrm{ReLU}\left(m_i \oplus f_b(b_i)\right) \tag{13}$$

其中 m_i 表示经过注意力机制处理后的敌方小组融合特征; $\boldsymbol{z}_u$ 为加权融合特征; $\oplus$ 运算符表示加权融合, 其作用是保留原始导弹状态信息, 确保在计算态势感知特征时不丢失导弹自身的状态数据.

3.2.2 差异化奖励函数设计

差异化奖励函数的核心在于给不同的敌方小组分配不同的奖励值. 本文将注意力权重 $\omega_{i,\,j}$ 引入奖励函数的设计中, 促使己方导弹有选择地拦截重要性较高的敌方目标, 实现敌我相对关系与集群拦截行为对齐. 对于己方导弹 i 和敌方小组 j 的成员:

$$R_{i,\,j} = \omega_{i,\,j}\left(R_{dist,\,ij} + R_{ang,\,ij} + R_{attack,\,ij}\right) \tag{14}$$

在多智能体强化学习模块的奖励函数设计中, $R_{i,\,j}$ 是敌方小组 j 关于己方导弹 i 奖励的一部分. 该部分通过融合三类子奖励函数来引导己方导弹的战术行为. $R_{dist,\,ij}$ 是距离奖励, 用来激励己方导弹 i 缩短与敌方小组 j 成员的物理距离, 促使导弹快速接近目标; $R_{ang,\,ij}$ 是角度奖励, 优化导弹 i 相对于敌方小组 j 成员的攻击方位, 引导己方导弹占据有利战术位置; $R_{attack,\,ij}$ 指攻击奖励, 在成功对敌方小组 j 成员实施有效攻击行为时给予显著的正向激励, 鼓励己方导弹完成最终的敌方目标拦截任务.

上述设计使得奖励函数 $R_{i,\,j}$ 动态依赖于不同敌方小组对于己方导弹的重要程度, 驱使己方导弹自适应地选择优先攻击的敌方小组, 实现资源分配与战术决策的协同优化. 图注意力模块并非独立运行, 而是与 MARL 模块在整个框架中进行端到端的联合训练. 两个模块共享相同的全局损失函数, 参见式 (15), 并基于此函数计算得到的梯度信息同步更新各自的网络参数.

3.3 多智能体强化学习模块

多智能体强化学习模块包含两个关键训练过程: 1) 结合图注意力模块生成的融合特征与差异化奖励函数进行己方导弹策略训练; 2) 采用 SAC 基线算法进行敌方策略训练.

3.3.1　己方导弹策略训练

己方导弹策略训练阶段采用 SAC 算法作为多智能体协同决策的基础架构, 通过构建集中式训练、分布式执行 (centralized training decentralized execution, CTDE) 范式实现高效学习. 所有己方导弹共享由 ϕ 参数化的集中式价值网络, 对由 θ 参数化的策略网络采用参数共享, 即 $\theta_1 = \theta_2 = \cdots = \theta_m$. 该参数共享机制促使各智能体在训练过程中同步高效地更新网络参数, 形成策略一致性. 训练过程的全局目标函数为

$$\mathcal{J}(\theta) = \max_{\theta}\ \mathrm{E}_{s_t\sim\mathcal{D}}\left[\sum_{i=1}^{m}\left(\mathrm{E}_{a_t^i\sim\pi_\theta}\left[Q_\phi\left(s_t,\ \boldsymbol{a}_t\right)\right] + \alpha\mathcal{H}\Big(\pi_\theta\left(\cdot\mid s_t\right)\Big)\right)\right] \tag{15}$$

其中 s_t 为环境状态向量, 包含敌方分组信息与融合特征; $\boldsymbol{a}_t = \{a_t^1,\ \cdots,\ a_t^m\}$ 代表联合动作向量; $\mathcal{D}$ 为经验回放池; $\mathcal{H}\left(\pi_\theta\left(\cdot\mid s_t\right)\right) = -\mathrm{E}_{a_t\sim\pi_\theta}[\log_2\pi_\theta(a_t\mid s_t)]$ 为熵项, 用来维持动作空间探索强度[32], 避免参数共享导致的策略僵化; α 为可调节的熵项系数; Q_ϕ 为 SAC 算法中用于评价状态–动作对好坏的 Q 函数.

与此同时, 结合由图注意力模块得到的融合特征与差异化奖励函数 (第 3.2.1 节和第 3.2.2 节详细描述), 利用己方导弹与敌方小组的相对关系训练策略, 提升其跨场景泛化性. 进一步利用差异化奖励函数鼓励己方导弹优先拦截重要性高的敌方目标, 实现敌我相对关系与集群拦截行为的对齐.

3.3.2　敌方策略训练

敌方策略训练阶段采用对抗性训练机制, 结合 SAC 算法对抗己方导弹策略以训练敌方策略, 使其针对不同水平的己方导弹都有较好的对抗表现. 在此过程中引入阶段性存档机制, 该机制周期性地保存不同训练阶段的敌方策略参数, 从而获得不同的敌方策略模型, 将其整合构建敌方策略模型池. 该策略池汇集的敌方策略模型以多样化的形式影响训练环境, 给己方导弹执行任务的稳定性和泛化性均带来新挑战. 通过迫使己方导弹与多版本敌方策略轮番对抗, 并在差异化奖励函数作用下, 显著提升了己方导弹策略的跨场景泛化性和战术适应性.

4　实验与分析

实验分析旨在回答以下几个核心问题: 所提方法在拦截效率上是否优于基线方法? 所提方法在不同规模和不同敌方策略下的跨场景泛化性如何? 动态分组、图注意力和差异化奖励函数等模块是否提升了整体性能? 本节首先描述实验设置, 包含具体环境设计、算法训练与神经网络超参数设置等; 随后评估算法在所设计环境中的性能表现, 重点测试任务成功率 (SR)、平均击杀数 (AK) 以及平均步数 (AS) 三个关键指标; 与此同时验证算法在跨场景任务中的泛化性表现, 将训练好的模型应用在不同的场景设置中以测试其性能指标的变化; 最后进行整体框架的消融实验, 验证动态分组、图注意力、差异化奖励函数等模块的效果.

4.1　实验设置

基于第 1 节的集群拦截环境介绍, 本文设计了一个己方导弹集群 (20 个) 拦截敌方无人机目标集群 (10 架) 的训练仿真环境. 仿真环境与算法均被部署在配备了 Intel Core i5-13600KF CPU 和 NVIDIA RTX 4070Ti Super GPU 的计算机上. 神经网络参数和算法训练超参数分别汇总于表 1 和表 2, 关于环境设置的细节如下.

表 1　神经网络配置

Table 1　Neural network configuration

名称	设定值
目标网络隐藏层层数	2
目标网络隐藏层宽度	256 × 256
Critic 网络隐藏层层数	2
Critic 网络隐藏层宽度	256 × 256
Actor 网络隐藏层层数	2
Actor 网络隐藏层宽度	256 × 256
激活函数	ReLU
优化器	Adam

表 2　算法训练超参数

Table 2　Hyperparameters for algorithm training

参数	设定值
批量大小 (batch size)	4 096
经验池大小 (buffer size)	1 536 000
学习率	0.000 1
折扣率	0.99
熵项系数	0.2
目标网络更新参数	0.02

4.1.1　对抗规则设置

首先设置任务场景边界为横纵坐标 $x,\ y\in[-4\,000\ \mathrm{m},\ 4\,000\ \mathrm{m}]$. 实际仿真中为了避免神经网络的输入值太大, 将边界归一化为 $x,\ y\in[-1.0,\ 1.0]$ 的正方形区域. 在每个回合开始时, 己方导弹集群和敌方目标集群均被初始化到固定位置, 其中第 j 架敌方目标的位置坐标为 $(x_{\min},\ y_{\min}+j\times\Delta_y)$, 第 i 个导弹的位置坐标为 $(x_{\max},\ y_{\min}+i\times\Delta_y)$. Δ_y 表示

相邻敌方目标或己方导弹在垂直方向 (y 轴方向) 上的间隔. $\Delta_y = 2.0/(N_{agent}+1)$, 其中 N_{agent} 表示对应己方导弹或敌方目标的数量, 且双方的初始设定为面对面朝向对抗态势.

在任务执行过程中, 如果敌方目标进入导弹的攻击范围, 则目标会以预先设定的概率被击毁 (本文设定为 60%). 此处攻击范围定义为 $d_{ij} < R$, $\theta_{ij} < \beta/2$, 其中 d_{ij} 代表导弹 i 到目标 j 的欧氏距离, θ_{ij} 代表对应导弹 i 相对于目标 j 的攻击角度, 攻击范围半径 $R = 0.05$ m, $\beta = \pi/10$ (等比例缩小取值) 则对应着导弹的伤害范围. 当某一方完成任务, 即导弹成功拦截所有敌方目标或敌方成功入侵目标区域时, 该对抗回合结束; 或者在达到最大训练步数时该回合自动结束.

4.1.2 状态空间

单个己方导弹的状态空间由三部分构成: 自身特征、队友特征以及敌方目标特征. 其中, 自身特征包含飞行器的二维坐标位置、二维速度向量、飞行器朝向与姿态参数以及飞行器健康状态; 队友特征涵盖各队友相对于该智能体的相对位置向量、相对距离、相对方位角、姿态参数以及队友健康状态; 敌方目标特征包含各敌方目标与该导弹的相对位置向量、相对距离、攻击角、方位角、姿态参数以及各敌方目标健康状态.

4.1.3 动作空间

为了与第 1.1 节敌方目标模型保持一致, 将敌方目标的动作空间定义为 $\{a_u\} \times \{\omega_u\}$, 其中 $\{a_u\}$、$\{\omega_u\}$ 代表轴向加速度与滚转角速度的取值集合, 分别影响目标无人机的线速度和朝向, 从而控制敌方目标运动. 轴向加速度与滚转角速度取值均限制在 $[-1, 1]$ 区间, 且目标速度限制在 $[0.005, 0.010]$ 区间, 滚转角度限制在 $[-\pi/6, \pi/6]$. 同理, 为了与第 1.2 节中的导弹模型对应, 将己方导弹的动作空间定义为 $\{a_A\} \times \{a_L\}$, 其中 $\{a_A\}$、$\{a_L\}$ 代表纵向加速度与侧向加速度的取值集合, 分别影响导弹的线速度和朝向, 取值同样限制在 $[-1, 1]$ 区间. 但考虑导弹相较于目标无人机速度快、转弯半径大, 将其速度区间设置为 $[0.005, 0.020]$, 偏航角区间设置为 $[-\pi/10, \pi/10]$.

4.1.4 奖励函数设计

本仿真需设计两套奖励函数, 分别用于己方导弹与敌方目标的训练. 己方导弹奖励用来指导学习对敌方目标的拦截策略, 敌方目标奖励则用来训练多样化的目标入侵策略, 然后通过多样化目标策略来训练和测试己方导弹集群的跨场景泛化性.

如表 3 所述, 在己方导弹执行拦截任务时, 己方奖励分为攻击奖励和约束惩罚两部分. 攻击奖励包括: 导弹靠近或者远离某个敌方目标, 对应奖励值为 $\Delta d_1 \cdot 10 \cdot \omega_{i,j}$, 其中 Δd_1 靠近目标时为正, 远离目标则为负, 用来鼓励接近目标; 当目标个体进入导弹伤害范围内且成功被击毁时, 给予较高奖励 $+20.0\omega_{i,j}$, 如果没有成功击毁, 则给予较低奖励 $+5.0\omega_{i,j}$, 以鼓励此类事件发生. 约束奖励包括: 惩罚导弹间碰撞, 当导弹间的距离小于设定的阈值时, 给予 -0.10 的惩罚; 惩罚任务执行步数, 每执行一步都会给予 -0.05 的惩罚, 用于鼓励己方导弹尽快完成拦截任务.

表 3 奖励函数设计

Table 3 Reward function design

己方导弹拦截任务	奖励设置
导弹靠近或者远离目标 j	$\Delta d_1 \cdot 10 \cdot \omega_{i,j}$
目标 j 进入导弹伤害范围且被成功击毁	$+20.0\omega_{i,j}$
目标 j 进入导弹伤害范围但未被成功击毁	$+5.0\omega_{i,j}$
队友碰撞	-0.10
步数惩罚	-0.05
敌方目标入侵任务	**奖励设置**
目标靠近或者远离目标区域	$\Delta d_2 \cdot 100/\text{Distance}$
目标到达目标区域完成任务	$+100.00$
目标进入导弹伤害范围且被击毁	-5.00
目标进入导弹伤害范围但未被击毁	-0.20
队友碰撞	-0.10
步数惩罚	-0.05

在敌方目标执行入侵任务时, 敌方奖励分为目标奖励、被击毁惩罚以及约束惩罚三个部分. 目标奖励包括: 敌方目标靠近或者远离目标区域, 对应奖励值为 $\Delta d_2 \cdot 100/\text{Distance}$, 其中 Δd_2 在靠近目标区域时为正, 远离目标区域时则为负; Distance 表示当前位置到目标区域中心的距离. 敌方目标到达目标区域即完成入侵任务, 将给予最大的奖励 $+100.00$, 以引导敌方目标接近并到达入侵目标区域. 如表 3 所述, 被击毁惩罚、约束惩罚的设计与己方导弹拦截任务的思路相似, 只存在部分数值上的差异.

4.2 实验结果

在 20 vs 10 集群拦截训练场景中, 所提算法在第 1 节介绍的环境中进行了训练和评估. 首先, 敌方集群在评估过程中采用由 checkpoints 机制获得的策略模型. 对于己方导弹集群而言, 将利用这些策略进行由易到难的测试, 以此来评估己方导弹对

不同敌方策略的适应性. 其次, 设置了己方导弹以不同规模来拦截相同规模的敌方目标, 以此测试己方策略在不同对抗规模下的适应性. 该模型训练最大回合数量设为 3000, 每个回合中步数上限为 1200 步. 基线算法 (Baseline) 设置为不添加任何模块和机制的 SAC 算法, 通过中心化训练与策略参数共享应用到多智能体环境中. 评估过程涉及到的指标包括任务成功率、平均击杀数以及平均步数.

1) SR (success rate) 是评估算法稳定性和有效性的重要指标. 如果所有敌方目标都被消灭, 则认为一次任务成功. 为了评估本文算法的性能, 本文在不同的仿真环境中统计了 100 次任务的成功率.

2) AK (average kills) 是指 100 次测试过程中每个回合击杀敌方目标的平均数量. 这一指标可有效反映任务的完成程度.

3) AS (average steps) 表示 100 次测试中己方完成任务的平均步数, 己方任务完成即所有敌方目标被消灭. 这一指标可有效反映完成任务所需时间.

如图 4 所示, 实验表明动态分组图注意力集群拦截算法能高效地完成大规模集群拦截任务并适应跨场景设置, 训练 3 000 回合后能收敛到一个稳定的集群拦截策略. 如表 4 所示, 在 20 vs 10 的训练环境中, 本文所提算法任务成功率高于 Baseline, 提升了集群拦截策略的学习效率. 此外, 针对拦截策略的跨场景泛化性同样进行多维度的测试.

首先测试本文所提算法在不同对抗规模下的适应性, 实验结果如表 4 所示. 在 20 vs 10 的训练环境中, 本文所提算法与基线算法相差不大, 任务成功率分别为 97% 与 93%, 且基线算法的平均步数甚至更少. 当训练好的策略直接应用到 15 vs 10 的测试环境时, 虽然本文所提算法与基线算法均有不同程度的性能下降, 但仍然较高效地完成了集群拦截任务. 任务成功率分别为 95% 与 91%, 但由表 4 可见, 平均步数明显增加. 当应用到 10 vs 10 的场景中, 本文所提算法与基线算法之间的差距开始显著扩大. 本文所提算法成功率为 34%, 而基线算法的任务成功率已经降至个位数 4%. 此外, 本文所提算法每回合平均击杀个数为 8.81, 明显高于基线算法 7.88. 因此, 本文基于图注意力机制与动态分组的算法在跨场景任务中表现出较好的泛化能力, 能够有效适应不同敌方策略以及不同对抗规模.

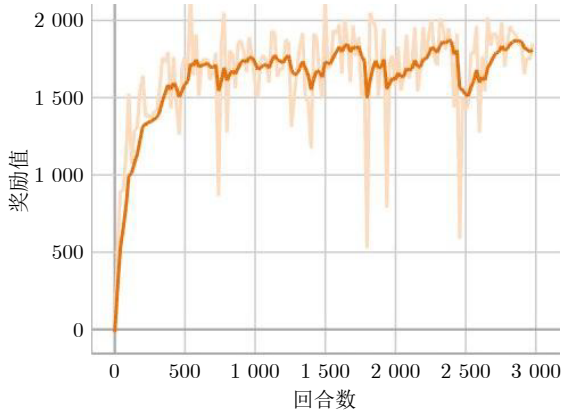


图 4　所提方法奖励函数曲线

Fig. 4　Reward function curve of proposed method

表 4　不同对抗规模实验

Table 4　Experiments with different scales of combat

己方 vs 敌方	Ours (SR/AK/AS)	Baseline (SR/AK/AS)
20 vs 10	97% / 9.97 / 220.00	93% / 9.93 / 197.48
15 vs 10	95% / 9.91 / 317.34	91% / 9.87 / 279.72
10 vs 10	34% / 8.81 / 509.00	4% / 7.88 / 455.37

随后测试己方导弹在不同敌方策略下的任务表现, 实验结果如表 5 所示. 当敌方策略设置为训练阶段的策略时, 本文训练框架得到的策略与基线算法训练得到的策略表现相当, 任务成功率分别为 97% 和 93%, 并且均具有很高的平均击杀数及很少的平均步数. 当敌方策略转变为训练时未见过的策略, 即训练外策略时, 根据难易程度 (与训练策略分布的差距), 设置“训练外策略 1”、“训练外策略 2”及“训练外策略 3”. 己方导弹在拦截相对简单的训练外策略 1 时, 因为该策略与训练策略相似度较高, 本文策略与基线算法均具有良好的任务表现. 在拦截敌方训练外策略 2 时, 基线算法相比于本文提出的算法成功率降低. 尽管在平均击杀数上相差不大 (9.99 与 9.62), 但实验发现基线算法执行任务时总会遗漏个别的敌方目标, 而且基线算法需要耗费更长的任务完成时间. 当己方导弹拦截难度更大的训练外策略 3 时, 虽然本文提出的算法也有相应的指标下降, 如任务成功率降低, 但基线算法性能下降程度更大, 任务成功率降至 62%, 且平均步数急剧增加.

最后, 从表 6 和表 7 中可以看出, 随着敌方速度 v_{train} 的提升或初始队形复杂化, 任务难度显著增加, 成功率均呈下降趋势. 然而, 本文方法在各类复杂场景下均显著优于基线方法: 在敌方速度提升

表 5　不同敌方策略实验

Table 5　Experiments with different enemy policies

敌方策略	Ours (SR/AK/AS)	Baseline (SR/AK/AS)
训练策略	97% / 9.97 / 220.00	93% / 9.93 / 197.48
训练外策略 1	100% / 10.00 / 210.35	96% / 9.96 / 186.41
训练外策略 2	99% / 9.99 / 199.08	71% / 9.62 / 420.87
训练外策略 3	95% / 9.94 / 274.66	62% / 9.30 / 962.25

100% 时仍保持 40% 的成功率, 而基线仅为 20%; 在 3-3-4 敌方队形下, 本文方法成功率为 64%, 而基线几乎完全失效 (0%). 本文所提出的动态分组图注意力拦截算法在这些跨场景测试中均表现出很好的泛化性.

与此同时还进行了失败案例分析. 如图 5(a) 所示, 在某些情况下动态分组可能未能充分反映敌方分布特征, 导致拦截分配不均; 在另外一些情况下,

表 6 不同敌方速度实验

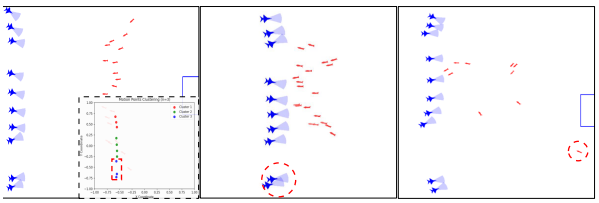
Table 6 Experiments with different enemy speed

敌方速度	Ours (SR/AK/AS)	Baseline (SR/AK/AS)
v_{train}	97% / 9.97 / 220.00	93% / 9.93 / 197.48
$v_{train}(1+50\%)$	85% / 9.81 / 198.45	38% / 8.44 / 251.29
$v_{train}(1+100\%)$	40% / 8.60 / 197.75	20% / 7.25 / 196.30

表 7 不同敌方队形实验

Table 7 Experiments with different enemy formation

敌方队形	Ours (SR/AK/AS)	Baseline (SR/AK/AS)
一字	97% / 9.97 / 220.00	93% / 9.93 / 197.48
3-3-4	64% / 9.16 / 459.36	0% / 4.76 / 282.36



(a) 分布特征不合理 (a) Unreasonable distribution characteristics (b) 部分目标被忽视 (b) Ignore some targets (c) 奖励函数引导不足 (c) Insufficient guidance by the reward function

图 5 失败案例可视化

Fig. 5 Failure case visualization

如图 5(b) 所示, 注意力机制存在权重过度集中现象, 从而使部分目标被忽视; 此外, 如图 5(c) 所示, 观察到奖励函数在特定场景下引导不足, 可能导致个体导弹行为偏离全局最优. 因此本文所提算法在动态分组与关系捕获方面仍存在改进空间.

4.3 消融实验

本部分进行消融实验, 主要包含以下几种设置:

1) 包含 MARL、动态分组以及图注意力模块的完整框架;

2) 包含 MARL、动态分组以及图注意力模块, 但去除差异化奖励函数机制;

3) 包含 MARL、动态分组以及图注意力模块, 无差异化奖励函数, 但 MARL 模块的输入同时包含了原始观测与图注意力模块的融合特征;

4) 仅使用 MARL 与图注意力模块;

5) 无动态分组模块与图注意力模块的基础框架.

消融实验中己方导弹的策略由上述设置对应的框架训练得到, 训练曲线如图 6 所示. 因设置 1) 修改了奖励函数, 所以不参与奖励函数曲线的对比, 设置 1) 的奖励函数曲线见图 4. 在消融实验中, 敌方目标使用固定不变的训练外策略.

从表 8 中的测试结果可以看出, 差异化奖励函数机制对于整体任务的成功率提升幅度较小, 但设置 2) 中去除该机制后完成任务所需的平均步数增加, 因此差异化奖励函数的设计能在一定程度上加快任务执行速度, 优化目标分配效率. 此外, 在去除差异化奖励函数的情况下, 通过原始观测增强策略网络输入, 设置 3) 虽然完成任务的平均步数仍然较多, 但任务成功率和平均击杀数指标会优于设置 1) 的完整框架. 这说明使用图注意力网络融合信息与

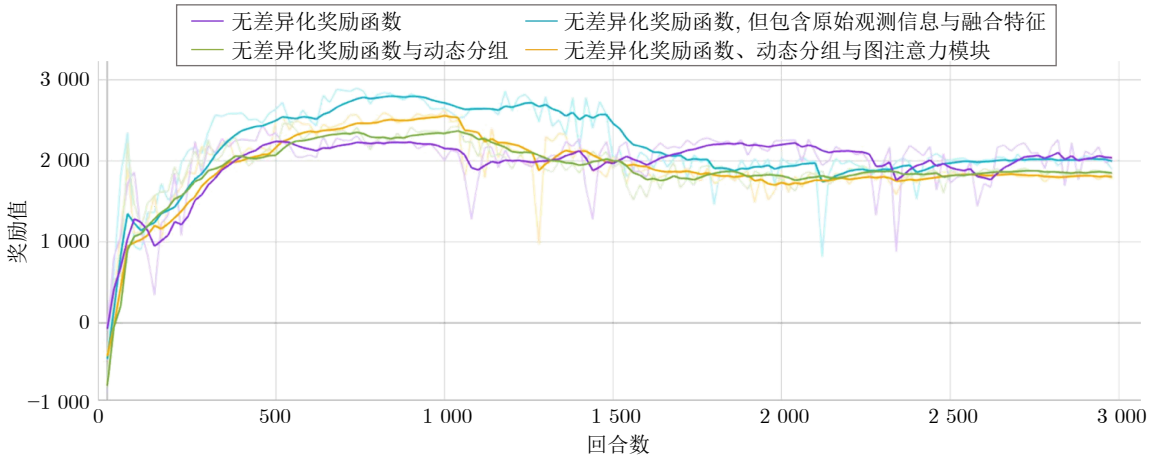


图 6 消融实验不同设置奖励函数曲线对比

Fig. 6 Comparison of reward function curves in ablation experiments with different settings

表 8　不同消融设置实验

Table 8　Experiments of different ablation settings

消融设置	SR (%)	AK	AS
设置 1)	98	9.96	187.60
设置 2)	95	9.90	315.33
设置 3)	100	10.00	201.18
设置 4)	80	9.33	387.47
设置 5)	65	9.31	783.25

原始观测相比可能存在信息丢失, 但使用原始观测进行增强的代价就是状态–动作空间随着规模不断增加, 不利于提升学习效率.

设置 4) 去除了动态分组模块, 虽然每回合平均拦截敌方目标数超过 9 个 (一共 10 个敌方目标), 但对于任务成功率和平均步数的影响较大. 由于没有动态分组模块, 每个己方导弹面对的目标数量较多, 目标分配的难度也相应增加, 因此出现目标遗漏的概率也在增加, 直接影响任务完成效率. 设置 5) 中进一步去除了图注意力模块, 己方导弹在面对大规模目标的同时也缺乏发现目标相对关系的能力, 进一步影响己方导弹的目标分配效率, 如图 7 所示. 更重要的是, 在敌方测试策略与训练策略分布相差较大时, 图注意力机制能够通过发掘敌我相对关系来进行高效的策略迁移, 增强跨场景泛化性. 去除图注意力模块之后将直接影响己方导弹策略面对训练场景外策略的性能, 导致更低的任务完成率以及更多的任务完成步数.

4.4　算法复杂度

本节分别对所提算法在训练与执行阶段的复杂度进行简要分析. 训练阶段的计算主要来源于图注意力网络与 MARL 中策略–价值网络的更新过程, 引入参数共享与动态分组机制降低输入维度, 有效降低了计算复杂度; 而引入图注意力模块与对抗性策略池在提升策略泛化性的同时会增加计算复杂度. 相比于基线算法, 本文方法在训练阶段增加了一定的计算复杂度, 但其显著提升了策略的跨场景泛化性, 达到了计算复杂度与算法性能的平衡.

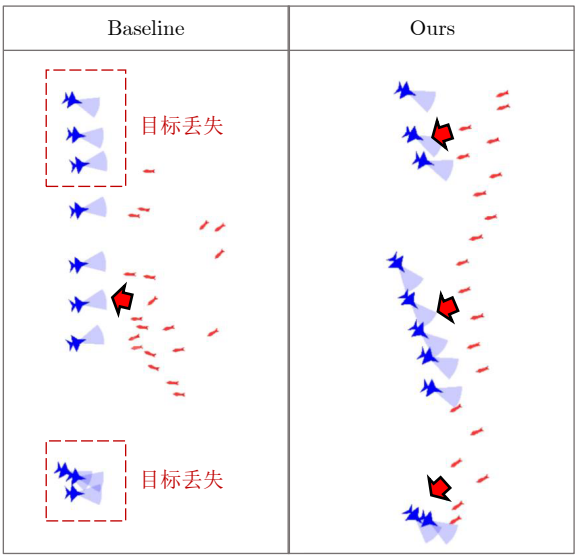


图 7　目标分配可视化对比

Fig. 7　Target allocation visualization comparison

执行阶段的复杂度主要来自于动态分组、图注意力模块推理以及策略网络前向计算, 整体复杂度为 $\mathrm{O}(h((m+n)d^2+mnd)+(d_{in}d_h+(L_a-2)d_h^2+d_hd_{out}))+\mathrm{O}((T_{\max}n_end_x)/\tau)$, 其中 h 为多头注意力机制头数, d_{in}、d_h、d_{out} 和 L_a 分别为策略网络的输入、隐藏层、输出维度以及网络层数, n_e 为敌方目标数量, τ 为聚类分组周期, d_x 为每个敌方目标参与聚类的原始空间特征维度. 事实上, 由于网络规模较小, 各参数取值均很小, 因而执行阶段的计算开销非常小.

5　结束语

本文提出一种基于动态分组与图注意力机制的集群协同拦截框架, 通过对敌方集群进行目标分组, 降低决策维度, 提升了算法学习效率. 利用图注意力网络捕获敌我对抗关系, 实现精准目标分配, 提升策略跨场景泛化性, 并设计差异化奖励函数将敌我关系与拦截行为深度耦合. 实验通过多维度对比验证了框架在策略泛化性与策略稳定性上的优势, 并结合消融实验验证了模块的有效性. 同时实验发现当前图注意力网络的特征融合能力仍有优化空间, 后续工作将继续关注在复杂动态场景下的高效特征融合方法. 此外, 当前方法局限在简化动力学模型的二维场景, 与实际大规模集群拦截任务设定仍有差距, 后续工作将进一步研究高保真环境中的大规模集群对抗算法, 以提升算法实用性.

参考文献

1　Jia Yong-Nan, Tian Si-Ying, Li Qing. Recent development of unmanned aerial vehicle swarms. *Acta Aeronautica et Astronautica Sinica*, 2020, **41**(S1): 723−738
(贾永楠, 田似营, 李擎. 无人机集群研究进展综述. 航空学报, 2020, **41**(S1): 723−738)

2　Xue Jian, Zhao Lin, Xiang Xian-Cai, Lv Ke, Hong Chen, Zhang Bao-Lin, et al. A review of the research on UAV swarm confrontation under incomplete information. *Journal of Electronics and Information Technology*, 2024, **46**(4): 1157−1172
(薛健, 赵琳, 向贤财, 吕科, 宏晨, 张宝琳, 等. 非完全信息下无人机集群对抗研究综述. 电子与信息学报, 2024, **46**(4): 1157−1172)

3　Gao Shu-Yi, Lin De-Fu, Zheng Duo, Hu Xin-Yu. Intelligent cooperative interception strategy of aircraft against cluster attack. *Acta Aeronautica et Astronautica Sinica*, 2023, **44**(18): 276−291
(高树一, 林德福, 郑多, 胡馨予. 针对集群攻击的飞行器智能协同

拦截策略. 航空学报, 2023, **44**(18): 276−291)

4 Duan H B, Zhang D F, Fan Y M, Deng Y M. From wolf pack intelligence to UAV swarm cooperative decision-making. *Science China Information Sciences*, 2019, **49**: 112−118

5 Zhou Mo, Sun Hai-Wen, Wang Liang, Yu Shao-Zhen, Meng Xiang-Yao, Li Dan. Research on foreign anti-UAV swarm warfare. *Command Control and Simulation*, 2023, **45**(2): 24−30
(周末, 孙海文, 王亮, 于邵祯, 孟祥尧, 李丹. 国外反无人机蜂群作战研究. 指挥控制与仿真, 2023, **45**(2): 24−30)

6 Luo R N, Huang S C, Zhao Y. Guidance strategy of mother-son missile against unmanned aerial vehicle cluster. *Systems Engineering and Electronics*, 2023, **45**(10): 3249−3258

7 Burgin G H, Sidor L B. Rule-based Air Combat Simulation, Technical Report TITAN-TLJ-H-1501, Titan Systems Inc., USA, 1988.

8 Park H, Lee B Y, Tahk M J. Differential game based air combat maneuver generation using scoring function matrix. *International Journal of Aeronautical and Space Sciences*, 2016, **17**(2): 204−213

9 Sunehag P, Lever G, Gruslys A, Czarnecki W M, Zambaldi V, Jaderberg M, et al. Value-decomposition networks for cooperative multi-agent learning based on team reward. In: Proceedings of the 17th International Conference on Autonomous Agents and Multi-agent Systems. Stockholm, Sweden: ACM, 2018. 2085–2087

10 Rashid T, Samvelyan M, de Witt C S, Farquhar G, Foerster J, Whiteson S. Monotonic value function factorisation for deep multi-agent reinforcement learning. *Journal of Machine Learning Research*, 2020, **21**(178): 1−51

11 Yu Wen-Wu, Yang Xiao-Ya, Li Hai-Chang, Wang Rui, Hu Xiao-Hui. Attentional intention and communication for multi-agent learning. *Acta Automatica Sinica*, 2023, **49**(11): 2311−2325
(俞文武, 杨晓亚, 李海昌, 王瑞, 胡晓惠. 面向多智能体协作的注意力意图与交流学习方法. 自动化学报, 2023, **49**(11): 2311−2325)

12 Wang Yao-Nan, Hua He-An, Zhang Hui, Zhong Hang, Fan Ye-Xin, Liang Hong-Tao, et al. Performance function-guided deep reinforcement learning control for UAV swarm. *Acta Automatica Sinica*, 2025, **51**(5): 905−916
(王耀南, 华和安, 张辉, 钟杭, 樊叶心, 梁鸿涛, 等. 性能函数引导的无人机集群深度强化学习控制方法. 自动化学报, 2025, **51**(5): 905−916)

13 Zheng Z, Wei C, Duan H. UAV swarm air combat maneuver decision-making method based on multi-agent reinforcement learning and transferring. *Science China Information Sciences*, 2024, **67**(8): 180−204

14 Yu C, Velu A, Vinitsky E, Gao J, Wang Y, Bayen A, et al. The surprising effectiveness of PPO in cooperative multi-agent games. In: Proceedings of 2022 Advances in Neural Information Processing Systems. New Orleans, USA: Curran Associates, Inc., 2022. 24611–24624

15 Xuan S Z, Ke L J. UAV swarm attack-defense confrontation based on multi-agent reinforcement learning. In: Proceedings of the 2020 International Conference on Guidance, Navigation and Control. Singapore: Springer, 2021. 5599–5608

16 Lowe R, Wu Y, Tamar A, Harb J, Abbeel P, Mordatch I. Multi-agent actor-critic for mixed cooperative-competitive environments. In: Proceedings of 2017 Advances in Neural Information Processing Systems. Long Beach, USA: Curran Associates, Inc., 2017. 6379–6390

17 Wang B, Li S, Gao X, Xie T. UAV swarm confrontation using hierarchical multiagent reinforcement learning. *International Journal of Aerospace Engineering*, 2021, **2021**: Article No. 3360116

18 Pope A P, Ide J S, Mićović D, Diaz H, Rosenbluth D, Ritholtz L, et al. Hierarchical reinforcement learning for air-to-air combat. In: Proceedings of the 2021 International Conference on Unmanned Aircraft Systems. Athens, Greece: IEEE, 2021. 275–284

19 Cai H, Li X, Zhang Y, Gao H. Interception of a single intruding unmanned aerial vehicle by multiple missiles using the novel EA-MADDPG training algorithm. *Drones*, 2024, **8**(10): Article No. 524

20 Han Y, Piao H, Hou Y, Sun Y, Sun Z, Zhou D. Deep relationship graph reinforcement learning for multi-aircraft air combat. In: Proceedings of the International Joint Conference on Neural Networks. Padua, Italy: IEEE, 2022. 1–8

21 Scarselli F, Gori M, Tsoi A C, Hagenbuchner M, Monfardini G. The graph neural network model. *IEEE Transactions on Neural Networks*, 2008, **20**(1): 61−80

22 Sun Z, Wu H, Shi Y, Yu X, Gao Y, Pei W, et al. Multi-agent air combat with two-stage graph-attention communication. *Neural Computing and Applications*, 2023, **35**(27): 19765−19781

23 Foerster J, Nardelli N, Farquhar G, Afouras T, Torr P, Kohli P, et al. Stabilising experience replay for deep multi-agent reinforcement learning. In: Proceedings of the International Conference on Machine Learning. Sydney, Australia: PMLR, 2017. 1146–1155

24 Koh W, Oh W, Kim S, Shin S, Kim H, Jang J, et al. FlickerFusion: Intra-trajectory domain generalizing multi-agent reinforcement learning. In: Proceedings of the 13th International Conference on Learning Representations. Singapore: OpenReview.net, 2025. 59197–59234

25 Cao Y, Kou Y, Xu A, Xi Z. Target threat assessment in air combat based on improved glowworm swarm optimization and ELM neural network. *International Journal of Aerospace Engineering*, **2021**(1): Article No. 4687167

26 Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph attention networks. arXiv preprint arXiv: 1710.10903, 2017.

27 Haarnoja T, Zhou A, Hartikainen K, Tucker G, Ha S, Tan J, et al. Soft actor-critic algorithms and applications. arXiv preprint arXiv: 1812.05905, 2018.

28 Liu E, Zhu J, Lin Z, Ning X, Wang S, Blaschko M B, et al. Linear combination of saved checkpoints makes consistency and diffusion models better. arXiv preprint arXiv: 2404.02241, 2024.

29 Vinyals O, Babuschkin I, Czarnecki W M, Mathieu M, Dudzik A, Chung J, et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 2019, **575**(7782): 350−354

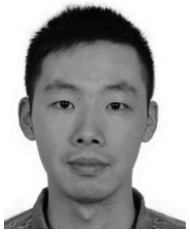
30 Yuan G, He M, Ma Z, Zhang W, Liu X, Li W. Multiagent following multileader algorithm based on K-means clustering. *Journal of System Simulation*, 2023, **35**(3): 616−622

31 Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, et al. Attention is all you need. In: Proceedings of Advances in Neural Information Processing Systems. Long Beach, USA: Curran Associates, Inc., 2017. 6000–6010

32 Tao F, Wu M, Cao Y. Generalized maximum entropy reinforcement learning via reward shaping. *IEEE Transactions on Artificial Intelligence*, 2023, **5**(4): 1563−1572

吴子鹏 中国科学技术大学自动化系博士研究生. 主要研究方向为多智能体强化学习和人-AI 协作.
E-mail: zipengwu@mail.ustc.edu.cn
(**WU Zi-Peng** Ph.D. candidate in the Department of Automation, University of Science and Techno-

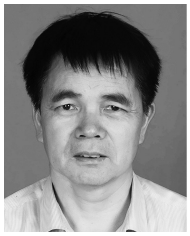
logy of China. His research interests include MARL and human-AI collaboration.)



马麒超 中国科学技术大学自动化系副教授. 主要研究方向为自主智能集群系统决策与控制, 多智能体博弈与强化学习.
E-mail: qcma@ustc.edu.cn
(**MA Qi-Chao** Associate professor in the Department of Automation, University of Science and Technology of China. His research interests include decision-making and control of autonomous intelligent swarm systems, multi-agent game and RL.)

秦家虎 中国科学技术大学自动化系教授. 主要研究方向为自主智能系统, 移动机器人自主导航与具身操作, 人–机交互. 本文通信作者.
E-mail: jhqin@ustc.edu.cn
(**QIN Jia-Hu** Professor in the Department of Automation, University of Science and Technology of China. His research interests include autonomous intelligent systems, autonomous navigation and embodied manipulation of mobile robots, and human-machine interaction. Corresponding author of this paper.)

詹晨光 空基信息感知与融合全国重点实验室工程师. 主要研究方向为任务规划和需求论证.
E-mail: zhanchenguang@qq.com
(**ZHAN Chen-Guang** Engineer at the National Key Laboratory of Air-based Information Perception and Fusion. His research interests include task planning and requirements justification.)



张金鹏 空基信息感知与融合全国重点实验室研究员. 主要研究方向为制导, 导航与控制.
E-mail: zhangapengly@163.com
(**ZHANG Jin-Peng** Researcher at the National Key Laboratory of Air-based Information Perception and Fusion. His research interests include guidance, navigation and control.)