XXXX

博弈协调增强的多智能体强化学习：UAV集群任务卸载优化

刘铭[1]，杜彦熹[2]，熊凯[2]，廖旭[3]，李志斌[3]，冷甦鹏[2]

（1. 中国电子科技集团公司第七研究所，广东省广州市 邮编510310；
2. 电子科技大学 信息与通信工程学院，四川省成都市 邮编61173；
3. 广州市弘宇科技有限公司，广东省广州市 邮编 510310）

中图分类号：

文献标志码：

doi: 10.11959/j.issn.1000−0801.

Game-Coordinated Multi-Agent Reinforcement Learning for Task Offloading Optimization in UAV Swarms

Liu Ming[1], Du Yanxi[2], Xiong Kai[2], Liao Xu[3], Li Zhibin[3], Leng Supeng[2]

1. The 7th Research Institute of China Electronics Technology Group Corporation, Guangdong Guangzhou 510310, China
2. School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu Sichuan 611731, China
3. Hong Yu Science & Technology Co., Ltd., Guangdong Guangzhou 510310, China

随着低空经济的蓬勃兴起，先进空中交通系统（AAM）正成为智能交通体系的重要组成部分。UAV集群作为低空智能网络的核心载体，通过与地面基础设施的协同，在智慧物流、应急救援等场景中发挥关键作用。然而，在边缘-云协同架构中，UAV集群面临通感算资源受限、网络拓扑动态变化与多目标冲突等核心挑战。本文提出一种融合博弈协调的多智能体强化学习框架，以实现低空场景下UAV集群的高效任务卸载与资源优化。具体而言，构建"UAV集群-边缘服务器-云数据中心"三级协同架构，将每架UAV建模为具备边缘计算能力的智能网联节点；设计"分散学习-博弈协调"交替决策机制，分散阶段各智能体基于策略网络独立学习卸载策略，协调阶段通过构建融合负载、通信、能耗与截止时间的综合支付矩阵，结合匈牙利算法实现任务-资源全局

收稿日期：XXXX−XX−XX；**修回日期**：XXXX−XX−XX

通信作者：熊凯，xiongkai@uestc.edu.cn

基金项目：四川省自然科学基金项目（批准号 2026NSFSC1431），航空工业集团飞机智能决策与协同控制机理模型联合技术中心资助，国家自然科学基金重点项目（批准号 62541103），以及国家自然科学基金青年项目（批准号 62201122）

Foundation Items: the Sichuan Provincial Natural Science Foundation under Grant 2026NSFSC1431, the AVIC United Technology Center for Intelligent Decision-making and Coordinated Control Mechanism Model Research grant, the Youth Program of the National Natural Science Foundation of China under Grant 62201122, and the Key Program of the National Natural Science Foundation of China under Grant 62541103.

最优匹配。此外，设计了融合任务完成率、负载均衡与QoS的复合奖励函数，引导系统向帕累托最优演进。仿真结果表明，所提方法在各核心指标上显著优于现有基线，为低空智能网络部署与AAM协同调度提供了有效技术支撑。

低空智能；UAV集群；多智能体强化学习；博弈论；边缘-云协同；任务卸载；资源优化

With the rise of the low-altitude economy, Unmanned Aerial Vehicle (UAV) swarms are becoming essential carriers of Advanced Air Mobility (AAM) systems. However, UAV swarms in air-ground integrated edge computing architectures face challenges including limited resources, dynamic topology, and multi-objective conflicts. This paper proposes a game-coordinated multi-agent reinforcement learning framework for efficient task offloading in low-altitude scenarios. We design a three-tier "UAV swarm – edge server – cloud" collaborative architecture and a "decentralized learning – game coordination" alternating mechanism. During the coordination phase, a comprehensive payoff matrix incorporating load, communication, energy, and deadline costs is constructed, and the Hungarian algorithm is employed for globally optimal task-resource matching. A composite reward function balancing task completion rate, load balancing, and QoS guides the system toward Pareto optimality. Simulation results demonstrate that the proposed method significantly outperforms baselines in task completion rate, load balancing, and energy efficiency, providing effective support for AAM system deployment.

Low-Altitude Intelligence; UAV Swarm; Multi-Agent Reinforcement Learning (MARL); Game Theory; Edge-Cloud Collaboration; Task Offloading; Resource Optimization

1 引言

随着人工智能、物联网、大数据以及5G/6G通信技术的迅速发展，全球交通运输系统正经历深刻的结构性变革。智能交通系统（Intelligent Transportation Systems, ITS）作为未来交通发展的核心方向，依托高性能感知、互联通信与算力协同，推动交通运行向安全化、智能化与可持续化的持续演进[1]。在此背景下，低空经济正成为战略性新兴产业的重要组成部分，先进空中交通系统（Advanced Air Mobility, AAM）作为其核心应用，正在拓展未来空中出行的新模式[2]。无人机（Unmanned Aerial Vehicle, UAV）集群作为低空智能网络的关键载体，因其独特的三维机动性与灵活部署能力，在智慧物流、城市空中交通（Urban Air Mobility, UAM）、应急救援与智能交通监测等领域展现出广阔的应用前景[3]。在这些应用场景中，UAV集群通常需要实时处理大量计算密集型任务，包括图像视频分析、多传感器数据融合、目标检测识别、实时航迹规划等。然而，受限于UAV自身的载荷约束与能耗限制，单架UAV难以独立完成复杂计算任务，亟需借助外部计算资源实现任务卸载与协同处理。

智能网联技术（如V2X、IoV、C-V2X）作为ITS的重要支撑，构建了"车-路-云"一体化的智慧交通生态。在低空交通领域，这一理念自然延伸为"机-地-云"协同架构——UAV集群通过与地面通信基站、边缘服务器的实时交互，形成空地一体化的智能网联系统。边缘计算作为一种新型计算范式，通过将计算资源下沉至网络边缘，能够有效降低数据传输延迟、减轻核心网络负载，成为解决UAV集群计算瓶颈的重要技术途径。在这种边缘-云协同架构下，UAV可将计算密集型任务（如实时目标检测、多传感器数据融合、航迹规划等）卸载至近端边缘节点或远程云数据中

心进行处理，从而突破自身资源限制，提升任务处理效率与系统响应速度[4][5]。然而，UAV 集群作为低空智能交通系统中的移动边缘计算节点，面临着与地面车联网系统既相似又独特的挑战，这些挑战使得 UAV 集群在边缘-云环境下的任务卸载与资源优化成为一个极具理论价值与工程意义的研究课题。

在 UAV 集群边缘-云协同任务卸载系统中，多目标冲突与动态优化问题是制约系统整体性能提升的关键瓶颈[6]-[9]。具体而言，该问题可从系统动态性、多目标协同与多智能体协同三个维度进行深入分析。

从**系统动态性维度**来看，UAV 集群的移动特性引发了一系列连锁反应，这与车联网中车辆高速行驶时的通信切换问题相类似但更为复杂。首先，UAV 的动态移动特性导致网络拓扑与通信条件实时变化——UAV 的位置变化导致其与地面基站、地面终端及其他 UAV 之间的通信链路质量发生动态波动，无线信道条件的时变性进一步加剧了卸载决策的不确定性。传统的静态资源分配策略难以适应这种高度动态的环境。其次，UAV 执行任务过程中产生的负载分布呈现时空异质性——某些 UAV 可能因承担计算密集型任务而处于高负载状态，而另一些 UAV 可能仅承担感知任务而处于相对空闲状态。UAV 集群在执行任务过程中会产生差异化的资源需求与负载分布，集群内部易出现资源竞争与负载不均衡现象。这种负载动态变化要求资源分配策略具备实时感知与自适应调整能力，传统的静态优化方法难以应对如此复杂多变的系统环境。

从**多目标协同维度**来看，UAV 集群任务卸载涉及多个相互制约的性能指标，这些指标之间存在复杂的耦合关系与潜在冲突。具体而言：任务完成率反映了系统对用户请求的响应能力，高任务完成率意味着更多任务能够在截止时间内完成；负载均衡度关乎集群资源利用的公平性与效率，过度集中可能导致部分 UAV 过载而另一些 UAV 闲置；能耗效率直接影响 UAV 的续航能力与任务执行的可持续性，这对于低空交通系统尤为关键——UAV 集群的能耗管理直接关系到续航时间与任务执行连续性，资源优化决策必须在任务性能与能耗效率之间寻求平衡；服务质量（Quality of Service, QoS）则综合考量了延迟、可靠性等用户感知指标。

从**多智能体协同维度**来看，UAV 集群可视为由多个自主决策智能体组成的分布式系统，这与地面交通中的拥堵博弈现象具有高度的理论同构性。当多架 UAV 同时竞争有限的边缘计算资源时，个体理性决策的叠加可能导致集体非理性的"囚徒困境"——每架 UAV 都倾向于将任务卸载至资源最丰富的节点，但这种局部最优策略的集体执行将造成局部热点与资源浪费，反而降低了系统整体性能。此外，不同 UAV 可能具有差异化的任务特征、能量状态与服务能力，这种异质性进一步增加了协同决策的复杂度。如何设计有效的协调机制，引导各 UAV 在追求个体利益的同时实现系统全局最优，是解决多智能体协同困境的关键。

针对上述挑战，学术界已开展了大量研究工作。文献[10]为解决无人机辅助空地集成网络中局部拥堵和资源分配不均的挑战，提出了一个两阶段优化框架，降低了延迟和能耗；文献[11]提出了基于非正交多址（NOMA）的多无人机协作 MEC 网络模型，目标是最小化整体系统延迟；文献[12]提出了一种基于微分演化和贪婪算法的算法，旨在最小化能耗；文献[13]提出了一种联合优化策略，最大化无人机飞行轨迹和用户设备任务卸载率的能源效率。但上述研究多聚焦于单一或部分性能指标的优化，未能充分考虑 UAV 集群作为低空智能网联节点时的动态性、多智能体协同复杂性与多目标冲突特性，且忽视了任务卸载过程中的负载均衡与全局资源优化等关键性能的

协同实现。在军事侦察、灾害救援、环境监测等实际应用场景中，负载均衡直接关系到UAV集群的续航能力与任务执行连续性，可避免部分UAV因过载提前退出任务、部分UAV闲置的资源浪费问题[14]；而全局资源优化则能突破个体理性导致的“囚徒困境”，实现系统层面任务处理效率与资源利用效率的双赢[15]。

在算法层面，单智能体强化学习方法通过智能体与环境的持续交互学习最优卸载策略，在静态或准静态环境中展现出一定优势。然而，这类方法将其他UAV的决策视为环境噪声，忽视了多智能体系统的策略互动本质。当系统规模扩大或环境动态性增强时，单智能体方法难以捕捉复杂的状态-动作空间分布，容易陷入局部最优解[16]。而多智能体强化学习（Multi-Agent Reinforcement Learning, MARL）通过"分散执行、集中训练"或"完全分散"等框架为多智能体协同决策提供了新的技术路径。然而，现有MARL方法在处理UAV集群任务卸载问题时仍面临诸多挑战：全局奖励的稀疏性与延迟性问题导致智能体难以建立行为与长期收益之间的有效映射；信用分配问题使得智能体难以准确评估自身贡献与系统整体性能之间的关系；部分可观测与通信约束进一步加剧了策略学习的难度。更关键的是，当智能体数量增加时，状态空间与动作空间的维度爆炸问题使得MARL方法的计算复杂度急剧上升，难以满足AAM系统对实时决策与全局协调的双重需求[17][18]。此外，现有研究在UAV集群作为移动边缘节点的场景下存在明显的理论空白。多数研究将UAV仅视为任务卸载的请求方或中继节点，忽视了UAV集群作为分布式计算资源的潜在价值。实际上，当UAV集群协同执行任务时，集群内部同样存在计算任务的再分配与资源协调需求，这一问题尚未得到充分研究。

综上所述，现有研究虽然在UAV任务卸载领域取得了丰富成果，但在应对低空智能交通场景的复杂性与动态性方面仍存在共性问题：一方面，多数研究聚焦于单一或部分性能指标的优化，缺乏对任务完成率、负载均衡与能耗效率等多目标冲突的协同处理机制；另一方面，将UAV集群视为静态或孤立个体的简化假设，难以适应网络拓扑动态变化与多智能体策略交互带来的环境非平稳性。针对上述核心挑战与现有研究的局限性，本文提出一种融合博弈协调机制的多智能体强化学习方法，旨在实现低空智能交通场景下UAV集群的高效任务卸载与资源优化。核心思路是将UAV集群构建为空地协同的移动边缘计算资源池，通过博弈论方法协调集群内部的任务分配，结合强化学习实现动态环境下的自适应决策。本文的主要贡献可归纳为以下三个方面：

第一，构建面向低空智能交通的UAV集群移动边缘计算模型。该模型将UAV集群纳入空地一体化的边缘计算资源体系，提出"UAV集群-边缘节点-云数据中心"的三级协同架构，系统刻画UAV的移动特性、能耗约束与V2X通信动态性，为智能网联场景下的任务卸载决策提供统一的形式化描述。

第二，设计基于博弈论的多层协调机制。针对UAV集群内部的任务分配问题，构建非合作博弈与最优匹配相结合的双层协调模型：第一层博弈建模各UAV的任务卸载决策，通过支付矩阵量化不同卸载选择的效用；第二层采用匈牙利算法求解任务与UAV资源的最优匹配，实现纳什均衡稳定状态下的全局资源优化。

第三，提出融合分散学习与集中协调的多智能体强化学习框架。在分散执行阶段，各UAV作为智能网联节点基于策略网络独立做出卸载决策；在协调阶段，博弈协调器根据支付矩阵计算最优匹配结果，对分散决策进行修正与优化。同时，设计了融合任务完成率、负载均衡度与QoS满意度的复合奖励函数，引导智能体在个体利益与系统全局效益之间达成平衡，为AAM系统的

高效协同调度提供技术支撑。

2　边缘-云融合架构模型

本研究设计了边缘-云融合任务卸载系统，采用"云中心-无人机集群-地面终端"三级分布式架构。该架构充分考虑了低空交通场景中UAV的三维机动性与空地协同特性，通过多智能体决策与博弈协调实现资源高效分配。无人机集群作为移动边缘计算资源池，为覆盖范围内的地面终端设备提供计算卸载服务，突破传统固定边缘节点的覆盖限制与资源瓶颈，整体架构如图1所示。

在融合架构中，云中心提供全局算力支撑与资源管理，无人机集群作为移动边缘层实现空域灵活部署与近端服务，用户设备（User Equipment，UE）作为任务生成端可按需卸载。三层通过层级协作实现任务合理分配：终端可卸载至覆盖无人机，无人机集群内部通过博弈协调进行任务再分配，资源不足时任务可进一步上云处理。

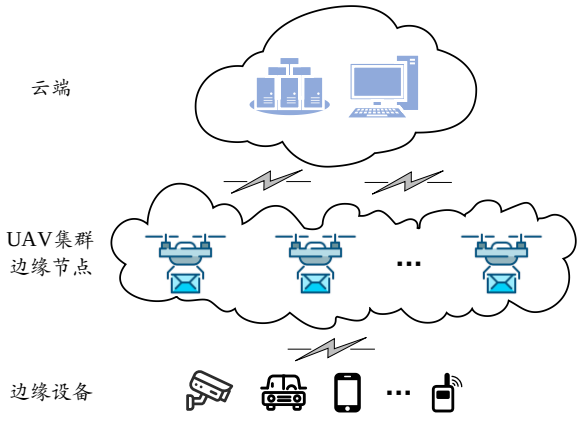


图1　边缘-云架构

2.1　节点模型

在低空智能交通场景中，节点模型需充分反映AAM系统的运行特征。与传统地面车联网不同，低空交通网络具有以下特点：（1）UAV节点具备三维空间机动能力，可根据任务需求动态调整服务区域；（2）空地通信链路受大气传播特性影响，信道条件随飞行高度变化；（3）UAV能耗包含悬停、飞行与计算三种模式，续航约束对任务卸载决策具有重要影响。本文将所有计算与通信实体抽象为统一的节点模型，不同类型节点具有差异化的物理属性与资源能力。为便于后续数学建模与算法设计，对三类节点的特性进行定义。

（1）云中心节点位于整个架构的顶层，作为低空交通管理系统的核心算力支撑，具有预设的计算容量、通信带宽与能耗系数。在AAM系统中，云中心承担全局态势感知、空域资源调度与跨域协同等关键功能。

（2）考虑由N_u架无人机组成的集群系统，第i架无人机的位置随时间动态变化，位置向量表示为$\mathrm{p}_i(t)=\left(x_i(t),y_i(t),z_i(t)\right)$，其中$t$表示时间步。无人机的移动采用恒定速度直线飞行简化分析：$\mathrm{p}_i(t)=\mathrm{p}_i(0)+\mathrm{v}_i\cdot t$，其中$\mathrm{v}_i=\left(v_{ix},v_{iy},v_{iz}\right)$为恒定速度向量，$v_i=/\!/\,\mathrm{v}_i\,/\!/$为无人机的巡航速度。

每架无人机配置一定规模的机载计算资源，CPU计算容量设定为$C_{u,i}$，与云中心及地面终端的上行/下行传输速率分别设定为$B_{u,i}^{\mathrm{up}}$和$B_{u,i}^{\mathrm{down}}$，无人机的能耗综合考虑悬停能耗、飞行能耗与计算能耗，总能耗$E_{u,i}(t)$可表示为：

$$E_{u,i}(t)=E_{u,i}^{\mathrm{hover}}+E_{u,i}^{\mathrm{flight}}+E_{u,i}^{\mathrm{compute}} \quad (1)$$

其中，悬停能耗$E_{u,i}^{\mathrm{hover}}$与飞行能耗$E_{u,i}^{\mathrm{flight}}$由无人机的动力学特性决定，计算能耗与任务处理量成正比。

无人机服务覆盖范围是一个以无人机当前位置为圆心、半径为$R_{u,i}$的圆形区域。覆盖半径取决于无人机的通信发射功率与传播环境。只有位于覆盖范围内的地面终端才能与该无人机建立通信链路并发起任务卸载请求。

（3）UE设备节点位于架构的底层，是计算任务的主要生成端。在低空交通场景中，UE设备不仅包括传统的智能手机、平板等终端，还涵盖地面物联网传感器、智能交通设施以及需要实

时数据处理的AAM地面支撑设备。在本文模型中，UE设备不具备自主计算卸载决策能力，其任务请求由所绑定的无人机边缘节点代为处理。

考虑由N_e个用户设备组成的终端群体，第j个用户设备的位置固定或缓慢移动，位置向量表示为$\mathrm{p}_j^e=\left(x_j^e,y_j^e,0\right)$。UE设备的CPU计算能力设定为$C_e$，通信带宽设定为$B_e$。UE设备不参与任务处理，仅负责任务生成与结果接收。

节点的实时负载状态对于任务卸载决策至关重要。本文采用CPU利用率作为负载状态的核心度量指标。设t时刻节点k已分配的任务集合为$\mathrm{T}_{k(t)}$，任务$i\in\mathrm{T}_{k(t)}$的CPU计算需求为C_i，节点k的CPU总容量为C_k，则t时刻节点k的CPU利用率$\rho_k(t)$定义为：

$$\rho_k(t)=\frac{\sum_{i\in T_k(t)}c_i}{C_k} \tag{2}$$

其中$\rho_k(t)\in[0,1]$，$\rho_k(t)=0$表示节点空闲，$\rho_k(t)=1$表示节点满负荷运行。当$\rho_k(t)$大于负载阈值，该节点被视为过载节点，不宜继续接收新的计算任务。

对于无人机边缘节点，还需定义电量状态。设t时刻无人机i的剩余电量为$B_{u,i}(t)$，则其归一化电量状态$\phi_i(t)$定义为：

$$\phi_i(t)=\frac{B_{u,i}(t)}{B_{u,i}^{max}} \tag{3}$$

其中$\phi_i(t)\in[0,1]$，$\phi_i(t)<0.2$时无人机进入低电量模式，需优先考虑任务卸载以避免电量耗尽导致的任务中断，$B_{u,i}^{\max}$为UAV的电池容量。

2.2 任务模型

系统采用离散时间步长进行仿真，仿真时间$t=1,2,3,\ldots,T_{max}$。每个时间步内，UE以一定概率生成新任务，任务生成过程建模为泊松过程。在时间步t，UE生成新任务的概率p_e设定为（0.1~0.3），单个时间步内系统总任务数N_t服从参数为λ_t的泊松分布，其中$\lambda_t=N_e\cdot p_e\cdot\left(1+\sigma\cdot\sin\left(2\pi t/24\right)\right)$，$\sigma$=0.3引入任务负载的周期性波动以模拟实际应用场景中的峰谷变化。

任务生成后立即进入待处理队列等待卸载决策。队列管理采用先进先服务（First-Come-First-Served，FCFS）策略，但高优先级任务可插队处理。

在融合架构中，任务有二种卸载路径：

路径一为UAV边缘处理路径。任务卸载至覆盖该UE的无人机边缘节点进行处理。卸载至无人机的执行时间包括传输时间与处理时间两部分：传输时间$t_i^{\mathrm{tx}}=d_i/B_{e,u}$，其中$B_{e,u}$为UE与UAV之间的通信带宽；处理时间$t_i^{\mathrm{proc}}=c_i/C_{u,i}$，其中$C_{u,i}$为目标UAV的计算容量。

路径二为云中心处理路径。任务经UAV中继或直接卸载至云中心进行处理。适用于计算需求极高或集群资源紧张的大型任务，但传输延迟较大。总延迟包括UE到UAV的传输时间、UAV到云中心的传输时间以及云中心的处理时间。

3 多智能体决策模型

将UAV集群任务卸载问题建模为分散式部分可观测马尔可夫决策过程（Decentralized Partially Observable Markov Decision Process，Dec-POMDP），每架UAV作为独立智能体与环境持续交互，通过策略优化实现长期累积奖励最大化[19]。

3.1 决策问题定义

由N架UAV组成的集群系统$\mathcal{U}=\{u_1,u_2,\ldots,u_N\}$执行协同任务，系统运行环境包含$M$个用户设备$\mathcal{E}=\{e_1,e_2,\ldots,e_M\}$与一个云中心节点$C$。每个时间步$t$，环境状态包含所有节点的位置信息、资源状态与任务队列状态。由于通信范围限制与隐私保护约束，单架UAV无法获取全局完整信息，仅能观测到局部状态$\mathrm{o}_i^t\in O$，形成部分可观测的决策环境。

每架无人机u_i基于局部观测o_i^t独立选择动作

$a_i^t \in A_i$（服务区域约束，表示只有覆盖该区域内的UAV才能处理该任务），动作空间定义了所有可能的卸载决策选项。智能体的策略 π_i: $O \to \Delta(A_i)$ 将观测映射到动作的概率分布，其中 $\Delta(A_i)$ 表示动作空间上的概率单纯形。策略优化的目标是最大化长期累积奖励，即：

$$\max_{\pi} \mathrm{E}_{\tau\sim\pi}\left[\sum_{t=0}^{T} \gamma^t R(\mathrm{s}_t, \mathrm{a}_t)\right] \quad （4）$$

其中 $\gamma\in[0,1)$ 为折扣因子，$R(\mathrm{s}_t, \mathrm{a}_t)$ 为 t 时刻的即时奖励，τ 表示智能体与环境交互产生的轨迹。

3.2　状态空间设计

状态向量融合了UAV集群状态、UE状态、任务队列状态与环境状态四类信息，以求在信息完整性与计算复杂度之间取得平衡。

UAV集群状态：UAV集群状态反映了集群内部各UAV的实时运行状况，状态包含以下维度：

位置状态维度 $s_i^{\mathrm{pos}} = \left(x_i(t), y_i(t), z_i(t)\right)$ 表示UAV的三维空间坐标，其中 $x_i(t), y_i(t)$ 为水平位置坐标，$z_i(t)$ 为飞行高度。位置信息决定了无人机的服务覆盖范围与通信链路质量，是卸载决策的关键参考因素。

资源状态维度包含CPU利用率 $\rho_i(t)$ 与电量状态 $\phi_i(t)$ 两个指标。CPU利用率 $\rho_i(t) = \sum_{j\in T_i(t)} c_j/C_i$ 反映UAV的当前负载水平，其中 $T_i(t)$ 为 t 时刻分配给 $\mathcal{U}_i$ 的任务集合，c_j 为任务 j 的计算需求，C_i 为 $\mathcal{U}_i$ 的CPU总容量，$\phi_i(t)$ 反映无人机的剩余电量比例。

负载预测维度 $s_i^{\mathrm{load\text{-}pred}} = \left(\hat{\rho}_i(t+1), \hat{\rho}_i(t+2)\right)$ 给出未来两个时间步的负载预测值，基于历史负载曲线与任务到达率进行估计。负载预测为UAV提供前瞻性信息，有助于避免将任务卸载至即将过载的节点。对于包含 N 架UAV集群，UAV集群状态的总体维度为 $3N+3N=6N$，其中 $3N$ 为位置状态维度，$3N$ 为资源与预测状态维度。

UE状态：UE状态反映地面终端的任务生成情况与服务需求。用户设备 e_j 的状态包含以下维度：

任务队列维度 $q_j(t)$ 表示 t 时刻 e_j 待处理的任务队列长度，反映了设备的任务生成速率与处理能力的匹配程度。队列长度越大，表示设备越迫切需要将任务卸载至UAV进行处理。

覆盖状态维度 $c_{i,j}(t) \in \{0,1\}$ 表示 e_j 是否处于 $\mathcal{U}_i$ 的服务覆盖范围内，覆盖状态决定了卸载选择的可行域。对于包含 M 个UE的群体，UE状态的总体维度为 $2M$。

任务队列状态：任务队列状态描述系统中待处理任务的分布情况，为卸载决策提供全局视野。每架 $\mathcal{U}_i$ 的任务队列状态包含以下信息：

待处理任务数 $n_i^{\mathrm{pending}}(t)$ 表示当前分配给 $\mathcal{U}_i$ 但尚未开始处理的任务数量，反映了任务积压程度。

任务特征统计维度 $s_i^{\mathrm{taskstat}} = \left(\bar{c}_i(t), \bar{T}_i^{max}(t), \bar{\tau}_i(t)\right)$ 汇总了待处理任务的特征参数，其中 $\bar{c}_i(t)$ 为平均计算需求，$\bar{T}_i^{max}(t)$ 为平均截止时间上限，$\bar{\tau}_i(t)$ 为平均截止时间紧迫度。

任务优先级分布维度 $s_i^{\mathrm{prio\_dist}}(t) = \left(n_i^1(t), n_i^2(t), n_i^3(t), n_i^4(t), n_i^5(t)\right)$ 为各优先级等级的任务数量分布。对于整个UAV集群，任务队列状态的总体维度为 $N+3N+5N=9N$。

环境状态：环境状态描述了影响系统运行的外部因素，包括时间因素与干扰因素：

时间因素维度 t_{global} 表示全局仿真时间，用于捕捉任务的周期性到达模式与UAV电量的时间累积效应。

干扰状态维度 $I(t)\in[0,1]$ 表示 t 时刻的通信干扰水平，数值越大表示信道条件越差，通信成本越高。环境状态的总体维度为2。

环完整状态向量：

综合以上四类状态，智能体观测到的完整状态向量定义为：

$$o_i(t)=\left\{\begin{array}{l}\underbrace{p_1(t),\ldots,p_N(t)}_{\text{位置状态}}\mid\ \underbrace{\rho_1(t),\ldots,\rho_N(t)}_{\text{CPU利用率}}\mid\ \underbrace{\phi_1(t),\ldots,\phi_N(t)}_{\text{电量状态}}\mid \\ \underbrace{q_1(t),\ldots,q_M(t)}_{\text{任务队列}}\mid\ \underbrace{n_1^{\text{pending}}(t),\ldots,n_N^{\text{pending}}(t)}_{\text{待处理任务}}\end{array}\right\}\sum_{k=1}^{4}\omega_k=1。$$

（5）

状态向量的总维度为D=6N+2M+9N+2=15N+2M+2。以N=10架UAV、M=20个UE为例，状态维度为15×10+2×20+2=182，在深度强化学习的处理能力范围内。

3.3 动作空间设计

动作空间定义了智能体在给定状态下可采取的所有决策选项，当UE将计算任务卸载至UAV后，UAV需要决策如何处理该任务，确定任务是在本地执行还是转发至云中心处理。动作选择需满足以下物理约束：

覆盖约束要求目标节点必须能够与源节点建立通信链路。对于UE到UAV的卸载，UAV需满足$\mathbin{/\!/}p_i(t)-p_k(t)\mathbin{/\!/}\le R_{\text{comm}}$，其中$R_{\text{comm}}$为通信范围上限。

容量约束要求目标节点的资源分配不超过其剩余容量。设UAV的剩余CPU容量为$C_k^{\text{rem}}(t)$，任务所需CPU资源为c_j，则资源分配需满足$a_i^{\text{resource}}(t)\cdot C_k\le C_k^{\text{rem}}(t)$。

电量约束要求低电量UAV优先卸载任务。当$\phi_k(t)<0.2$时，UAV不再接收新的卸载任务，以避免电量耗尽导致的任务中断。

3.4 奖励机制设计

由于UAV集群任务卸载问题的多目标特性，设计融合多维度指标的复合奖励函数，平衡任务完成率、负载均衡度、服务质量与能耗效率四个优化目标。

t时刻智能体的奖励$R_i(t)$由四项子奖励加权合成：

$$R_i(t)=\omega_1R_i^{\text{completion}}(t)+\omega_2R_i^{\text{balance}}(t)+\omega_3R_i^{\text{qos}}(t)+\omega_4R_i^{\text{energy}}(t)\quad（6）$$

其中$\omega_1,\omega_2,\omega_3,\omega_4$为归一化权重系数，满足$\sum_{k=1}^{4}\omega_k=1$。

第一项任务完成奖励$R_i^{\text{completion}}(t)$衡量任务分配与处理的成功程度：

$$R_i^{\text{completion}}(t)=\begin{cases}+10 & \text{if task successfully allocated}\\ -5 & \text{if task allocation failed}\\ 0 & \text{otherwise}\end{cases}\quad（7）$$

任务成功分配的判定标准为：任务在截止时间内获得目标节点的处理承诺，且目标节点具有足够的资源承载该任务。

第二项负载均衡奖励$R_i^{\text{balance}}(t)$鼓励任务在集群内的均匀分布，避免局部过载：

$$R_i^{\text{balance}}(t)=\frac{\sigma(t)}{\bar{\rho}(t)}\quad（8）$$

其中$\sigma(t)$为t时刻集群各无人机CPU利用率的标准差，$\bar{\rho}(t)$为均值。负载均衡奖励在$\sigma(t)$接近0时取得最大值，表明UAV负载完全均衡；当$\sigma(t)$增大时，奖励值降低，偏离程度越大惩罚越重。

第三项服务质量奖励$R_i^{\text{qos}}(t)$考量任务处理的时效性与可靠性：

$$R_i^{\text{qos}}(t)=1-\frac{T_i^{\text{actual}}-T_i^{\min}}{T_i^{\max}-T_i^{\min}}\quad（9）$$

其中T_i^{actual}为任务的实际处理延迟，$T_i^{\min}$为理论最小延迟，$T_i^{\max}$为截止时间上限。当任务在截止时间内完成时，$R_i^{\text{qos}}(t)\in[0,1]$，延迟越接近理论最小值奖励越高；当任务超时失效时，$R_i^{\text{qos}}(t)=-1$。

第四项能耗效率奖励$R_i^{\text{energy}}(t)$引导智能体在完成任务的同时控制能耗：

$$R_i^{\text{energy}}(t)=\frac{E_i^{\text{task}}}{E_i^{\text{total}}}\quad（10）$$

其中E_i^{task}为处理分配任务消耗的能耗，E_i^{total}为

UAV 在 t 时刻的总能耗。

4　博弈协调机制设计

MARL 的分散学习使各 UAV 能够在部分可观测环境中独立学习卸载策略，但多个智能体基于局部观测做出决策时，容易同时选择同一目标节点，造成资源冲突与负载失衡。博弈协调通过构建全局支付矩阵并求解纳什均衡，为分散决策提供冲突消解与全局匹配能力。另一方面，博弈论的均衡分析依赖系统参数的准确建模，而 UAV 集群拓扑与任务负载持续变化，静态博弈模型难以跟踪实时状态。MARL 的在线学习为支付矩阵提供动态更新的状态估计，避免了对固定参数的依赖。基于上述考虑，本文将两者结合为"分散学习—博弈协调"交替机制：MARL 负责环境适应与策略探索，博弈协调负责决策冲突的全局化解。

在多 UAV 集群边缘-云任务卸载系统中，各无人机作为独立决策智能体，为覆盖范围内的用户设备提供计算卸载。当多个 UE 同时向无人机集群发起卸载请求时，UAV 的接收策略叠加可能导致系统陷入次优状态。个体理性决策的集体执行可能引发“囚徒困境”—每架无人机倾向于接收尽可能多的任务以获取奖励，但这种局部最优策略的叠加将导致部分无人机过载而其他无人机闲置，反而降低系统整体性能[20][21]。

4.1　双层博弈模型定义

本文采用双层博弈模型，第一层为 UE-UAV 卸载博弈，描述地面终端与 UAV 边缘节点之间的任务分配关系；第二层为集群内部资源协调博弈，描述 UAV 之间的任务再分配与负载均衡关系。

在第一层博弈中，地面终端 e_j 作为任务请求方选择目标 UAV，UAV 作为资源提供方接受或拒绝任务请求。对于 UE 产生的任务，策略空间为 $N(j)=\left\{u_i\in\mathcal{U}\,\middle|\,/\!/\,\mathrm{p}_j^e-\mathrm{p}_i(t)\,/\!/\le R_{u,i}\right\}$，即 e_j 可选择将任务卸载至服务覆盖范围内的任意无人机。对于 u_i，其策略空间为接收或拒绝来自各用户设备的卸载请求，即 $A_i=\{\mathrm{accept},\mathrm{reject}\}$。博弈的效用函数综合考量任务的处理延迟、资源消耗与服务覆盖，对于 e_j 选择 u_i 执行任务的效用定义为：

$$U_j\left(a_j=u_i,\mathrm{a}_{-j}\right)=-C_{j,i}^{\mathrm{total}} \tag{11}$$

其中 $C_{j,i}^{\mathrm{total}}$ 为任务卸载至 u_i 的综合成本，纳什均衡条件要求在均衡状态下，u_i 不愿单方面改变其接收或拒绝策略，e_j 也不愿单方面改变其目标无人机选择，即：

$$\begin{aligned}U_i\left(\mathrm{a}_i^*,\mathrm{a}_{-i}^*\right)&\ge U_i\left(\mathrm{a}_i,\mathrm{a}_{-i}^*\right),\forall\mathrm{a}_i\in A_i\\U_j\left(\mathrm{a}_j^*,\mathrm{a}_{-j}^*\right)&\ge U_j\left(\mathrm{a}_j,\mathrm{a}_{-j}^*\right),\forall\mathrm{a}_j\in N(j)\end{aligned} \tag{12}$$

第二层博弈建模为集群内部的任务再分配与资源协调。当 UAV 接收到来自 UE 的卸载请求后，可能因自身负载过高或电量不足而选择将任务转发给集群内其他无人机。对于无人机接收的任务集合，策略空间为任务分配方案 $X_i=(x_{i1},x_{i2},\ldots,x_{iN})$，其中 x_{iN} 表示分配给 UAV 的任务比例，无人机的策略空间也可表示为离散选择：对于每个待分配任务，选择目标无人机作为处理节点。博弈的效用函数基于集群整体性能定义：

$$\begin{aligned}U_i(x_i,x_{-i})=&\alpha\cdot TR_i(x_i)+\beta\cdot LB_i(\mathrm{x}_i)\\&-\gamma\cdot EC_i(x_i)-\delta\cdot DL_i(x_i)\end{aligned} \tag{13}$$

其中 TR_i 为任务完成率贡献，LB_i 为负载均衡贡献，EC_i 为能耗成本，DL_i 为截止时间延迟成本，$\alpha,\beta,\gamma,\delta$ 为权重系数。第二层博弈的纳什均衡要求在均衡状态下，UAV 不愿单方面改变其任务分配策略。

4.2　支付矩阵构建方法

支付矩阵能量化各参与者在不同策略组合下的收益或成本，本文构建的支付矩阵综合考量负载成本、通信成本、能耗成本与截止时间紧迫性成本四项指标，通过加权融合形成综合效用，为博弈决策提供量化依据。

（1）负载成本量化目标节点的资源占用程度

对任务执行效率的影响。当目标节点的CPU利用率较高时，新任务的处理延迟将显著增加，甚至可能导致任务排队超时。设任务T_j卸载至无人机u_i，u_i当前CPU利用率为$\rho_i(t)$，任务的计算需求为c_j，则负载成本定义为：

$$C_{j,i}^{\text{load}} = w_1 \cdot \rho_i(t) \cdot \frac{c_j}{C_i} \tag{14}$$

其中C_i为UAV的CPU总容量，w_1为负载成本权重系数。

为体现负载的动态演化特性，本文还引入负载预测因子$\hat{\rho}_i(t+1)$，将预测负载纳入当前负载成本的计算：

$$C_{j,i}^{\text{load}} = w_1 \cdot \left[\lambda \cdot \rho_i(t) + (1-\lambda) \cdot \hat{\rho}_i(t+1)\right] \cdot \frac{c_j}{C_i} \tag{15}$$

其中λ为当前负载与预测负载的权重系数。

（2）通信成本包括传输时间与带宽占用两个方面，通信成本与通信距离正相关，与可用带宽负相关。设任务源自主机k，目标节点为u_i，通信距离为$d_{k,i}(t) = /\!/ \, \mathrm{p}_k(t) - \mathrm{p}_i(t) \, /\!/$，任务数据量为$d_j$，源节点到目标节点的可用带宽为$B_{k,i}(t)$，则通信成本定义为：

$$C_{j,i}^{\text{comm}} = w_2 \cdot \frac{d_{k,i}(t)}{R_{\max}} \cdot \frac{d_j \cdot 8}{B_{k,i}(t)} \tag{16}$$

其中$R_{\max}$为最大通信范围，w_2为通信成本权重系数。

对于集群内部的通信，还需考虑UAV间通信链路的稳定性。设t时刻无人机u_k与u_i之间的链路质量因子为$L_{k,i}(t) \in [0,1]$，则通信成本修正为：

$$C_{j,i}^{\text{comm}} = w_2 \cdot \frac{d_{k,i}(t)}{R_{\max}} \cdot \frac{d_j \cdot 8}{B_{k,i}(t) \cdot L_{k,i}(t)} \tag{17}$$

其中链路质量因子可根据历史通信成功率或实时信噪比测量值进行估计。

（3）能耗成本量化任务处理过程中的能量消耗，能耗成本与目标节点的能耗系数及任务处理时间正相关。设目标节点的能耗系数为η_i，目标节点分配给任务的CPU资源比例为α_j，则能耗成本定义为：

$$C_{j,i}^{\text{energy}} = w_3 \cdot \eta_i \cdot c_j \cdot \alpha_j \tag{18}$$

对于UE节点，能耗成本还需考虑电池电量约束。设UE的剩余电量为$E_k(t)$，电池容量为E_k^{max}，则当$E_k(t)/E_k^{max} < 0.2$时，通信能耗成本需乘以惩罚因子$w_3^{\text{penalty}} = 2.0$：

$$C_{j,i}^{\text{energy}} = \begin{cases} w_3 \cdot \eta_i \cdot c_j \cdot \alpha_j & \text{if } E_k(t)/E_k^{\max} \geq 0.2 \\ 2w_3 \cdot \eta_i \cdot c_j \cdot \alpha_j & \text{if } E_k(t)/E_k^{\max} < 0.2 \end{cases} \tag{19}$$

对于UAV节点，能耗成本还需考虑无人机的飞行状态。无人机在悬停状态、执行任务状态与移动状态下的能耗系数不同，设悬停能耗系数为η_i^{hover}，任务处理能耗系数为η_i^{task}，移动能耗系数为η_i^{move}，则能耗成本修正为：

$$C_{j,i}^{\text{energy}} = w_3 \cdot \eta_i^{\text{state}} \cdot c_j \cdot \alpha_j \tag{20}$$

其中$\text{state} \in \{\text{hover}, \text{task}, \text{move}\}$表示无人机当前状态。

（4）截止时间紧迫性成本量化任务的时间约束对决策的影响。任务剩余时间越紧，决策的时效性要求越高。设任务的截止时间为$T_j^{\max}$，任务已等待时间为$t - t_j^{\text{arrival}}$，则剩余时间可表示为$T_j^{\text{rem}} = T_j^{\max} - \left(t - t_j^{\text{arrival}}\right)$。截止时间紧迫性成本定义为：

$$\begin{aligned} C_{j,i}^{\text{deadline}} &= w_4 \cdot \frac{T_j^{\max} - T_j^{\text{rem}}}{T_j^{\max}} \\ &= w_4 \cdot \frac{t - t_j^{\text{arrival}}}{T_j^{\max}} \end{aligned} \tag{21}$$

其中w_4为截止时间紧迫性成本权重系数，取值范围为[0,1]。

将以上四项成本加权融合，得到综合支付函数：

$$\begin{aligned} P_{j,i} = -(&\alpha \cdot C_{j,i}^{\text{comm}} + \beta \cdot C_{j,i}^{\text{load}} \\ &+ \gamma \cdot C_{j,i}^{\text{energy}} + \delta \cdot C_{j,i}^{\text{deadline}}) \end{aligned} \tag{22}$$

其中$\alpha, \beta, \gamma, \delta$为各成本项的权重系数，综合支付函数取负值形式，使得成本越高的选择对应越低的

支付。

4.3　纳什均衡求解及证明

对于用户设备-无人机卸载博弈，纳什均衡的求解算法流程如下。首先，初始化所有UE的策略为随机选择，然后，对于每一轮迭代：

步骤一，更新UAV的资源可用状态。更新UAV的CPU利用率、电量状态与任务队列长度。

步骤二，计算UE的最佳响应。对于每个UE，遍历其可行目标UAV集合，选择具有最高支付值的节点作为最佳响应策略。

步骤三，更新策略并检查收敛。如果所有UE的策略在本次迭代中均未改变，或迭代次数达到上限，则算法收敛，输出当前策略作为纳什均衡；否则，返回步骤二继续迭代。

对于UAV集群内部资源协调博弈，纳什均衡的求解采用类似的迭代最佳响应算法，但策略空间为离散的任务分配选择。算法流程如下。

首先，初始化UAV的任务分配策略。对于每个无人机接收的任务集合，初始化将所有任务分配给自身处理。然后，对于每一轮迭代：

步骤一，更新资源状态。根据当前任务分配方案，更新UAV的CPU利用率与电量消耗率。

步骤二，计算最佳响应。对于每个UAV，对于其待分配任务集合中的每个任务，遍历可行的目标UAV集合（包括自身与其他无人机），选择具有最高综合效用的目标节点。

步骤三，更新任务分配方案并检查收敛。如果所有任务的目标节点在本次迭代中均未改变，或迭代次数达到上限，则算法收敛；否则，返回步骤二继续迭代。

博弈论的基本定理保证：若策略空间为非空紧凸集，且效用函数为连续拟凹函数，则纳什均衡一定存在（Debreu-Glicksberg定理）。本文设计的博弈满足这些条件：(1) 策略空间性质：第一层博弈的策略空间为离散有限集，第二层博弈的资源分配比例空间为闭区间[0,1]的子集，均为非空紧集；对于连续策略空间，其凸性显然满足。(2) 效用函数性质：本文构建的效用函数为各项成本函数（负载、通信、能耗等）的线性加权和。由于各项成本函数关于资源分配量连续，且权重系数均为正数，故效用函数保持连续性；同时，成本函数通常被建模为资源消耗的凸函数（如负载成本随利用率呈指数增长），这保证了效用函数关于策略变量的拟凹性。因此，根据Debreu-Glicksberg定理，该双层博弈的纳什均衡一定存在。

关于纳什均衡的唯一性，本文进一步通过效用函数的严格拟凹性证明其唯一。本文构建的效用函数为各项成本/收益函数的线性加权和，其中负载成本、通信成本、能耗成本和截止时间紧迫性成本均为资源分配量的凸函数（如负载成本随利用率呈指数增长，通信成本随传输量呈线性增长）。由于权重系数均为正数，效用函数关于资源分配策略的Hessian矩阵为负定，从而保证效用函数严格拟凹。根据Debreu的纳什均衡唯一性定理，当效用函数严格拟凹且策略空间为凸集时，纳什均衡唯一。综上，该双层博弈的纳什均衡不仅存在，且唯一。

4.4　匈牙利算法最优匹配

在博弈协调机制中，纳什均衡求解仅能保证策略的局部最优，无法保证全局最优[22][23]。为实现任务与节点的最优匹配，本文引入匈牙利算法求解二分图的最优匹配问题，在多项式时间复杂度内找到使总支付最大化的任务-节点分配方案。

将任务-节点匹配问题建模为二分图匹配问题。设左侧顶点集合为待处理任务 $T=\{T_1,T_2,\ldots,T_J\}$，右侧顶点集合为可用计算节点 $\mathcal{N}=\{N_1,N_2,\ldots,N_K\}$，边 $\left(T_j,N_k\right)$ 的权重为任务 T_j 分配给节点 N_k 的综合支付值 $P_{j,k}$。目标是在满足节点容量约束的前提下，选择一组匹配边使总支付最大化：

$$\begin{aligned} &\text{P1: } \max \sum_{j=1}^{J} \sum_{k=1}^{K} P_{j,k} x_{j,k} \\ &\text{s.t. C1: } \sum_{k=1}^{K} x_{j,k} = 1, \forall j = 1,2,\ldots,J \\ &\text{C2: } \sum_{j=1}^{J} c_j x_{j,k} \le C_k, \forall k = 1,2,\ldots,K \\ &\text{C3: } x_{j,k} \in \{0,1\}, \forall j,k \end{aligned} \quad (23)$$

其中，$x_{j,k}$为二元决策变量，C_k为节点的计算容量。

算法流程如下。首先，构建支付矩阵P，若$J<K$则添加$K-J$行虚拟任务，若$J>K$则添加$J-K$列虚拟节点；然后初始化可行标记$l(v)=\max_k P_{v,k}$与$l(u)=0$；接着构建相等子图，包含所有满足$l(v)+l(u)=P_{v,u}$的边(v,u)。在相等子图中寻找增广路径，若找到则更新匹配；若未找到，则调整可行标记：

$$\begin{aligned} l(v) &\leftarrow l(v) - \theta, \ \forall v \in S \backslash T \\ l(u) &\leftarrow l(u) + \theta, \ \forall u \in T \end{aligned} \quad (24)$$

其中S为增广路径搜索中访问的左侧顶点集合，T为访问的右侧顶点集合，$\theta = min\{l(v)+l(u)-P_{v,u} | v \in S \backslash T, u \notin T\}$为最小调整量。重复以上步骤直至找到完美匹配。

5 算法实现

图2展示了本文所提基于博弈协调的多智能体强化学习任务卸载框架的完整流程。主要分为三个阶段：在分散学习阶段，各UAV基于局部观测通过策略网络独立生成初步卸载决策，形成可能存在冲突的初步任务分配矩阵；在博弈协调阶段，协调器收集全局信息，根据负载、通信、能耗和截止时间成本构建支付矩阵，并使用匈牙利算法求解全局最优匹配，得到无冲突的最优任务-UAV分配方案；在执行反馈阶段，UAV按最终方案执行卸载决策，环境计算复合奖励并更新状态，用于更新各UAV的策略网络。

本文设计的模型结构包含策略网络（Actor）和价值网络（Critic）两个核心组件。策略网络为

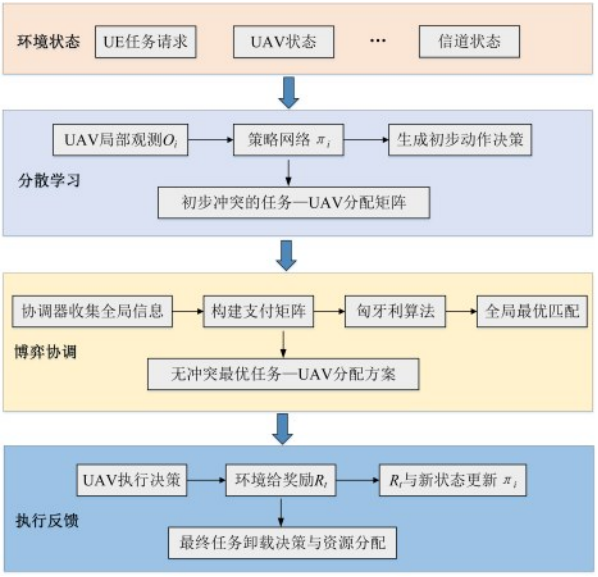


图2 基于博弈协调的强化学习任务卸载系统框图

三层全连接神经网络，输入为系统状态（节点CPU利用率、位置信息等），依次经过两个128神经元的隐藏层（ReLU激活），最后通过Softmax输出动作概率分布。价值网络同样为三层全连接结构，输入维度与策略网络相同，经过相同的隐藏层结构，输出单一标量值作为状态价值评估。两个网络均采用Adam优化器（学习率0.001）进行参数更新。

具体算法可以通过算法1来描述。每步迭代中，各UAV首先基于局部观测通过策略网络生成初步动作；协调器随后构建支付矩阵并运行匈牙利算法，得到全局最优的最终动作（博弈协调）；UAV执行最终动作后获得复合奖励，并存储经验用于更新策略网络。

算法1：博弈协调增强的多智能体强化学习算法

输入：UAV数量N，环境Env，协调器$Coord$，最大回合数n_{ep}，每回合步数T输出：训练好的策略网络参数$\{\theta_i\}_{i=1}^{N}$

1. 初始化每个UAV的策略网络$\pi_i(\theta_i)$和经验回放池D

2. for episode =1 ton_{ep}do

3. 获取初始观测 $\{o_i\}_{i=1}^{N}$

4. for step =1 to T do

5. // 分散学习：各UAV并行生成初步动作

6. for each UAV i in parallel do

7. $a_i^{\text{pre}} \sim \pi(\cdot|o_i)$// 根据策略采样

8. end for

9. // 博弈协调：全局最优匹配

10. 协调器构建支付矩阵 P（公式14-22），运行匈牙利算法，得到最终动作 a_i

11. // 执行与反馈

12. 执行 a_i，获得奖励 R_t 和下一观测 o_i'

13. 存储 (o_i, a_i, R_t, o_i') 到 D

14. // 网络更新

15. if $|D| \geqslant$ 批量大小 then

16. 从 D 采样小批量，更新 θ_i

17. end if

18. $'o_i \leftarrow o_i'$

19. end for

20. end for

21. 返回 $\{\theta_i\}_{i=1}^{N}$

算法1中，分散学习阶段（步骤5-8）各UAV策略网络输出初步动作，随后在博弈协调阶段（步骤9-10）由协调器构建支付矩阵并求解最优匹配，对初步动作中存在的冲突进行修正。修正后的动作经环境执行产生奖励信号，再用于策略网络的参数更新。由于奖励信号同时反映了博弈协调的修正效果，策略网络在训练中会逐步倾向于生成与全局最优匹配一致的初步动作。

6　仿真实验分析

6.1　仿真参数设置

本文构建的仿真场景模拟典型的UAV集群应急通信与数据处理应用。在一个1000×1000平方米的方形区域，地面部署有20个UE，均匀分布于区域外围。空中部署有10架无人机组成的移动边缘计算集群，UAV在高度80米处呈环形巡航飞行，巡航速度为10米/秒。云中心位于区域中心位置。

仿真环境涉及通信模型、计算模型、任务模型和算法四类核心参数。通信模型参数描述无线信道的物理特性，计算模型参数描述各节点的算力与能耗特性，任务模型参数描述任务的生成规律与属性分布，算法核心参数描述仿真算法运行过程的参数取值。四类参数的具体配置如表1至表4所示。

仿真过程中信道模型采用对数距离路径损耗模型，UE与UAV之间的路径损耗表示为：

$$PL(d) = PL(d_0) + 10\alpha \log_{10}\left(\frac{d}{d_0}\right) + X_\sigma \quad （25）$$

其中 d 为通信距离，X_σ 为服从高斯分布的阴影衰落因子，标准差 σ=6 dB。传输速率根据香农公式计算：

$$R = B\log_2\left(1 + \frac{P_t \cdot G_r \cdot G_t}{N_0 \cdot B \cdot 10^{\frac{PL(d)}{10}}}\right) \quad （26）$$

其中 P_t=23 dB 为发射功率，G_r=2 dBi 为接收天线增益，G_t=2 dBi 为发射天线增益。

表1　通信模型参数

参数名称	符号	取值	单位
载波频率	f_c	2.4	GHz
信道带宽	B	20	MHz
噪声功率谱密度	N_0	-174	dBm/Hz
路径损耗指数	α	2	-
参考距离	d_0	1	m
参考路径损耗	$PL(d_0)$	40	dB
最大通信距离	$R_{\max}$	150	m

7　仿真结果分析

所有实验均在配置为NVIDIA RTX A5000 GPU的计算机上运行,编程语言是Python，版本为3.10。选用的对比算法包括双延迟深度确定性策略梯度算法（Twin Delayed Deep Deterministic Policy Gradient, TD3）、异步优势演员-评论家算

表2　计算模型参数

节点类型	CP容量	通信带宽	能耗系数	部署特征
云中心	1000 GFLOPS	1000 Mbps	0.01 J/GFLOP	固定坐标 (0,0,H_c)，高度10~20 m
无人机边缘	10~30 GFLOPS	50~100 Mbps	0.05 J/GFLOP	动态位置，速度5~15 m/s
用户设备	1~5 GFLOPS	50 Mbps	0.1 J/GFLOP	固定位置，分布于80~120 m环形区域

表3　任务模型参数

参数名称	符号	取值范围	分布类型
任务计算量	c_i	5~50 GFLOPS	均匀分布
任务数据量	d_i	1~10 MB	正比分布
任务截止时间	$T_i^{\max}$	100~500ms	均匀分布
任务优先级	prio_i	1~5	泊松分布
任务到达率	λ	0.5~3任务/步	对数正态分布

法（Asynchronous Advantage Actor-Critic, A3C）和随机算法。TD3是一种基于Actor-Critic的深度强化学习算法，通过引入双Q网络和延迟更新机制有效缓解值函数过估计问题，并采用目标策略平滑正则化提升算法稳定性；A3C则采用异步并行训练框架，多个Actor-learner线程同时与环境交互，通过优势估计降低梯度方差，具有收敛速度快、训练稳定的优点。为适配多智能体UAV集群场景，本文采用独立学习（Independent Learning, IL）框架实现TD3与A3C的多智能体扩展：将其他UAV的决策行为纳入环境动态，每架UAV视为独立智能体，维护各自的策略网络与价值网络，在共享环境中并行训练。随机算法在每个时间步随机选择卸载目标，作为性能下界参考。为保证对比公平性，所有算法的超参数均经网格搜索调优，状态空间、动作空间与奖励函数设置与本文方法保持一致。

图3为UAV集群辅助计算场景中各算法的平均累积奖励收敛对比。受奖励函数中惩罚机制的影响，纵坐标的平均累积奖励呈现出包含负值的分布特征，反映了系统在决策过程中由于任务处理超时违约、集群负载不均衡以及过高的能耗成本等因素触发而受到的惩罚。观察曲线可知，本文所提方法收敛最快、最终性能最优，1000回合

表4　算法核心参数表

参数类别	参数名称	取值
MARL基础参数	状态维度	182
	动作维度	26
	隐藏层维度	256
	学习率	0.001
	折扣因子	0.95
	探索率	0.9→0.01
	批次大小	64
博弈协调参数	支付矩阵权重系数	α=0.2（负载）、β=0.2（通信）、γ=0.3（能耗）、δ=0.3（截止时间）
	负载阈值	0.3
	匈牙利算法迭代次数	100
奖励函数参数	任务完成权重	0.4
	负载均衡权重	0.3
	QoS权重	0.2
	能效权重	0.1

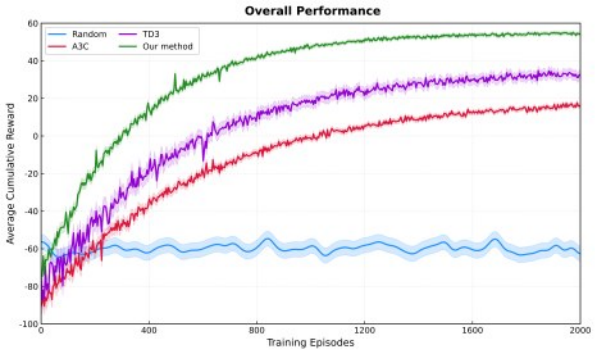


图3　本模型、TD3、A3C 和随机算法表现对比

后趋于稳定；TD3 上升速度与奖励次之且前期波动大；A3C 收敛慢且波动显著；Random 始终徘徊在低奖励区间。结果表明，所提方法在边缘-云任务卸载及资源优化中，具备更优的收敛效率、性能上限与稳定性，显著优于对比算法。从图 3 的收敛过程还可观察到在训练前期，策略网络尚未收敛，博弈协调在此阶段承担了主要的冲突消解与性能保障作用；随着训练推进，策略网络通过奖励反馈逐步学习到与全局最优一致的决策模式，博弈协调对动作的修正幅度随之减小。这说明随着训练的进行，策略网络能够生成接近全局最优的动作，博弈协调的作用从初期的大幅修正转变为微调补充。

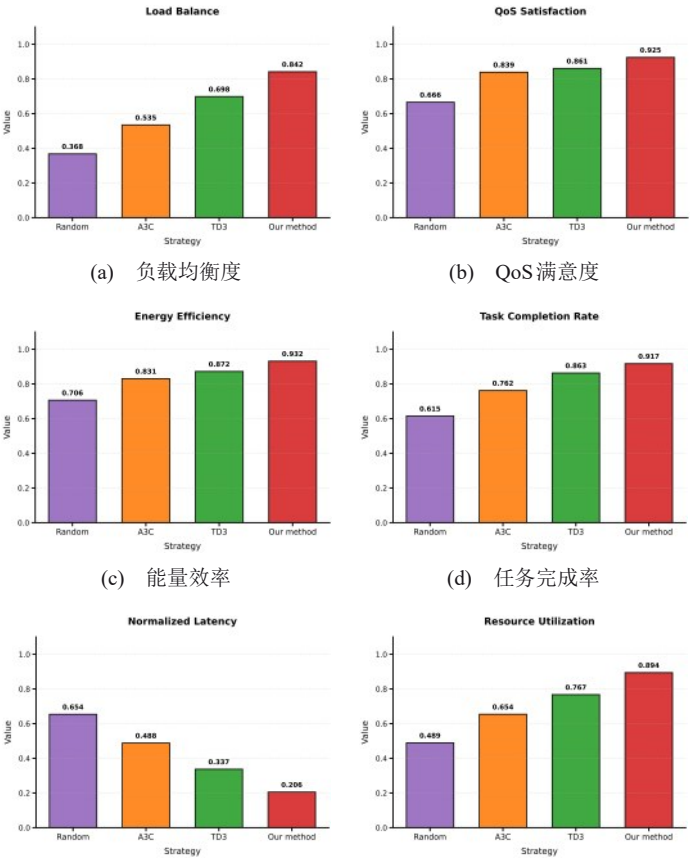
图 4 为多指标性能对比柱状图，从负载均衡、QoS 满意度、能效、任务完成率、归一化时延及资源利用率六个核心维度的对比结果可见，本文方法在所有评估指标上均展现出最优性能：负载均衡值达 0.87、QoS 满意度为 0.93、能效高达 0.94、任务完成率提升至 0.93，归一化时延仅 0.21，资源利用率也达到 0.90 的最高水平；TD3 策略各项指标均优于 A3C 和 Random；A3C 策略相较于 Random 虽有明显提升，但在时延控制、负载均衡等维度仍存在差距；Random 则在所有维度表现最差，显著低于其他策略。整体而言，本文方法具备显著性能优势，适用于对时延、能效和任务完成率有高要求的任务卸载场景。

图 5 为任务-UAV 节点分配支付矩阵热力图，数值分布契合博弈协调机制与实际场景逻辑：正支付对应高适配性任务-节点组合，负支付标识为高成本约束型任务，负支付是对低效分配的量化惩罚。实验结果表明，算法基于“总支付最大化”目标成功筛选出全局最优任务-节点匹配方案，有效提升了系统整体收益与资源配置效率。

图 6 为 UAV 效用值分布直方图。结果显示，效用值集中于 0.6 至 1.0 区间的样本占比达 82%，表明博弈协调机制通过支付矩阵多目标加权设计与匈牙利算法全局最优匹配，实现个体效用与全局效益的平衡，规避 UAV 收益极端化情况，保障了 UAV 集群系统协同稳定。

8　结束语

面向低空经济与先进空中交通系统（AAM）的快速发展需求，本文针对 UAV 集群作为低空智能网联节点在边缘-云协同环境中面临的多目标冲突与动态优化难题，提出了一种融合博弈协调机制的多智能体强化学习框架。所构建的"UAV 集群-边缘服务器-云数据中心"三级协同架构，将 UAV 集群纳入空地一体化的智能交通资源体系；所设计的"分散学习-博弈协调"交替决策机制，通过支付矩阵量化 V2X 通信成本、负载均衡与能耗约束间的多目标权衡，结合匈牙利算法实现了任务卸载的全局最优匹配。仿真结果表明，所提方法在任务完成率、负载均衡与能效等核心指标上均显著优于现有基线算法，验证了博弈论与多智能体强化学习融合方法在低空智能交通场景中的有效性。未来工作将进一步拓展至空地协同的多模态交通场景，探索 UAV 集群与地面智能网联车辆（CAV）的跨域协同调度问题，并考虑复杂气象条件与通信干扰对系统性能的影响，推动低空交通系统与地面交通体系的深度融合，助力智能交通系统的立体化发展。



(a) 负载均衡度 (b) QoS满意度

(c) 能量效率 (d) 任务完成率

(e) 响应时间 (f) 资源利用率

图4 多指标性能对比柱状图

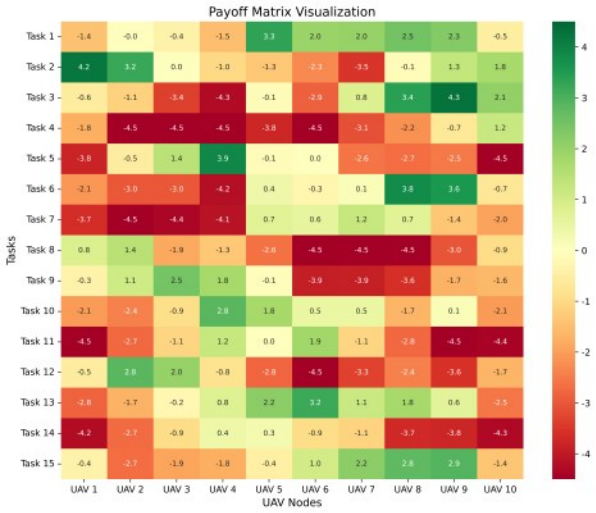


图5 支付矩阵热力图

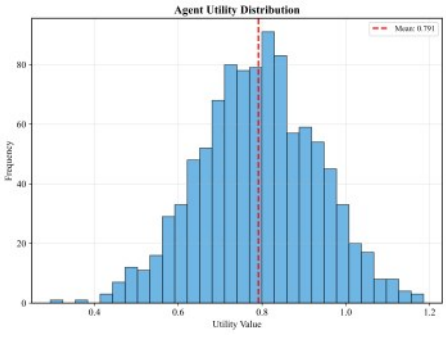


图6 UAV效用分布

（Reference）：

[1] Tao M ,Zhu Q .Joint Optimization of Serial Task Offloading and UAV Position for Mobile Edge Computing Based on Multi-Agent Deep Reinforcement Learning[J].Applied Sciences,2025, 15(23):12419-12419.

[2] Liang Z ,Ni Y ,Tao H , et al. Dependency-aware task offloading and energy optimization in UAV-assisted MEC systems[J].Peer-

to-Peer Networking and Applications,2025,19(1):13-13.

[3] He X ,Ni Y ,Tao H .Lyapunov-based queue stability optimization for task offloading in UAV-assisted VEC[J].Pervasive and Mobile Computing,2026,115102126-102126.

[4] Zhao L ,Zhu X ,Cai J .Task Offloading Algorithm for Multiple Unmanned Aerial Vehicles Based on Temporal Graph[J].Sensors,2025,25(21):6759-6759.

[5] Liu Z ,Gao L ,Ma Z , et al. Joint task offloading and resource allocation scheme with UAV assistance in vehicle edge computing networks[J].Computer Networks,2025,273111746-111746.

[6] Abdmeziem R M ,Nacer A A ,Demil S .Proactive handover for task offloading in UAVs[J]. Computer Communications, 2025, 242108282-108282.

[7] Gu W ,Qin T ,Chen D , et al.A dual-layer UAV-assisted mobile edge computing system for disaster rescue: Coordinated optimization of coverage, obstacle-avoidance path planning and task offloading[J].Ad Hoc Networks,2025,178103981-103981.

[8] Wang C ,Liu K ,Yuan Y , et al.Joint trajectory and offloading optimization in UAV-assisted MEC via federated multi-agent reinforcement learning and potential fields[J]. Computer Networks,2025,272111681-111681.

[9] Li H ,Li H .Enhanced energy efficiency in UAV-assisted Mobile Edge Computing through improved hybrid nature-inspired algorithm for task offloading[J].Journal of Network and Computer Applications,2025,243104290-104290.

[10] Yao Z ,Chang P ,Khalil K F , et al.Joint node selection and task offloading via evolutionary game and MATD3 in UAV-assisted MEC networks[J]. Journal of King Saud University Computer and Information Sciences,2025,37(8):220-220.

[11] Liu J ,Hu H ,Bai X , et al.Multi-UAV-Assisted Task Offloading and Trajectory Optimization for Edge Computing via NOMA [J].Sensors,2025,25(16):4965-4965.

[12] Fang Y ,Kuang Z ,Wang H , et al.Minimizing energy consumption of collaborative deployment and task offloading in two-tier UAV edge computing networks[J].Journal of Systems Architecture,2025,167103511-103511.

[13] Zheng Y ,Li A ,Wen Y , et al.A UAV Trajectory Optimization and Task Offloading Strategy Based on Hybrid Metaheuristic Algorithm in Mobile Edge Computing[J].Future Internet,2025, 17(7):300-300.

[14] Nguyen N A ,Hoang C D ,Ngo S T , et al.Secure task offloading and optimization for UAV-assisted NOMA-MEC system with friendly jamming[J]. Physical Communication, 2025, 72102707-102707.

[15] Liu Y ,He Y ,Zhao H , et al.Meta-reinforcement learning-based task offloading method for UAV-enabled mobile edge computing[J].The Journal of Supercomputing,2025,81(8):925-925.

[16] Choi W ,Ahn I .Evolutionary dispersal of ecological species via Multi-Agent Deep Reinforcement Learning[J].Ecological Complexity,2025,64101146-101146.

[17] Zuo P ,Miao C ,Fu C , et al.SMAPPO: A security-aware multi-agent reinforcement learning framework for secure computation offloading in SAGIN[J]. Journal of King Saud University Computer and Information Sciences,2025,37(10):336-336.

[18] Zhang L ,Lu X ,Liu J , et al.Multi-agent reinforcement learning for resource allocation in NOMA-enhanced aerial edge computing networks[J]. Journal of Systems Architecture, 2026, 170103634-103634.

[19] Mei Z ,Zhang Y ,Jiang H , et al. Multi-agent heterogeneous graph reinforcement learning for electric vehicle routing and charging scheduling in coupled power-transportation networks [J].Applied Energy,2026,403(PA):126958-126958.

[20] Wang T ,Na X ,Nie Y , et al.Parallel Task Offloading and Trajectory Optimization for UAV-Assisted Mobile Edge Computing via Hierarchical Reinforcement Learning[J].Drones,2025,9 (5):358-358.

[21] Huang Z ,Kuang Z ,Xu B , et al.Dependency-aware task collaborative offloading and resource allocation in UAV enabled edge computing[J]. Peer-to-Peer Networking and Applications, 2025, 18(3):118-118.

[22] Liu J ,Xie P ,Lin K , et al.Trust-aware task offloading for cost-effective UAV-based edge computing based on reinforcement learning[J]. Neural Computing and Applications, 2024, 37(20): 1-18.

[23] Wang R ,Huang Y ,Lu Y , et al.Robust Task Offloading and Trajectory Optimization for UAV-Mounted Mobile Edge Computing[J].Drones,2024,8(12):757-757.

附录：

主要变量定义

变量符号	含义	变量符号	含义
N_{u}	UAV集群中无人机的总数量	T	仿真总时间步
N_e	用户设备(UE)的总数量	p_e	t时刻UE生成新任务的概率
i,j	UAV或UE的索引编号	N_t	t时刻系统的总任务数
t	时间步索引	λ_t	泊松分布的参数（任务到达率）
$\mathrm{p}_i(\mathrm{t})$	UAV i在时刻t的三维位置向量	σ	任务负载周期性波动系数
v_i	UAV i的恒定速度向量	t_i^{tx}	UE到无人机i的任务传输时间
v_c	UAV的巡航速度	t_i^{proc}	无人机i处理任务的时间
$c_{u,i}$	UAV i的CPU计算容量	π_i	智能体的策略
$B_{u,i}^{\mathrm{up}}$	UAV i的上行传输速率	γ	折扣因子
$B_{u,i}^{\mathrm{down}}$	UAV i的下行传输速率	$R(s_t,a_t)$	t时刻的即时奖励
$E_{u,i}(t)$	UAV i的总能耗	$s_i^{\mathrm{load,pred}}$	负载预测维度
$E_{u,i}^{\mathrm{hover}}$	UAV i的悬停能耗	$q_j(t)$	任务队列维度
$E_{u,i}^{\mathrm{flight}}$	UAV i的飞行能耗	$c_{i,j}(t)$	覆盖状态维度
$E_{u,i}^{\mathrm{compute}}$	UAV i的计算能耗	$n_i^{\mathrm{pending}}(t)$	待处理任务数
$R_{u,i}$	UAV的服务覆盖半径	s_i^{taskstat}	任务特征统计维度
p_j^e	UE的位置向量	$T_i^{\min}$	理论最小延迟
C_e	UE的CPU计算能力	$T_i^{\max}$	截止时间上限
B_e	UE的通信带宽	$C_{j,t}^{\mathrm{total}}$	任务卸载至UAV的综合成本
$T_k(t)$	时刻t分配给节点i的任务集合	x_{iN}	分配给UAV的任务比例
c_j	任务j的CPU计算需求	TR_i	任务完成率贡献
c_i	节点i的CPU总容量	LB_i	负载均衡贡献
$\rho_k(t)$	节点i在时刻t的CPU利用率	EC_i	能耗成本
$B_i^{\mathrm{bat}}(t)$	UAV i在时刻t的剩余电量	DL_i	截止时间延迟成本
$\phi_i(t)$	UAV i归一化电量状态	$\alpha,\beta,\gamma,\delta$	权重系数
$B_{u,i}^{\max}$	UAV i的电池容量	$o_i(t)$	完整状态向量
$R_i^{\mathrm{completion}}$	任务完成率	R_i^{balance}	负载均衡度
R_i^{qos}	服务质量	R_i^{energy}	能耗效率
$\omega_1,\omega_2,\omega_3,\omega_4$	归一化权重系数	T_i^{actual}	任务的实际处理延迟