

面向有限通信约束下多智能体对抗的团队博弈决策研究

任智¹, 张栋^{1†}, 唐硕¹, 熊威¹, 杨书恒¹, 刘文逸²

(1. 西北工业大学 航天学院, 陕西 西安 710129; 2. 西北机电工程研究所, 陕西 西安 712000)

摘要: 随着无人系统平台的快速发展, 无人集群对抗是未来体系能力发展与建设的关键。然而, 在应对集群目标时, 集群间的对抗博弈加剧了有限通信约束下通信带宽压力、链路动态变化和消息时延积累等问题, 集群对抗决策面临态势信息过载、协同能力下降与团队响应滞后的挑战。针对上述问题, 首先构建了多智能体对抗的运动学模型与对抗损耗模型, 并在此基础上提出了有限通信约束下多智能体对抗的团队博弈建模与求解方法。通过设计小世界网络通信的团队联盟策略, 将有限通信约束引入多智能体对抗的团队博弈模型, 给出了基于关联团队最大最小均衡解概念的团队博弈均衡定义与存在性条件。然后, 基于异构智能体强化学习算法设计了交替冻结训练的团队博弈均衡求解算法。最后, 仿真结果表明, 所提出的博弈模型与交替冻结训练框架能够获得稳定的策略剖面, 实现了有限通信条件下的集群自适应动态决策与对抗策略优化。

关键词: 集群对抗; 团队博弈; 小世界网络; 有限通信; 强化学习; 群体决策

中图分类号: TP18 文献标志码: A

DOI: 10.13195/j.kzyjc.2025.1188

引用格式: 任智, 张栋, 唐硕, 等. 面向有限通信约束下多智能体对抗的团队博弈决策研究 [J]. 控制与决策, xxxx, x(x): xxxx-xxxx.

Research on team game decision-making for multi-agent confrontation under limited communication constraints

REN Zhi¹, ZHANG Dong^{1†}, TANG Shuo¹, XIONG Wie¹, YANG Shu-heng¹, LIU Wen-yi²

(1. School of Astronautics, Northwestern Polytechnical University, Xi'an 710129, China; 2. Northwest Institute of Mechanical and Electrical Engineering, Xianyang 712000, China)

Abstract: With the rapid development of unmanned system platforms, unmanned swarm confrontation has become crucial for future development and construction of system capabilities. However, when engaging swarm targets, adversarial interactions between swarms exacerbate issues such as communication bandwidth pressure, dynamic link variations, and accumulated information delays under limited communication constraints. Swarm confrontation decision-making faces challenges including overloaded situational information, degraded collaborative capabilities, and delayed team responses. To address these problems, this study first establishes kinematic models and attrition models for multi-agent confrontation, and subsequently proposes a team game modeling and solution approach for multi-agent confrontation under limited communication constraints. By designing a team alliance strategy based on small-world network communication, limited communication conditions are incorporated into the team game model for multi-agent confrontation. The definition and existence conditions of team game equilibrium are provided, based on the concept of correlated team max-min equilibrium solutions. Then, a heterogeneous multi-agent reinforcement learning algorithm is developed to design an alternating freeze-training method for solving the team game equilibrium. Finally, the simulation results demonstrate that the proposed game model and alternating freezing training framework can obtain stable strategy profiles, achieving adaptive dynamic decision-making and adversarial strategy optimization under limited communication conditions.

Keywords: swarm confrontation; team game; small world network; limited communication; reinforcement learning; group decision-making

收稿日期: xxxx-xx-xx; 录用日期: 2026-03-25.

基金项目: 国家自然科学基金资助 (NO: 52472417).

†通信作者. E-mail: zhangdong@nwpu.edu.cn.

0 引言

随着多智能体系统和无人集群平台的快速发展,智能体集群具备高效的执行能力与环境适应性,能够在复杂环境中完成广域巡查、动态追踪与多点协作等多种任务,在未来动态空域争夺、对抗性集群围捕与多目标抢占驱逐等典型场景展现了巨大潜力与广阔前景^[1].因此,多智能体集群的动态博弈对抗不仅是智能集群的必然发展趋势,更是未来无人系统平台发展与体系能力建设的核心要求.然而,随着集群规模扩展与目标复杂度提高,现有单平台节点通信带宽与计算资源有限,难以满足智能体集群对抗任务的实际需求.此外,高动态复杂对抗环境中通信链路易受扰动,使得智能体集群协同效率下降,决策响应滞后.因此,如何在上述有限通信约束下实现集群自适应决策与最优响应,是未来多智能体对抗应用亟需解决的关键问题.

多智能体对抗问题最初以疆域防御、追逃博弈和团队博弈为代表,通过矩阵博弈、微分对策等博弈论方法^[2-4],建立面向敌我之间对抗的最优响应关系模型.在这些经典框架下,学者们致力于通过解析纳什均衡等理论结果,揭示有限局中人个体间的策略互动机理^[5].然而,多智能体集群对抗的实时态势高度动态变化且参与个体数量庞大,传统博弈模型难以应对策略空间的爆炸性增长,对应纳什均衡策略的求解也面临维度灾难与收敛慢等问题,难以支持大规模多智能体集群的实时决策.随着深度强化学习技术的发展,神经网络因其较强的拟合能力被引入均衡策略的求解过程中^[6-8].多智能体博弈问题能够在强化学习架构的基础上,通过构建博弈双方的策略空间,训练对应敌方集群策略最佳响应的策略模型,实现对应博弈模型均衡的求解.

现有多智能体对抗研究多以分层优化的决策框架为基础,自上而下地建立多智能体对抗具体任务目标与约束模型,采用博弈论、优化理论^[9-11]、人工智能^[12-15]等方法实现集群系统的动态分工与协同对抗.其本质均在于建立全局统一的目标函数与协同机制,将多智能体系统视为多个功能单元,基于集中式的信息感知和分布式的任务规划,对个体进行目标分配、航迹规划、队形保持以实现集群的动态分工与协同对抗^[16-19].然而,在实际多智能体对抗场景中,态势信息随集群规模指数级增长,需要处理和转发的信息容易超出单个节点的通信负载与计算资源.而关键节点的失效与通信链路的动态变化也容易导致关键信息的延迟、丢失或冗余,影响集群整体的协同决

策效率.此外,现有的有模型决策方法依赖于先验的任务类型与协同策略^[20-23],而多智能体对抗的专家经验尚未形成共识,对应模型决策结果的可信度低与可扩展性低.

围绕有限通信约束问题,现有国内外研究集中于网络化多智能体系统的协同控制,将有限通信抽象为邻域节点间交互拓扑,研究拓扑切换、链路失效与信息时延等对一致性控制性能的影响^[24-26].然而,此类研究多以单个集群系统为对象,将通信网络视为一致性收敛的约束条件.当协同任务转为对抗任务时,通信拓扑与敌我交互、联盟形成以及观测信息相互耦合,传统多智能体系统在有限通信约束下的协同行为一致性控制方法难以适应集群对抗任务需求.对此,复杂网络领域中广泛采用的小世界网络模型^[27,28],恰好贴近了集群对抗过程中全局协同、分簇对抗的结构特征,为分析集群对抗中的协同对抗行为与通信拓扑演化提供了工具.在此基础上,网络博弈领域将个体之间的策略交互约束在给定网络结构上,研究对应网络中个体的最优响应与博弈均衡^[29,30].然而,现有的网络博弈研究结论往往基于随机生成的静态网络,较少考虑通信距离受限与链路时变的动态网络,难以直接映射到集群对抗中拓扑信息与对抗响应耦合演化的决策过程.

针对上述问题,本文基于自组织规则构建面向多智能体对抗的群体动力学模型,通过引入小世界网络表征有限通信约束下的集群相互作用关系与团队联盟机制,将具体的多智能体对抗任务转化为个体邻域交互的自组织规则,构建有限通信约束下的团队博弈模型,自下而上地优化集群对抗的战术策略,实现多智能体对抗的自适应动态决策与最优响应.本文的主要贡献在于:1)提出了一种以协同对抗交互权重为核心的多智能体决策参数化建模方法,通过动态参数配置实现复杂集群行为与协同对抗策略,相较于直接在高维联合动作空间中求解的方法降低了决策变量维度;2)提出了一种基于小世界网络的团队博弈建模方法,通过小世界网络表征团队联盟通信拓扑,并据此构造通信邻域与局部信息集,将有限通信约束显式纳入博弈要素与可行策略域,实现了有限通信约束下团队博弈的模型扩展与均衡定义;3)提出了一种基于交替冻结训练的博弈均衡求解算法,并在大规模多智能体对抗仿真中验证了其能够获得稳定策略剖面,实现了有限通信约束下多智能体对抗策略的近似均衡求解.

1 问题描述与建模

为刻画有限通信约束下多智能体对抗的动态演化与决策耦合关系,本节从对抗场景出发,依次建立智能体运动学模型、局部交互机制与攻击损耗模型,为后续团队博弈建模与策略求解提供模型基础。

1.1 对抗场景想定描述

本文以空中无人集群平台为对象,研究的对抗场景考虑典型二维空域内多智能体集群对多智能体集群的攻防对抗,如图1所示。

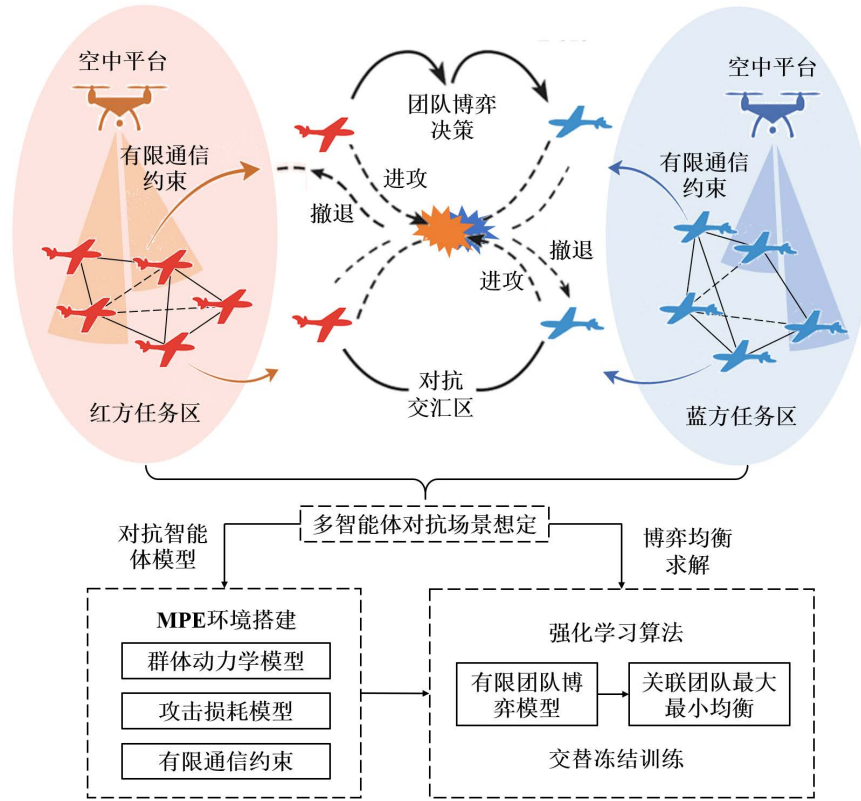


图1 多智能体对抗示意图

在传统单个多智能体系统模型的基础上,将集群对抗系统整体划分为红方和蓝方两个相互对抗的子群,分别代表两支规模相等且能力一致的智能体集群。智能体集群由空中平台在任务区投放,并通过有限距离通信实现信息共享与策略协同,最终依据对抗集群的群体动力学模型进行对抗交互。记对抗集群的智能体集合为 $\mathcal{N}\{r_1, r_2, \dots, r_n, b_1, b_2, \dots, b_n\}$,其中, r 为红方集群个体, b 为蓝方集群个体。假设红蓝双方智能体在初始阶段随机分布在各自的任务区内。对抗集群双方均需要在保证自身存活率的前提下实现对敌方多智能体的拦截,避免敌方多智能体对我方任务区的攻击。为便于讨论分析,在后续的论述中以红方为己方视角,研究以蓝方作为敌方目标集群的多智能体对抗。在对集群中某一个具体智能体个体分析时,记其与所属阵营标识为 i ,对应敌对阵营标识为 $-i$,对应相同阵营友方智能体个体标识为 j ,对应不同阵营敌方智能体个体标识为 k 。

针对上述多智能体对抗场景想定,通过结合集

群对抗的群体动力学模型、攻击损耗模型与有限通信交互场景,在多智能体粒子环境框架 (Multi-Agent Particle Environment, MPE) 中重新搭建对应多智能体对抗的各类模型与对抗环境,以便于强化学习算法求解对应团队博弈均衡时的动作状态转移与效用函数计算。在此基础上,依据有限通信交互条件建立团队博弈模型,引入关联团队最大最小均衡解的概念,并通过交替冻结训练实现对应团队博弈均衡的求解,实现多智能体对抗博弈决策。

1.2 智能体运动学模型

在群体动力学模型中,智能体间的相互作用关系依赖于智能体的感知能力及通信机制。在实际对抗过程中,无人平台的速度通常稳定在巡航速度,其机动主要通过航向调整来实现。设初始时刻对抗集群系统中智能体 i 以固定速度 v_0 前进,在 t 时刻的位置向量为 $\mathbf{x}_i(t) \in \mathbb{R}^2$,对应智能体的运动方向为 $\theta_i(t) \in (-\pi, \pi]$,其动作状态转移的方向角更新由友方协同交互与敌方对抗交互共同决定。为表征实际

集群对抗系统中智能体复杂关系,通过引入通信半径 R_c 、感知半径 R_s 与攻击半径 R_a 三个关键个体能力参数,分别对应智能体在集群对抗中的通信交互、态势感知与对抗响应.其中, $R_c, R_s, R_a > 0$ 为标量常量,分别表示通信、感知与攻击作用的最大距离.

设 $\|\cdot\|$ 表示 2-范数,以智能体 i 为例,其通信半径 R_c 决定了其与友方邻居智能体 j 间信息交互的邻域 $\mathcal{N}_i^u\{j\|\mathbf{x}_j - \mathbf{x}_i\| \leq R_c, j \neq i\}$,集群内部智能体仅在相对距离小于通信半径时进行信息同步;感知半径 R_s 确定了与探测范围内敌方智能体 k 之间存在对抗作用的邻域 $\mathcal{N}_i^v\{k\|\mathbf{x}_k - \mathbf{x}_i\| \leq R_s, k \neq i\}$,集群间敌对智能体仅在相对距离小于感知半径时存在对抗关系响应;攻击半径 R_a 则决定了对敌方智能体 $\mathcal{N}_i^a\{k\|\mathbf{x}_k - \mathbf{x}_i\| \leq R_a, k \neq i\}$ 的毁伤范围,智能体仅能够在攻击范围内对敌方智能体造成毁伤.

在上述空间尺度划分的基础上,为刻画有限通信约束下的集群协同与对抗交互行为,参考 flocking 二维运动学模型^[31],依据智能体的邻域航向信息,构造智能体的局部作用规则.记智能体 i 在时刻 t 的速度向量为 $\mathbf{v}_i(t) = v_0[\cos \theta_i(t), \sin \theta_i(t)]^T$,则对应个体间协同与对抗的局部相互作用规则可建模为期望速度方向的融合.具体而言,智能体 i 在时刻 t 通过获取其通信邻域与感知邻域内的局部信息生成两类期望方向:一类是基于通信邻域内友方信息融合得到的友方协同交互作用项 ψ_i ,用于表征类 flocking 集群的一致性机动趋势;另一类是基于感知邻域内敌方相对距离与方位构造的敌方对抗交互作用项 ϕ_i ,用于表征“近斥远引”的追逃机动趋势.其中,智能体 i 的友方协同交互作用项 ψ_i 依据友方平均运动方向确定,如式 (1) 所示:

$$\psi_i(t+1) = \arctan\left(\frac{\sum_{j \in \mathcal{N}_i^u(t)} \sin \theta_j(t)}{\sum_{j \in \mathcal{N}_i^u(t)} \cos \theta_j(t)}\right). \quad (1)$$

其中, ψ_i 表征对应智能体 i 通信半径范围内友方智能体的平均运动方向.

当智能体 i 感知范围内的对抗邻域存在敌方智能体 $k \in \mathcal{N}_i^v$ 时,敌方对抗交互作用项需要依据攻击半径确定近斥远引的对抗作用,如式 (2) 所示:

$$\phi_i(t+1) = \arctan\left(\frac{\sum_{k \in \mathcal{N}_i^v(t)} \mathbf{d}_{ik,y}(t)}{\sum_{k \in \mathcal{N}_i^v(t)} \mathbf{d}_{ik,x}(t)}\right). \quad (2)$$

其中, ϕ_i 表征对应智能体 i 感知半径范围内敌方智能体的对抗交互响应方向, \mathbf{d}_{ik} 为智能体 i 对应敌方智

能体 k 推拉对抗作用响应的单位方向向量.当双方距离小于攻击半径时,智能体 i 需要规避并远离敌方智能体 k ,反之则靠近,具体推拉对抗作用向量的分段计算公式如下:

$$\mathbf{d}_{ik}(t) = \begin{cases} \frac{\mathbf{x}_i(t) - \mathbf{x}_k(t)}{\|\mathbf{x}_i(t) - \mathbf{x}_k(t)\|}, & \|\mathbf{x}_i(t) - \mathbf{x}_k(t)\| < R_a \\ \frac{\mathbf{x}_k(t) - \mathbf{x}_i(t)}{\|\mathbf{x}_i(t) - \mathbf{x}_k(t)\|}, & \|\mathbf{x}_i(t) - \mathbf{x}_k(t)\| \geq R_a \end{cases} \quad (3)$$

由此,将两类相互作用通过加权参数进行线性融合,确定对抗集群系统智能体的运动学方程,即:

$$\begin{cases} \theta_i(t+1) = \alpha_i \cdot \phi_i(t+1) + (1 - \alpha_i) \cdot \psi_i(t+1) \\ \mathbf{x}_i(t+1) = \mathbf{x}_i(t) + v_0[\cos \theta_i(t+1), \sin \theta_i(t+1)]^T \end{cases} \quad (4)$$

其中, $\alpha_i(t) \in (0, 1)$ 为对应智能体 i 友方协同交互作用项与敌方对抗交互作用项的比例因子,反映了对抗集群智能体对友方协同与敌方对抗响应的决策偏好: α_i 越小,表明智能体 i 更注重与友方智能体策略一致性的同步, α_i 越大,表明智能体 i 更注重与敌方的对抗响应.由此,能够通过控制权重因子 α_i 实现集群系统整体自组织行为涌现的优化控制.

需要说明的是,在上述建立的二维离散运动学设定下,智能体的运动状态由位置 \mathbf{x}_i 与航向角 θ_i 表征,式 (1)–(3) 构建了有限通信约束下的局部相互作用机制,即友方协同交互作用项 ψ_i 与敌方对抗交互作用项 ϕ_i ,作为行为规则表征了智能体在集群对抗过程中局部协同与对抗交互对航向的影响.在此基础上,式 (4) 中权重参数 α_i 用于在协同与对抗两类期望速度方向的加权融合,进而驱动智能体的位置与航向更新.由此,在后续的团队博弈建模与强化学习求解中,将权重参数 α_i 作为控制输入的动作变量进行优化,对应的博弈策略 $\pi_i(\alpha_i | H_i)$ 即为智能体在有限通信约束的信息集 H_i 下对 α_i 的选择规律.

1.3 对抗攻击损耗模型

在智能体对抗过程中,智能体会持续对攻击半径范围内的敌方智能体发起攻击.为描述多智能体对抗过程中的个体攻击损耗与死亡机制,参考传染病模型中的损伤扩散思想,引入基于概率的攻击损耗过程,描述红蓝双方智能体的攻击损耗演化.

假设每个集群智能体均具有有限攻击半径,当敌方智能体 $k \in \mathcal{N}_i^a$ 进入自身攻击半径范围内,即可对其实施攻击.记任意时刻 t 集群系统内的每个智能体 i 生命值为 $L_i(t) \in [0, H]$,其中 H 为常数,表示初始时刻智能体 i 的生命值.在初始时刻,所有智能体

处于相同的存活状态,且具备自主运动与对敌攻击的能力.随着集群对抗进程的推进,当敌方智能体 k 进入的智能体 i 攻击半径时依概率 p_{ik} 触发一次攻击事件,同时智能体 i 可依据相对运动关系以概率 μ_{ik} 规避毁伤,否则其生命值降低,直至累计为零时进入“死亡”状态.死亡的智能体将失去运动与攻击能力,不再参与后续集群的对抗博弈过程.设智能体 i 与其攻击半径内的敌方智能体 k 之间的相对位置为 \mathbf{r}_{ik} ,则攻击事件的触发概率可表示为:

$$p_{ik}(t) = \Psi\left(\frac{\|\mathbf{r}_{ik}(t)\|^2}{R_a^2}\right). \quad (5)$$

其中, $\Psi(\cdot)$ 为服从指数分布的概率分布函数,表征智能体攻击造成毁伤的成功概率随距离的增大而减小.Theodoros等^[32]指出,攻击造成毁伤的概率分布函数的具体数学表达形式并不影响个体间的攻击损耗结果,仅需要满足如下变化趋势:当自变量取零时,函数值应为1,且随着该变量的逐渐增大,函数值需呈现平滑且连续的递减特性.

同时,智能体 i 在遭受敌方智能体 k 攻击造成的毁伤时,受智能体间相对位置 \mathbf{r}_{ik} 与相对速度 \mathbf{v}_{ik} 影响而能够以概率 μ_{ik} 成功规避攻击免于毁伤损耗,即:

$$\mu_{ik}(t) = \Psi\left(2 - \frac{\mathbf{r}_{ik}(t) \cdot \mathbf{v}_{ik}(t)}{\|\mathbf{r}_{ik}(t)\| \cdot \|\mathbf{v}_{ik}(t)\|}\right). \quad (6)$$

基于上述攻击毁伤概率与规避机制,对抗集群系统智能体的攻击-损耗演化过程可表述为:

$$L_i(t+1) = L_i(t) - \sum_{k \in \mathcal{N}_i^a} a_{ik}(t) p_{ik}(t) (1 - \mu_{ik}(t)) D. \quad (7)$$

其中, a_{ik} 为攻击范围内敌对智能体攻击关系的邻接矩阵元素, D 为智能体单次攻击造成的基础伤害.

2 基于小世界网络的有限团队博弈模型

在前述智能体运动学与损耗演化模型的基础上,将小世界网络引入有限通信约束下的团队联盟结构,将集群对抗决策问题转化为可求解的团队博弈形式,构建基于小世界网络的有限团队博弈模型.

2.1 多智能体通信联盟构建规则

在多智能体对抗过程中,集群内部智能体的交互关系通常受到通信距离、带宽限制与链路动态变化的影响,实际空中无人集群平台的网络结构难以实现全局连接且始终保持稳定.小世界网络能够在有限局部连接的基础上,通过少量随机远程连边增强信息的交互能力和整体网络的连通效率,其结构特性与实际空中无人集群平台在组网层面表现出的局部密集、全局稀疏的混合通信模式高度吻合.因此,采用小世界网络建立友方智能体邻域的通信联

盟规则与信息交互机制,如图2所示.

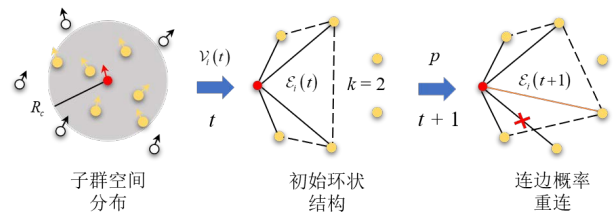


图2 多智能体对抗场景示意图

考虑通信距离对信息交互的影响,设对抗集群任意一方子群由智能体 i 与其通信半径范围内 m 个友方智能体 $j \in \mathcal{N}_i^u$ 共同构成.当且仅当友方智能体 i 与 j 的相对距离满足 $\|\mathbf{x}_i - \mathbf{x}_j\| \leq R_c$ 时,智能体间通信链路存在信息交互与同步,否则视为不可用.在此基础上,为进一步表征带宽限制导致的通信负载约束,通过引入小世界网络构造多智能体系统的集群通信拓扑网络.记时刻 t 集群的智能体节点集合为 $\mathcal{V}_i(t) = \{i, j_1, j_2, \dots, j_m\}$.在初始时刻,所有节点以环状结构排列,智能体 i 与左右两侧最近的至多 $2w$ 个邻居智能体 ($0 \leq 2w \leq m$) 建立连接并构成规则网络.由此,通过在每个通信更新时刻对智能体间的通信拓扑进行随机重连,表征高动态对抗场景链路的随机中断与重构.记任意时刻 t 下对应子群规则网络中以智能体 i 为端点的连边集合为 $\mathcal{E}_i(t)$,对于与智能体 i 连接的每条边,以重连概率 p 将其断开,并将智能体 i 随机重连至子群中除该连边端点外的其他任意友方节点.记任意时刻 t 下智能体 i 与友方协同的通信联盟图为 $\mathcal{G}_i(t) = (\mathcal{V}_i(t), \mathcal{E}_i(t))$,则对应集群系统小世界网络通信下的通信联盟图为:

$$\mathcal{G}^{sw}(t) = \bigcup_{i=1}^n \mathcal{G}_i(t) = \left(\bigcup_{i=1}^n \mathcal{V}_i(t), \bigcup_{i=1}^n \mathcal{E}_i(t) \right). \quad (8)$$

需要注意的是, $\mathcal{E}_i(t)$ 中的连边元素均为从友方智能体 $j \in \mathcal{N}_i^u$ 指向智能体 i 的有向边,对应集群系统整体的协同关系,可用邻接矩阵 $A = [a_{ij}]$ 表示.其中, $a_{ij} = 1$ 表明有向边由智能体节点 j 指向智能体节点 i ;若 $a_{ij} = a_{ji} = 1$,则表明智能体节点 i 与智能体节点 j 之间存在双向连接.由此,确定基于小世界网络的对抗集群通信联盟规则,通过在局部团簇结构的基础上引入稀疏长程连接,实现有限通信约束下集群信息异步扩散与协同决策效率的提升.

2.2 有限通信约束下团队博弈模型

在有限通信约束下的多智能体对抗场景中,每个智能体只能基于自身通信与感知能力获得信息进行决策,记智能体 i 在时刻 t 的信息集为 $H_i(t)$.在有限通信约束的信息条件下,智能体通过选取合适的

权重参数 $\alpha_i(t)$ 进行决策,进而调节协同与对抗响应项在航向更新中的相对作用强度.其中,智能体的两类响应由局部交互规则生成,进而依据运动学方程实现集群系统的状态演化.基于上述信息约束与决策表征,将小世界网络拓扑的有限通信约束引入团队博弈模型,采用不完全信息扩展型博弈来表示多智能体对抗团队博弈的动态演化过程.

设智能体 i 在时刻 t 的可观测信息包括通信邻居集合 \mathcal{N}_i^c 内的状态 $\{\mathbf{x}_j, \theta_j | j \in \mathcal{N}_i^c\}$ 与决策信息 $\{\alpha_j | j \in \mathcal{N}_i^c\}$,并据此形成对应信息集 $H_i(t)$.记 $\pi_i(\alpha_i | H_i)$ 表示在信息集 H_i 上对动作 α_i 的条件分布,则联合策略 $\Pi_T^i = \{\pi_1, \pi_2, \dots, \pi_n\}$ 为各个体策略在信息结构约束下的组合,对应通信拓扑约束下智能体依据自身信息集采取动作的决策映射可以用 $f(\cdot)$ 表示.在此基础上,团队博弈的目标是最大化整体的效用.其中,团队效用定义为对抗过程中动作状态转移的期望累计回报 $U_T^i: A_i \times \Pi_{T_i} \rightarrow \mathbb{R}$,且对抗双方的团队效用关系是一般和,这与多智能体对抗具体的任务目标有关.具体的团队效用函数将在后续博弈均衡求解一节中给出,与强化学习算法的奖励函数设计保持一致.在对抗过程中,有限通信约束导致团队智能体无法获取全局信息,每个智能体只能依赖通信网络内的邻居节点进行信息交互.因此,每个智能体均需要在通信网络拓扑结构约束下,通过优化智能体策略实现团队效用最大化,即:

$$\begin{aligned} & \max_{\Pi_{T_2}} \sum_{i \in T_2} u_i(\Pi_{T_2}) \\ & \text{s.t. } \pi_i(\alpha_i | H_i) = f\left(\sum_{j \in \mathcal{N}_i^c} \pi_j(\alpha_j | H_j)\right), \pi_i, \pi_j \in \Pi_{T_i} \\ & \sum_{i=1}^n \pi_i(\alpha_i | H_i) = 1, \pi_i(\alpha_i | H_i) \geq 0, \forall i \in \mathcal{N}. \end{aligned} \quad (9)$$

其中, $f(\cdot)$ 为决策射函数,表征了有限通信约束下团队博弈的动作状态转移机制,即团队智能体协同对抗的策略更新仅基于自身与通信邻居的信息,不依赖于全局状态或非直接通信智能体的信息.

综上,构建有限通信约束下团队博弈模型 $\Gamma_{\Pi}(A, H, \Pi, U)$,其中:

• $\mathcal{N} = T_r \cup T_b$ 为局中人集合, $T_r = \{r_1, r_2, \dots, r_n\}$ 表示红方智能体, $T_b = \{b_1, b_2, \dots, b_n\}$ 表示蓝方智能体.

• $A = \times_{\{r,b\}} A_i$ 为所有局中人的动作集合,其中 $A_i = \{\alpha_1, \alpha_2, \dots, \alpha_n\}, \alpha_i \in (0, 1)$ 表示了团队智能体在对抗过程中决策偏好选取的动作集.

• $H = \{H_{T_r}, H_{T_b}\}$ 为所有局中人的信息集合,

其中 $H_{T_i} = \cup_{i \in T_i} H_i$ 表示了各团队中所有智能体在博弈过程中有限通信约束下的局部信息联合, $H_i = \{\theta_i, \{\theta_j, \alpha_j | j \in \mathcal{N}_i^c\}\}$ 表示了团队智能体成员 i 在有限通信约束下的可用信息集, \mathcal{N}_i^c 表示了在小世界网络下成员 i 与其邻居成员 j 通信的节点集合.

• $\Pi = \{\Pi_{T_r}, \Pi_{T_b}\}$ 为所有局中人策略集合,其中 $\Pi_{T_i}^i = \{\pi_1, \pi_2, \dots, \pi_n\}$ 表示各团队的智能体策略联合, π_i 表示智能体 i 在有限通信约束下基于其可用信息集 H_i 对动作 $\alpha_i \in A_i$ 的条件概率分布,即 $\pi_i(\alpha_i | H_i)$.

• $U = \{U_{T_r}^r, U_{T_r}^b\}$ 为对应集群的团队效用集合,其中, $U_T^i: A_i \times \Pi_{T_i} \rightarrow \mathbb{R}$ 表示各团队根据每个智能体成员的动作选择而获得的团队整体效用.

2.3 关联团队最大最小均衡解概念

关联团队最大最小均衡 (Correlated Team Maxmin Equilibrium, CTME) 是完全通信条件下团队博弈模型的纳什均衡解^[33],所有团队智能体均能够共享信息策略协同决策,最大化我方团队在面对敌方团队最优策略时的最小效用.设完全通信条件下红蓝双方团队联合动作集合分别为 $A^r = \times A_{i \in T_r}$ 与 $A^b = \times A_{k \in T_b}$,对应双方联合策略 (π_r, π_b) 下红方团队期望效用为 $V^r(\pi_r, \pi_b)$,则称 $(\pi_r^*, \pi_b^*) \in \Pi_{T_r} \times \Pi_{T_b}$ 为关联团队最大最小均衡,当且仅当:

$$\begin{aligned} & V^r(\pi_r, \pi_b^*) \leq V^r(\pi_r^*, \pi_b^*) \leq V^r(\pi_r^*, \pi_b), \\ & \forall \pi_r \in \Pi_{T_r}, \forall \pi_b \in \Pi_{T_b}. \end{aligned} \quad (10)$$

对此,Zhang等^[34]指出:当团队效用函数关于双方混合策略分别线性连续且联合策略空间为非空凸紧集时,至少存在一组关联团队最大最小均衡且均衡值唯一,但实现该值的均衡策略对可能不唯一.

在实际集群对抗过程中,团队智能体的策略决策还受到通信约束、网络结构以及局部信息的共同作用影响.尤其是在有限通信环境下,团队智能体仅能基于局部通信的邻居信息做出决策,团队对抗实际上是在复杂网络上展开的.这种情况下,传统的关联团队最大最小均衡难以充分反映网络拓扑与通信约束对团队协同策略的影响.因此,需要针对本文建立的小世界网络团队博弈模型,在网络博弈的基础上提出有限通信约束下的关联团队最大最小均衡.

设初始时刻对抗集群系统中对应红蓝双方智能体集群通信拓扑的两个小世界网络相互独立,分别用 $\mathcal{G}_r^{sw} = (\mathcal{V}_r^{sw}, \mathcal{E}_r^{sw})$ 与 $\mathcal{G}_b^{sw} = (\mathcal{V}_b^{sw}, \mathcal{E}_b^{sw})$ 表示.随着多智能体对抗进程的推进,当 t 时刻存在智能体进入敌方集群智能体攻击半径时,两个小世界网络依据对抗关系建立连通路径 $\varepsilon_{r_b}^a$ 形成新的复杂网络对应

对抗集群系统整体的协同与对抗作用关系,即:

$$\mathcal{G}^{union} = (\mathcal{V}_r^{sw} \cup \mathcal{V}_b^{sw}, \mathcal{E}_r^{sw} \cup \mathcal{E}_b^{sw} \cup \mathcal{E}_{rb}^a). \quad (11)$$

设智能体 $i \in T_i$ 策略 $\pi_i(\alpha_i | H_i)$ 为对应图 \mathcal{G}^{union} 节点邻域信息集 H_j 上的可测函数,则全体团队智能体的策略空间 S_{T_i} 记为:

$$S_{T_i} = \{\Pi_{T_i} \mid \forall i \in T_i, \pi_i: H_i \rightarrow A_i\}. \quad (12)$$

若对所有团队 T_i 存在联合策略 $\Pi_{T_i}^*$ 使得在给定对手团队 T_{-i} 策略 $\Pi_{T_{-i}}^*$ 下都有:

$$\Pi_{T_i}^* = \arg \max_{\Pi_{T_i} \in S_{T_i}} U_{T_i}(\Pi_{T_i}, \Pi_{T_{-i}}^*). \quad (13)$$

则称联合策略 $(\Pi_{T_r}^*, \Pi_{T_b}^*)$ 为有限通信约束下的团队最大最小均衡.

有限通信约束下的团队最大最小均衡本质上是一种特殊结构下的纳什均衡,表明团队作为联合体在网络上进行局部最优响应决策.针对复杂网络上的博弈均衡存在性问题,Parise 等^[35]指出:只要博弈中个体成员的效用函数关于自身策略函数是强凹的且连续可微,同时每个个体成员的策略集合为凸闭有界集,则无论网络结构多么复杂,均可以保证对应的纳什均衡存在.由于本文研究中团队智能体策略为个体运动状态转移时友方协同交互作用项与敌方对抗交互作用项的比例因子 $\alpha \in (0, 1)$, 满足策略集合是凸闭有界集的条件.因此,仅需设计强凹且连续可微的智能体团队效用函数,则有限通信约束下的关联团队最大最小均衡必然存在.

3 基于强化学习的团队博弈均衡求解

在多智能体对抗过程中,团队成员仅能够依据局部可观测信息和有限通信交互,集群个体之间的通信邻域与信息结构存在异质性.传统依赖全局信息的博弈均衡求解方法难以直接适用于有限通信约束下团队最大最小均衡的求解.异构智能体强化学习方法^[36]摒弃了传统强化学习算法中参数共享的前提,能够针对具有不同观测空间、动作空间及功能角色的多类智能体进行独立策略优化,适用于复杂异构集群的协同决策任务.因此,本文以此作为团队博弈均衡求解的算法基础,设计对应有限通信约束下多智能体对抗团队均衡求解的强化学习状态空间、动作空间、奖励函数以及训练策略.

3.1 状态与动作空间定义描述

智能体的运动状态转移依赖于当前时刻通信邻域内友方智能体的平均运动方向与攻击半径范围内敌方智能体的空间分布与运动姿态.因此,对抗集群系统智能体 i 在 t 时刻的状态空间 $\mathbf{s}_i(t)$ 定义为:

$$\mathbf{s}_i(t) = [\bar{\mathbf{x}}_i(t), \bar{\theta}_i(t+1), \langle \bar{\theta}_j \rangle_i(t), \langle \bar{\phi}_k \rangle_i(t)]. \quad (14)$$

其中, $\bar{\mathbf{x}}_i$ 与 $\bar{\theta}_i$ 分别为对应智能体 i 线性归一化后的空间位置与运动方向角, $\langle \bar{\theta}_j \rangle_i$ 与 $\langle \bar{\phi}_k \rangle_i$ 分别为对应智能体 i 正弦值归一化后的友方协同交互作用角与敌方对抗交互作用角.其中智能体 i 的空间位置的线性归一化方法如下:

$$\bar{\mathbf{x}}_i(t) = 2 \cdot \frac{\mathbf{x}_i(t) - \|\mathbf{x}\|_{\min}}{\|\mathbf{x}\|_{\max} - \|\mathbf{x}\|_{\min}} - 1. \quad (15)$$

由此,智能体状态参数归一化后所有分量的取值范围均在 $[-1, 1]$ 之间.同理,智能体通过控制友方智能体协同交互作用项与敌方智能体对抗交互作用项的比例因子来调节多智能体对抗过程中进攻或撤退的协同策略.由于比例因子 $\alpha \in (0, 1)$, 故智能体 i 在 t 时刻的动作空间可直接定义为:

$$\mathbf{a}_i(t) = \alpha_i(t). \quad (16)$$

3.2 奖励函数定义描述

团队效用函数定义了对抗双方在动作状态转移过程中所获得的期望累计回报,可以依据具体对抗场景需求分解为协同、对抗、存活与任务等多个组成部分.在强化学习框架下求解对应团队博弈均衡时,奖励函数是对同一优化目标的逐时刻量化,通过将上述各项分解为即时奖励与终局奖励的线性组合,使得智能体的学习过程与团队博弈优化目标保持一致.具体的,智能体奖励 q_i 包含协同奖励、对抗奖励、存活奖励与任务奖励四个部分,即:

$$q_i = \sum_{j=1, j \in T_i} q_{ij}^u + \sum_{k=1, k \in T_{-i}} q_{ik}^v + q_h + q_f. \quad (17)$$

其中, q_{ij}^u 为与友方协同奖励, q_{ik}^v 为与敌方对抗奖励, q_h 为当前时刻存活奖励, q_f 为对抗结束任务奖励.

协同奖励用于引导集群内部友方个体的队形保持与同步,通过衡量智能体与其通信邻域内友方个体的相对距离,将其构造为对应相对距离的单调递减函数.设 q_{\max} 与 q_{\min} 分别为奖励函数设计的正值上限与负值下限,则智能体协同奖励为:

$$q_{ij}^u = \begin{cases} q_{\min} \cdot \frac{\|\mathbf{x}_{ij}\| - R_c}{R_c}, & \|\mathbf{x}_{ij}\| \leq R_c \\ q_{\max} \cdot \frac{\|\mathbf{x}_{ij}\| - R_c}{R_c}, & \|\mathbf{x}_{ij}\| > R_c \end{cases} \quad (18)$$

依据协同奖励的函数设计形式可知,当智能体位于友方通信半径内时获得正向奖励,反之则给定惩罚,进而将集群内部协同的队形保持与同步任务映射为对应友方空间态势的数值优化.

同理,通过构建智能体与敌方个体相对距离的分段函数设计对抗奖励,将持续追击与有效规避的

表1 算法伪代码

算法: 有限通信约束下的团队最大最小均衡求解算法	
Input: 红蓝双方集群初始位置、运动方向、作用范围等	
1:	初始化联合策略 $(\Pi_{T_r}, \Pi_{T_b}) \in S_{T_r} \times S_{T_b}$
2:	For 训练世代 $E = 1, 2, \dots, do$
3:	For 团队 $i = r, b$ do
4:	团队 i 策略固定
5:	For 回合时长 $t = 0, 1, \dots, do$
6:	For 团队- i 成员 $k = 1, 2, \dots, n$
7:	计算全部动作状态转移与所有对应的效用函数
8:	计算团队- i 成员 k 当前时刻最优策略 $\bar{\Pi}_{i,k}^t$
9:	策略更新 $\bar{\Pi}_{T-i}^t \leftarrow (\bar{\Pi}_{-i,1}^t, \bar{\Pi}_{-i,2}^t, \dots, \bar{\Pi}_{-i,k}^t)$
10:	End for
11:	End for
12:	团队- i 策略固定
13:	For 回合时长 $t = 0, 1, \dots, do$
14:	For 团队 i 成员 $k = 1, 2, \dots, n$
15:	计算全部动作状态转移与所有对应的效用函数
16:	计算团队 i 成员 k 当前时刻最优策略 $\bar{\Pi}_{i,k}^{t+1}$
17:	策略更新 $\bar{\Pi}_{T_i}^{t+1} \leftarrow (\bar{\Pi}_{i,1}^{t+1}, \bar{\Pi}_{i,2}^{t+1}, \dots, \bar{\Pi}_{i,k}^{t+1})$
18:	End for
19:	End for
20:	End for
21:	End for
Output: $(\Pi_{T_r}^*, \Pi_{T_b}^*)$	

战术目标转化为对敌我相对距离区间优化。当智能体与敌方个体的相对距离小于攻击半径或大于感知半径, 则给定负值惩罚, 反之给定正值奖励, 即:

$$q_{ik}^v = \begin{cases} q_{\min} \cdot \frac{R_a - \|\mathbf{x}_{ik}\|}{R_a}, & \|\mathbf{x}_{ik}\| \leq R_a \\ q_{\max} \cdot \frac{R_s - \|\mathbf{x}_{ik}\|}{R_s - R_a}, & R_a < \|\mathbf{x}_{ik}\| \leq R_s \\ q_{\min} \cdot \frac{\|\mathbf{x}_{ik}\| - R_s}{R_s}, & \|\mathbf{x}_{ik}\| > R_s \end{cases} \quad (19)$$

此外, 即时存活奖励 q_h 依据智能体存活状态给定, 若智能体存活则给定正值奖励 q_{\max} , 反之若死亡则给定单位负值奖励 q_{\min} . 当对抗结束时, 若成功拦截敌方使其未能进入任务区, 任务奖励 q_f 给定较大正值 κq_{\max} , 反之则给定较大负值 κq_{\min} , 其中 κ 为调节终局任务奖励与即时存活奖励间量级的常数。

3.3 交替冻结训练策略

由于异构智能体强化学习方法是针对单一集群的协同优化设计的, 在集群对抗场景中, 两个集群策略同时更新可能由于相互干扰导致训练不稳定. 对此, 本文采用交替冻结训练策略方法, 确定对应团队博弈均衡求解算法的伪代码如表 1 所示。

交替冻结训练策略的核心思想是: 每次更新时, 仅优化一个集群的策略, 另一个集群的策略保持冻结, 在避免集群间策略冲突的同时确保每个集群可以独立优化策略. 由于交替冻结训练本身是一种参数分组分阶段优化策略, 目前尚无严格的理论收敛证明. 然而, 类似的交替优化、交替训练、参数冻结策略在多智能体强化学习、生成对抗网络、多任务学习、协同博弈等领域得到了广泛应用^[37-39]. Wang 等^[40]指出: 交替冻结训练策略能够应对对手策略变化导致的非平稳性问题, 提升算法稳定性并加速局部最优点的收敛速度, 避免由于对抗双方策略同步演化导致的“红皇后效应”。

4 仿真实验

4.1 仿真环境设置

为验证本文所提出有限通信约束下的团队对抗建模与交替冻结强化学习框架, 仿真实验在多智能体粒子环境框架 (Multi-Agent Particle Environment, MPE) 中构建智能体并搭建 100 对 100 的红蓝双方无人集群二维空域内多智能体集群攻防对抗环境。

在仿真环境的初始设置中, 红蓝双方集群智能体规模相等且能力一致, 随机分布在 200×200 开放空间的两个独立区域依固定初始速度相向运动. 具体而言, 蓝方集群随机分布在 $U(0, 50)$ 正方形区域内, 红方集群随机分布在 $U(150, 200)$ 正方形区域内. 在对抗过程中, 红蓝双方集群依据小世界网络联盟策略构建集群相互作用与通信网络. 在此基础上, 根据对抗集群的运动学模型进行动作状态转移, 开展有限通信约束下多智能体对抗仿真实验. 红蓝双方集群智能体的参数设置, 如表 2 所示。

表2 智能体性能参数

参数	值	含义
R_c	30	智能体通信半径
R_s	20	智能体感知半径
R_a	15	智能体攻击半径
v_0	1	智能体的初始速度
$2w + 1$	7	小世界网络节点邻接数
p	0.1	小世界网络重连概率
T_{\max}	300	最大仿真时间
H	200	智能体初始生命值
D	1	智能体攻击基础伤害

在此基础上, 采用提出的基于交替冻结训练的团队博弈均衡求解方法, 训练使用的硬件配置包括 13 代 Intel i9-13900K 处理器和 NVIDIA GeForce RTX 4090 GPU, 并在该配置下完成了训练, 对应强化学习方法超参数设置如表 3 所示。

表3 超参数取值

参数	值
策略网络学习率	0.0005
价值网络学习率	0.0005
折扣因子	0.99
增益	0.01
单位正值奖励	0.05
单位负值奖励	-0.05
激活函数类型	ReLU
隐藏层的大小	[128, 128]
环境步数	120000
最大迭代步数	300
最大梯度范数	10
优化精度	1e-5

4.2 静态控制参数仿真结果分析

本节首先针对静态控制参数下的对抗集群进行仿真,验证本文提出多智能体对抗的群体动力学模型在有限通信约束下协同策略与对抗战术的涌现效果.通过选取典型的静态控制参数,揭示在不同参数配置下协同对抗行为的演化模式,为后续团队博弈均衡的动态控制参数分析提供参考基准.

对此,我们选定个体友方协同交互作用项与敌方对抗交互作用项的比例因子 α 为控制参数,系统性

地考察其在不同取值下对红蓝双方集群对抗行为的影响.图3展示了当 α 分别取0.1、0.3、0.6与0.9时红蓝双方集群对抗的关键态势快照组合.

仿真结果表明, α 的取值影响了红蓝双方集群在对抗过程中涌现的协同对抗战术与最终对抗态势.当 $\alpha = 0.1$ 时,如图3(a)所示,对抗集群双方智能体在对抗过程中受到敌方对抗交互作用项的影响高于友方协同交互作用项的影响,集群智能体对敌方智能体空间分布与运动状态的响应在协同对抗策略中占主导作用.在对抗集群相遇时,双方集群智能体自发地向一侧偏移聚集,涌现出空间错位侧翼包抄的对抗战术.随着对抗进程的推进,对抗双方智能体形成稳定的对抗界面与团簇结构并在持续牵制下整体平移.当 $\alpha = 0.3$,如图3(b)所示,对抗集群智能体的协同程度随着 α 的增大而提高.在对抗集群相遇时,对抗集群系统在空间错位侧翼包抄的对抗战术基础上,形成更快更明显的旋转趋势.随着对抗进程的推进,红方智能体分簇成两个团簇结构夹击蓝方智能体,形成夹击的稳定对抗形态并在持续牵制下整体平移.当 $\alpha = 0.6$,如图3(c)所示,集群双方智能体在对抗过程中受到友方协同交互作用项的影响高于敌方对抗交互作用项的影响,集群智能体

对友方空间分布与运动状态的响应在协同对抗

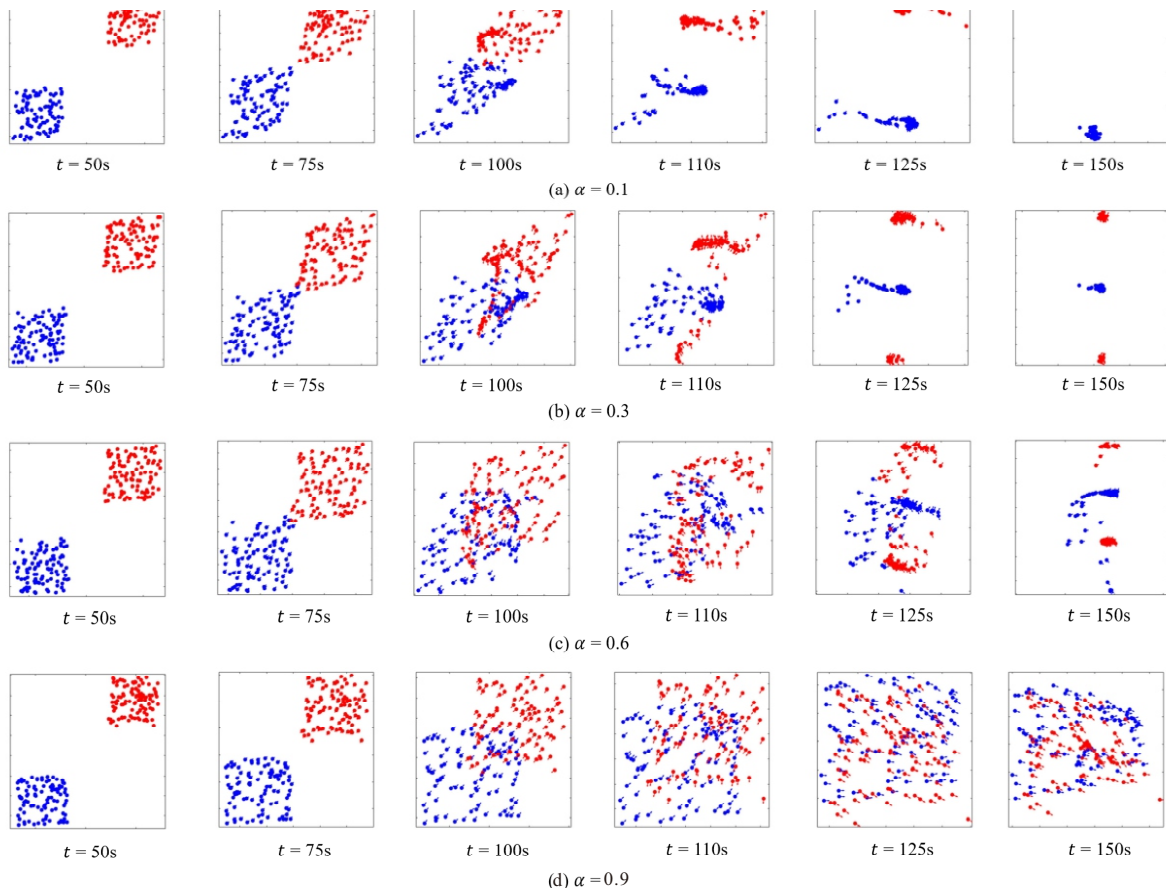


图3 静态参数集群对抗关键态势快照组合示意图

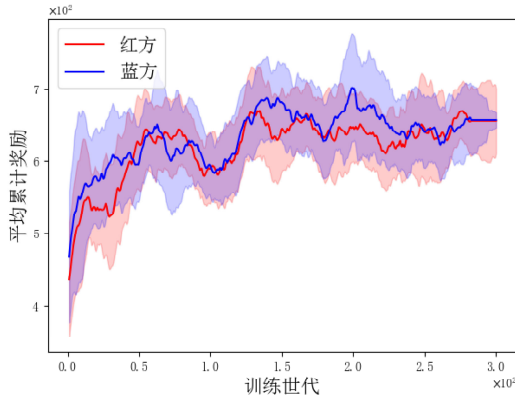


图4 有限通信场景训练奖励线示意图

策略中占主导作用. 在对抗集群相遇时, 双方集群涌现出明显的涡旋行为与空间分布. 随着对抗进程的推进, 红蓝双方智能体均分簇成两个团簇并形成相互夹击的对抗形态, 进而在持续牵制下整体平移. 当 $\alpha = 0.9$, 如图 3 (d) 所示, 对抗集群的之间的对抗程度随着 α 的增大而降低. 在对抗集群相遇时, 双方集群智能体散布在对抗范围内相互攻击并持续损耗, 没有形成明确的集群行为或对抗战术.

综上, 典型静态控制参数下的多智能体对抗结果表明, 友方智能体协同与敌方智能体对抗响应权重影响了集群对抗的行为模式与最终态势, 涌现出侧翼包抄、夹击对峙与持续牵制等协同策略与对抗战术, 验证了多智能体对抗的群体动力学模型在有限通信约束下能够涌现协同策略与对抗战术.

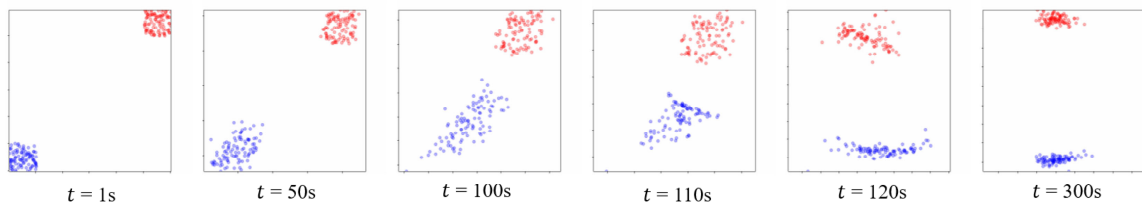
4.3 动态参数博弈均衡策略仿真结果分析

本节针对有限通信约束下的团队博弈问题, 采用本文提出的基于交替冻结训练的强化学习算法求解对应团队博弈均衡, 实现大规模多智能体对抗的动态控制参数优化与团队博弈均衡策略求解, 对应强化学习训练奖励曲线如图 4 所示.

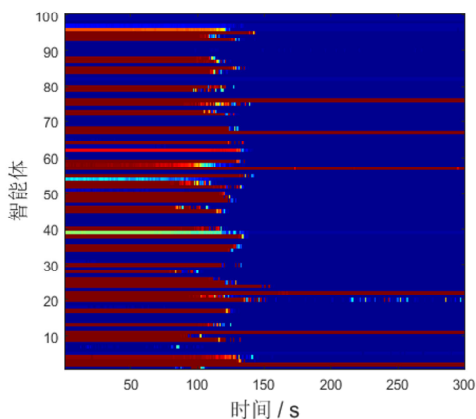
在训练初期, 红蓝双方奖励曲线存在明显交替上升现象, 这反映了交替冻结训练中, 智能体团队分别利用对手策略固定的窗口迭代训练, 策略探索呈现你追我赶的动态博弈特征. 随着训练推进, 红蓝双方回报曲线的波动幅度减小并进入稳态区间, 表明交替冻结训练能够持续逼近双方策略的最优响应, 从而获得更稳定的训练过程与策略剖面.

图 5 展示了最终训练收敛后团队博弈均衡策略对应的红蓝双方集群个体的空间分布演化结果与动态控制参数热图. 如图 5 (a) 所示, 在对抗集群相遇时, 红蓝双方集群边界处智能体开始对抗响应, 集

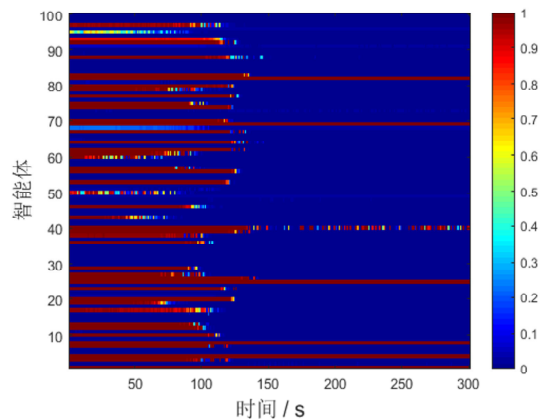
群队形由紧密团簇逐渐拉长并在中心区域形成对抗界面, 并在交界处向两侧分流扩散试图形成包围态势, 最终全部存活并形成线性对峙的对抗稳态. 分析图 5 (b) 和图 5 (c) 红蓝双方协同对抗的动态控制参数优化结果可知, 集群内部智能体之间出现明显的策略分化与阶段性变化规律. 一方面, 集群空间分布前沿的智能体在对抗初期即可保持较大的 α 值以加强对敌方的对抗响应, 在对抗相遇后, 大部分前



(a) 对抗集群行为与空间分布随时间演化示意图



(b) 红方集群智能体策略示意图



(c) 蓝方集群智能体策略示意图

图5 动态参数团队博弈均衡策略结果示意图

沿智能体 α 值减小以加强对方方的协同,而少部分前沿智能体继续保持较大的 α 值持续对敌响应;另一方面,后方集群智能体则持续保持较小的 α 值以保持集群内部的队形保持.在此基础上,依据集群对抗响应智能体的动态策略演化规律与特征,将集群对抗博弈进程划分为对抗响应阶段和态势稳定阶段.集群策略分化表征了在集群对抗过程中,集群内部智能体涌现出自适应的动态分工,由部分智能体专注于对敌响应,其他智能体进行群内协同与队形保持等任务.这种策略分化与集群对抗中智能体成员功能化、异构化分工协同的实际应用场景契合.

为进一步验证训练策略是否逼近团队博弈意义下的均衡解,通过在选定的策略检查点上固定对手策略继续训练,开展固定对手条件下的近似最优响应分析.设 $v_1(t)$ 为对应强化学习训练过程中的第 t 步检查点的团队效用评估序列,选取尾部20%的团队效用均值 $\bar{V}_{\varepsilon_{tail}}$ 作为近似最优响应的稳态团队效用估计.其中,尾段取20%作为均值统计区间是为了在保证稳态样本量的同时,降低噪声对均值估计的影响,比例的选取不改变可利用度估计的构造形式.由此,定义训练策略剖面的可利用度估计为:

$$\varepsilon(t) = d_r(t) + d_b(t) = (\bar{V}_{r,tail} - V_r(t)) + (V_b(t) - \bar{V}_{b,tail}). \quad (20)$$

其中, $\varepsilon(t)$ 为训练策略剖面的可利用度估计, $d_r(t)$ 表征红方集群在第 t 步检查点策略下团队效用可改进量的近似估计, $d_b(t)$ 表征蓝方集群在第 t 步检查点策略下团队效用可改进量的近似估计.当可利用度 ε 估计接近零时表明策略剖面接近博弈均衡.

对应博弈均衡训练过程中可利用度随策略检查点的变化曲线如图6所示.在训练初期,红蓝双方在固定对手条件下继续训练的团队效用可改进量较大,且伴随明显波动,这表明该阶段策略更新幅度

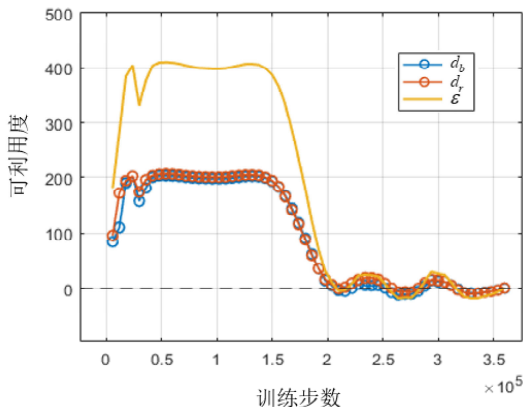


图6 可利用度曲线示意图

较大,对应团队效用的评估受采样噪声影响较强.随着训练步数的增加,红蓝双方团队效用的可改进量与最优响应训练的可利用度均快速衰减且在零附近小幅震荡.这表明随着策略迭代与学习训练的推进,对应当前训练策略剖面的单边团队效用的可改进空间被进一步压缩,双方训练策略持续逼近双方策略的最优响应.这一仿真结果验证了本文训练框架对团队博弈均衡的求解能力.

4.4 基线策略对比仿真结果分析

为进一步评估本文方法在有限通信约束下的多智能体对抗中的决策性能,本节在与前文一致的仿真环境与参数配置下,引入以下典型基线策略进行对比.记本文团队博弈均衡策略为H0;基于flocking模型的协同对抗基线策略记为H1;基于 ε -greedy贪心策略的启发式对抗基线策略记为H2;基于maximin的保守对抗基线策略记为H3.其中,H1以群体一致性协同为主要目标,通过邻域对齐、聚集与避碰等规则驱动集群行为,在对抗过程中优先维持群内队形与协同;而H2以即时对抗收益为导向,优先对最近的敌方目标做出追击或规避响应;H3则在个体可行动作集合中选取最不利情况下收益最大的响应策略.

在此基础上,分别从伤害优势、协同能力与通信负载三个角度设计指标对仿真结果分布进行统计对比.其中,伤害优势指标 J_a 定义为对抗双方累计损耗生命值 $\Delta L_i(t) = L_i(t+1) - L_i(t)$ 的差值,即 $J_a = \sum \sum \Delta L_i(t) - \sum \sum \Delta L_{-i}(t)$.若 $J_a > 0$,则表示己方在对抗过程中伤害占优;反之则敌方伤害占优.协同能力指标 J_s 则通过对抗双方终局构型的紧凑程度表示,具体定义为存活个体位置 x_i 到己方质心 \bar{x}_i 的平均距离,即 $J_s = \sum (x_i - \bar{x}_i)$.若 J_s 越小,则表示己方团队构型更紧凑,对应当前对抗策略下的相互协同支援能力更强.通信负载指标 J_c 定义为复杂网络中通用的节点平均度值,即 $J_c = \sum a_{ij} / n_i$.若 J_c 越大,则表示集群平均同时可通信邻居节点越多,对应集群个体的通信负载越大.

由此,对每种方法进行蒙特卡洛仿真实验,对应算法策略对比的箱线图统计结果如图7所示.首先从伤害优势来看,H0策略的中位数为正且上四分位水平显著高于其余基线,表明其在多数对抗中能够通过局部邻域的自组织交互形成更有效的协同压制与伤害.相较之下,H1策略与H3策略的伤害优势整体收敛于零附近,分布更为集中,表明两者更偏向稳

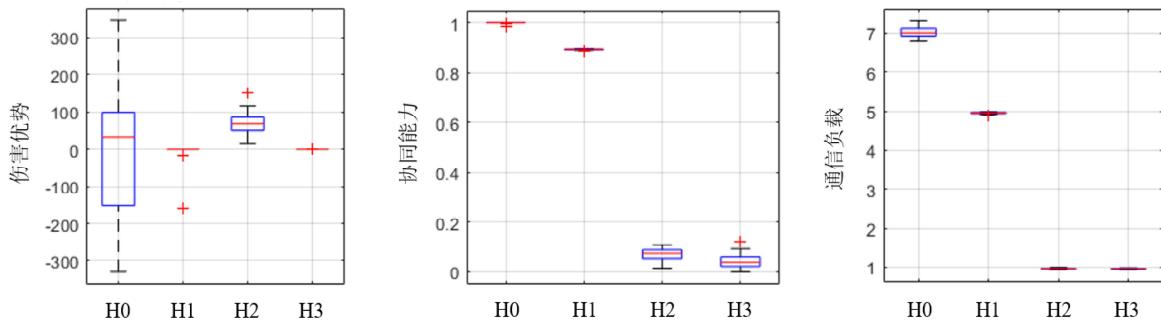


图7 不同算法策略对抗效能与通信代价箱线图统计结果示意图

健保守的策略,难以在对抗损耗上形成稳定增益.此外,H2策略的伤害优势同样呈现正向偏移且中位数高于H0,这是由于其局部贪婪响应带来更高的对抗伤害收益.其次,从协同能力来看,H0策略长期保持在接近最优水平且优于其他策略,这表明其能够在有限通信条件下稳定形成高一致性的紧凑构型.而H1策略的协同能力指标虽然同样具有较高的中位数但仍低于H0,这是由于其更倾向于群体内部一致性收敛,而非对抗导向的协同压制.相比之下,H2策略与H3策略的协同能力指标均较低,这表明两者受对抗响应的影响难以形成稳定的协同构型.最后,从通信负载来看,H0策略的节点平均度值高于其他方法,这与其协同能力一致,即团队构型越紧凑,通信半径内潜在可通信的邻居越多.需要强调的是,通信压力的上升并不等价于要求智能体全连接通信,而是体现有限通信半径条件下的潜在邻域节点规模变化.综上,本文方法在有限通信约束与相同配置场景下,相较于其他对比基线策略方法,能够通过合理的通信负载提升伤害优势与协同能力.

4.5 拓扑结构对团队博弈均衡的影响

为进一步验证有限通信约束下本文所建模型与方法的有效性,通过补充规则网络下团队博弈的强化学习训练结果,比较理想通信条件下的团队博弈均衡策略,验证拓扑结构对集群协同对抗策略的影响.在训练初期,红蓝双方的奖励曲线同样出现了明显的交替上升现象并随着训练推进收敛于接近水平.但与小世界网络下有限通信团队博弈的训练结果相比,规则网络下团队博弈强化学习训练的奖励曲线

的波动更小,收敛速度更快.对应规则网络下团队博弈强化学习训练的奖励曲线如图8所示.

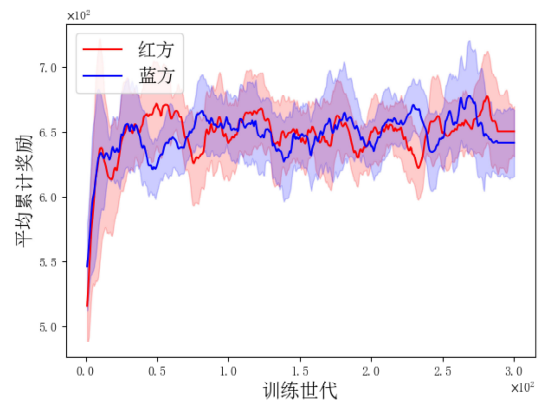
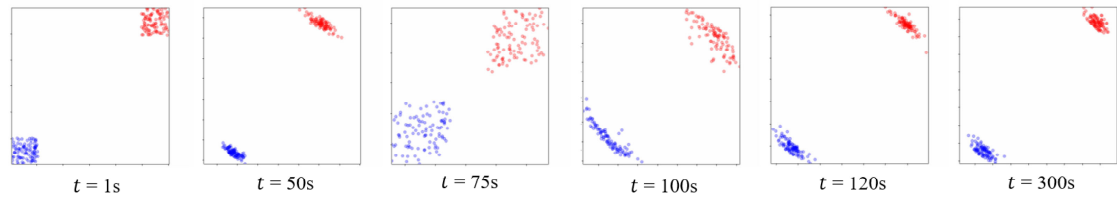
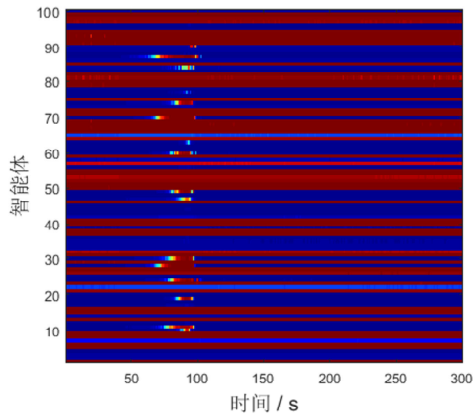


图8 理想通信场景训练奖励曲线示意图

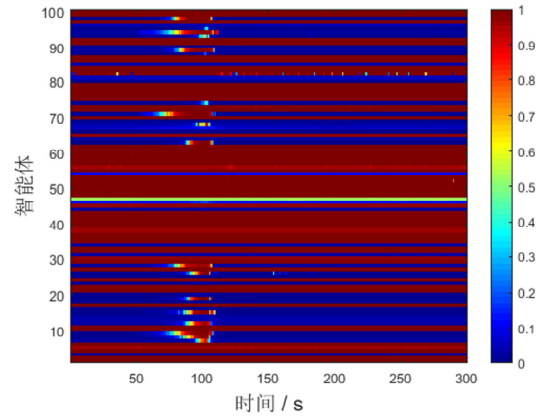
图9展示了最终训练收敛后团队博弈均衡策略对应的红蓝双方集群智能体的空间分布演化结果与动态控制参数热图.如图9(a)所示,区别于小世界网络下的团队博弈均衡策略,在规则网络下的对抗集群相遇后,集群队形在更为快速地向两侧分流扩散的同时迅速后撤并形成线性对峙.分析图9(b)和图9(c)红蓝双方协同对抗的动态控制参数优化结果可知,规则网络下的集群智能体之间也同样出现明显的策略分化与阶段性变化规律,同样可以将集群对抗博弈进程分为对抗响应阶段和态势稳定阶段.其中,对应对抗响应阶段相较于小世界网络的对抗响应阶段更为提前,且几乎全部智能体在对抗响应阶段保持较大的 α 值以加强对敌响应.这是由于规则网络表征的理想通信条件能够支持集群系统网络全局信息的交互共享与协同决策,集群对抗过程中智能体的响应速度更快,协同对抗交互过程中受随机噪声的影响更小,这也导致了强化学习训练过程中能够更快地探索到最优策略且奖励波动更小.在态势稳定阶段,双方集群在空间分布上呈现出分离对峙的稳态,且对抗双方在接触后快速形成线性构型并远离彼此并形成更为紧密的团簇结构.其中,大多数智能体在态势稳定阶段继续保持较大的 α 值以持续对敌响应,而小部分智能体减小 α 值以维系集群整体系统.这表明,规则网络下集群需要承担群内协同



(a) 对抗集群行为与空间分布随时间演化示意图



(b) 红方集群智能体策略示意图



(c) 蓝方集群智能体策略示意图

图9 规则网络团队博弈均衡策略结果示意图

与队形保持等任务所需的智能体减少,更多的智能体可以专注于对敌响应,这也侧面印证了理想通信条件下智能体的对抗响应速度更快,对应集群对抗响应阶段更短.对比仿真结果可知,集群的拓扑结构通过改变智能体的信息集,决定了集群对抗响应速度与态势稳定阶段的空间分布构型.

5 结论

本文针对有限通信约束下的多智能体集群对抗决策问题,构建了基于小世界网络的有限通信约束下团队博弈模型,设计了求解对应团队博弈均衡的交替冻结训练策略,研究了静态与动态控制参数对集群协同对抗决策与行为演化的影响,验证了网络拓扑结构对团队博弈均衡的影响.主要结论如下:

1) 友方协同与敌方对抗交互作用项的权重设置影响了集群行为和对抗态势,不同参数配置能够涌现空间错位、侧翼包抄、夹击对峙等协同对抗策略;

2) 交替冻结训练通过交替对手策略的固定窗口进行学习和调整,能够获得稳定的策略剖面实现有限通信约束下集群对抗策略的近似均衡求解;

3) 有限通信约束与理想通信条件下的团队博弈均衡在宏观结果上保持一致,涌现出策略分化的协同对抗战术,揭示了多智能体对抗响应的特征规律;

综上所述,本文提出的模型与方法能够在有限通信约束下驱动大规模集群协同对抗的自适应决策,为团队博弈优化与对抗决策机制提供了新的理论依

据和基准策略.在后续的研究中,可以进一步考虑研究其他各类网络拓扑形式,针对典型多智能体对抗场景和各类区域攻防任务,研究时延突变等更强通信约束条件下智能体协同对抗决策的应用效果,分析异构集群不同通信半径、感知半径与打击半径等能力参数对于团队博弈均衡策略的影响.

参考文献 (References)

- [1] 刘雷,刘大卫,王晓光,等.无人机集群与反无人机集群发展现状及展望[J].航空学报,2022,43(S1):4-20.
(Liu L, Liu D W, Wang X G, et al. Development status and outlook of UAV clusters and anti-UAV clusters[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(S1): 4-20.)
- [2] Vidal R, Shakernia O, Kim H J, et al. Probabilistic pursuit-evasion games: Theory, implementation, and experimental evaluation[J]. *IEEE Transactions on Robotics and Automation*, 2002, 18(5): 662-669.
- [3] von Stengel B, Koller D. Team-maxmin equilibria[J]. *Games and Economic Behavior*, 1997, 21(1/2): 309-321.
- [4] Shishika D, Paulos J, Kumar V. Cooperative team strategies for multi-player perimeter-defense games[J]. *IEEE Robotics and Automation Letters*, 2020, 5(2): 2738-2745.
- [5] 迟嵩禹,李帅,王晨,等.追逃博弈问题研究综述[J].自动化学报,2025,51(4):705-726.
(Chi S Y, Li S, Wang C, et al. A review of research on pursuit-evasion games[J]. Acta Automatica Sinica, 2025, 51(4): 705-726.)
- [6] 胡鹏林,潘泉,赵春晖.基于模糊强化学习和模型预测

- 控制的追逃博弈[J]. *控制与决策*, 2025, 40(6): 1855-1865.
(Hu P L, Pan Q, Zhao C H. Pursuit-evasion game based on fuzzy reinforcement learning and model predictive control[J]. *Control and Decision*, 2025, 40(6): 1855-1865.)
- [7] Li Z Y, Luo Y Z. Deep reinforcement learning for Nash equilibrium of differential games[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2025, 36(2): 2747-2761.
- [8] Alighanbari S, Azad N L. Deep reinforcement learning with NMPC assistance Nash switching for urban autonomous driving[J]. *IEEE Transactions on Intelligent Vehicles*, 2023, 8(3): 2604-2615.
- [9] Zuo J K, Liu Z Y, Chen J T, et al. A multi-agent cluster cooperative confrontation method based on swarm intelligence optimization[C]. 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering. Nanchang, 2021: 668-672.
- [10] Cai H, Luo Y G, Gao H L, et al. Bio-inspired intelligent swarm confrontation algorithm for a complex urban scenario[J]. *Electronics*, 2024, 13(10): 1848.
- [11] Cai H, Ma F, Ni R F, et al. Bio-inspired swarm confrontation algorithm for complex hilly terrains[J]. *Biomimetics*, 2025, 10(5): 257.
- [12] Wang B L, Li S G, Gao X Z, et al. UAV swarm confrontation using hierarchical multiagent reinforcement learning[J]. *International Journal of Aerospace Engineering*, 2021, 2021: 3360116.
- [13] Wang L H, Hu J W, Xu Z, et al. Autonomous maneuver strategy of swarm air combat based on DDPG[J]. *Autonomous Intelligent Systems*, 2021, 1(1): 15.
- [14] Wang B L, Li S G, Gao X Z, et al. Weighted mean field reinforcement learning for large-scale UAV swarm confrontation[J]. *Applied Intelligence*, 2023, 53(5): 5274-5289.
- [15] Wu Q Z, Liu K X, Chen L, et al. Hierarchical reinforcement learning for swarm confrontation with high uncertainty[J]. *IEEE Transactions on Automation Science and Engineering*, 2025, 22: 8630-8644.
- [16] Guo Z Q, Li Y K, Wang Y, et al. Group motion control for UAV swarm confrontation using distributed dynamic target assignment[J]. *Aerospace Systems*, 2023, 6(4): 689-701.
- [17] Zong J N, Gao X Z, Zhang Y, et al. Research on target allocation for hard-kill swarm anti-unmanned aerial vehicle swarm systems[J]. *Drones*, 2024, 8(11): 666.
- [18] 赵得霖, 寿莹鑫, 陈蓓, 等. 面向多目标侦察的多无人机分层任务规划方法[J]. *控制与决策*, 2025, 40(11): 3229-3238.
(Zhao D L, Shou Y X, Chen B, et al. Multi-UAVs hierarchical mission planning method for multi-target reconnaissance[J]. *Control and Decision*, 2025, 40(11): 3229-3238.)
- [19] Bu Y J, Yan Y, Yang Y N. Advancement challenges in UAV swarm formation control: A comprehensive review[J]. *Drones*, 2024, 8(7): 320.
- [20] Chen C, Bao W D, Men T, et al. A damage-tolerant task assignment algorithm for UAV swarm in confrontational environments[J]. *International Journal of Aerospace Engineering*, 2020, 2020: 8878136.
- [21] 张祥银, 张曦梁, 张天. 网络攻击下基于分布式意图识别的集群逃逸与汇聚控制[J]. *控制与决策*, 2024, 39(12): 4171-4180.
(Zhang X Y, Zhang X L, Zhang T. Swarm escape and convergence control based on distributed intent recognition under network attack[J]. *Control and Decision*, 2024, 39(12): 4171-4180.)
- [22] Shahid S, Zhen Z Y, Javaid U, et al. Offense-defense distributed decision making for swarm vs. swarm confrontation while attacking the aircraft carriers[J]. *Drones*, 2022, 6(10): 271.
- [23] Sun B, Zeng Y R, Zhu D Q. Dynamic task allocation in multi autonomous underwater vehicle confrontational games with multi-objective evaluation model and particle swarm optimization algorithm[J]. *Applied Soft Computing*, 2024, 153: 111295.
- [24] 程进, 胡寒栋, 江业帆, 等. 基于强化学习的通信受限环境多无人机协同策略[J]. *无人系统技术*, 2022, 5(5): 12-20.
(Cheng J, Hu H D, Jiang Y F, et al. Cooperative strategy of multiple unmanned aerial vehicles in limited communication environment based on reinforcement learning[J]. *Unmanned Systems Technology*, 2022, 5(5): 12-20.)
- [25] Mu J S, Zhang R H, Cui Y H, et al. UAV meets integrated sensing and communication: Challenges and future directions[J]. *IEEE Communications Magazine*, 2023, 61(5): 62-67.
- [26] Fei B W, Bao W D, Zhu X M, et al. Autonomous cooperative search model for multi-UAV with limited communication network[J]. *IEEE Internet of Things Journal*, 2022, 9(19): 19346-19361.
- [27] 刘常昱, 胡晓峰, 司光亚, 等. 基于小世界网络的舆论传播模型研究[J]. *系统仿真学报*, 2006, 18(12): 3608-3610.
(Liu C Y, Hu X F, Si G Y, et al. Public opinion propagation model based on small world networks[J]. *Journal of System Simulation*, 2006, 18(12): 3608-3610.)
- [28] Reis A S, Brugnago E L, Viana R L, et al. Effects of feedback control in small-world neuronal networks interconnected according to a human connectivity map[J]. *Neurocomputing*, 2023, 518: 321-331.
- [29] 张启亮, 武建荣. 不完全信息下基于网络演化博弈的微电网能源交易策略[J]. *控制与决策*, DOI: [10.13195/j.kzyjc.2025.0364](https://doi.org/10.13195/j.kzyjc.2025.0364).
(Zhang Q L, Wu J R. Networked Evolutionary Game-Based Energy Trading Strategy for Microgrids Considering Incomplete Information[J]. *Control and Decision*, DOI: [10.13195/j.kzyjc.2025.0364](https://doi.org/10.13195/j.kzyjc.2025.0364).)

- [30] Yan X, Huang C, Gan J Y, et al. Game theory-based energy-efficient clustering algorithm for wireless sensor networks[J]. *Sensors*, 2022, 22(2): 478.
- [31] Ginelli F. The physics of the vicsek model[J]. *The European Physical Journal Special Topics*, 2016, 225(11): 2099-2117.
- [32] Tsatsanifos T, Clark A H, Walton C, et al. Modeling large-scale adversarial swarm engagements using optimal control[C]. *The 60th IEEE Conference on Decision and Control*. Austin, 2022: 1244-1249.
- [33] Basilio N, Celli A, De Nittis G, et al. Team-maxmin equilibrium: Efficiency bounds and algorithms[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2017, 31(1): 356-362.
- [34] Zhang B H, Farina G, Sandholm T. Team belief DAG: generalizing the sequence form to team games for fast computation of correlated team max-min equilibria via regret minimization[C]. *Proc of the 40th International Conference on Machine Learning*. PMLR, 2023: 40996-41018.
- [35] Parise F, Ozdaglar A. Graphon games: A statistical framework for network games and interventions[J]. *Econometrica*, 2023, 91(1): 191-225.
- [36] Zhong Y F, Kuba J G, Feng X D, et al. Heterogeneous-agent reinforcement learning[J/OL]. 2023, arXiv: 2304.09870.
- [37] Foerster J, Farquhar G, Afouras T, et al. Counterfactual multi-agent policy gradients[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018, 32(1): 2974-2982.
- [38] Mescheder L M, Geiger A, Nowozin S. Which training methods for GANs do actually converge? [C]. *International Conference on Machine Learning*. Stockholm, 2018.
- [39] Li Q W, Zhu Z H, Tang G G. Alternating minimizations converge to second-order optimal solutions[C]. *International Conference on Machine Learning*. California, 2019.
- [40] Wang Z, Li H, Wu H L, et al. Improving maneuver strategy in air combat by alternate freeze games with a deep reinforcement learning algorithm[J]. *Mathematical Problems in Engineering*, 2020, 2020: 7180639.

作者简介

任智 (1999-), 男, 在读博士, 主要研究方向为无人集群系统对抗与博弈, E-mail: zhiren@mail.nwpu.edu.cn;

张栋 (1986-), 男, 教授, 博士, 博士生导师, 主要研究方向为无人集群智能规划与自主控制, E-mail: zhangdong@nwpu.edu.cn;

唐硕 (1963-), 男, 教授, 博士, 博士生导师, 主要研究方向为飞行器设计, E-mail: stang@nwpu.edu.cn;

熊威 (2000-), 男, 在读博士, 主要研究方向为有人无人集群协同决策, E-mail: xiong_wei@mail.nwpu.edu.cn;

杨书恒 (2001-), 男, 在读博士, 主要研究方向为无人集群决策与意图识别, E-mail: yangshuheng@mail.nwpu.edu.cn;

刘文逸 (1983-), 男, 副研究员, 在读博士, 主要研究方向为系统电气总体技术与集群智能协同控制技术, E-mail: 2309831451@qq.com.