

DOI:10.16356/j.1005-2615.2023.05.002

## 边缘网络下多无人机协同计算和资源分配 联合优化策略

郭永安<sup>1,2</sup>, 王宇翱<sup>1,2</sup>, 周沂<sup>1,2</sup>, 房晶晶<sup>1,2</sup>, 任保全<sup>3</sup>

(1. 南京邮电大学通信与信息工程学院, 南京 210003; 2. 南京邮电大学边缘智能研究院, 南京 210003;  
3. 军事科学院系统工程研究院, 北京 100101)

**摘要:** 针对边缘网络环境下多人机之间存在计算负载不均, 造成卸载任务失败的问题, 提出了一种多无人机间协作的智能任务卸载方案。通过联合考虑多无人机任务分配、计算资源分配和无人机飞行轨迹, 引入公平性指数建立了无人机公平负载最大化和能量消耗最小化问题。基于多智能体深度强化学习框架, 提出了融合轨迹规划和任务卸载的分布式算法。仿真结果表明, 所提出的多无人机协作方案可以显著提高任务完成率和负载公平度, 并且有效适用于大规模用户设备场景。

**关键词:** 边缘计算; 无人机协作; 任务卸载; 多智能体深度强化学习; 资源分配

**中图分类号:** TN929.52 **文献标志码:** A **文章编号:** 1005-2615(2023)05-0757-11

### Multi-UAV Collaborative Computing and Resource Allocation Joint Optimization Strategy in Edge Networks

GUO Yongan<sup>1,2</sup>, WANG Yuao<sup>1,2</sup>, ZHOU Yi<sup>1,2</sup>, FANG Jingjing<sup>1,2</sup>, REN Baoquan<sup>3</sup>

(1. College of Telecommunications & Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China; 2. Edge Intelligence Research Institute, Nanjing University of Posts and Telecommunications, Nanjing 210003, China; 3. Institute of Systems Engineering, Academy of Military Sciences, Beijing 100101, China)

**Abstract:** Aiming at the problem of uneven computational load among multi-UAVs in the edge network environment, which causes the failure of offloading tasks, a multi-UAV collaborative intelligent task offloading scheme is proposed. By jointly considering task allocation, computing resource allocation, and UAV trajectory, the optimization problem is established to maximize the fairness of computing load among UAVs and minimize the power consumption. Firstly, the optimization problem is modeled as multiple intelligent agent Markov decision process model. Then, a collaborative offloading algorithm based on multi-agent deep deterministic policy gradient (MADDPG) is proposed. The simulation results show that the proposed multi-UAV collaborative scheme can effectively improve task completion rate and fairness of computing load, and it is effectively applicable to large-scale user device scenarios.

**基金项目:** 江苏省前沿引领技术基础研究专项(BK20202001); 江苏省现代农机装备与技术示范推广项目(NJ2022-13)。

**收稿日期:** 2023-09-09; **修订日期:** 2023-10-03

**作者简介:** 郭永安, 男, 教授, 主要从事智能物联网、网络路由技术与协议等研究工作, 现任南京邮电大学边缘智能研究院院长, 泛在网络健康服务系统教育部工程研究中心副主任。发表高水平期刊或会议论文 40 余篇, 授权发明专利 50 余项, 获中国通信学会科学技术一等奖, 中国电子学会技术发明二等奖, 中国指挥与控制学会科技进步一等奖等省部级奖励 10 余项。

**通信作者:** 郭永安, E-mail: guo@njupt.edu.cn。

**引用格式:** 郭永安, 王宇翱, 周沂, 等. 边缘网络下多无人机协同计算和资源分配联合优化策略[J]. 南京航空航天大学学报, 2023, 55(5): 757-767. GUO Yongan, WANG Yuao, ZHOU Yi, et al. Multi-UAV collaborative computing and resource allocation joint optimization strategy in edge networks[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2023, 55(5): 757-767.

**Key words:** edge computing; UAV cooperation; task offloading; multi-agent deep reinforcement learning; resource allocation

随着5G的发展,各种新兴的计算密集型和延迟敏感型应用产生,这对物联网中计算受限的低功耗终端设备提出了严峻挑战<sup>[1]</sup>。移动边缘计算(Mobile edge computing, MEC)作为一种有前途的技术被提出,可以极大减轻物联网终端设备的计算负载<sup>[2]</sup>。无人机因其部署灵活、响应迅速、覆盖范围广等优点,已在各类边缘计算场景广泛应用<sup>[3]</sup>。特别是在地面的基础通信设施遭到毁坏时,无人机能够快速覆盖受灾区域,向地面用户提供通信和计算服务,已经成为应急通信领域的有效解决方案<sup>[4]</sup>。

目前对于单个无人机辅助MEC系统已经有了比较全面的研究,包括任务卸载<sup>[5]</sup>、轨迹规划<sup>[6]</sup>、资源管理<sup>[7]</sup>等。然而,用户的计算需求通常以随机或突发的方式产生,单个无人机完成任务的能力十分有限。因此,多个无人机可以通过协作的方式,为移动终端设备提供更广泛的覆盖范围,以及更灵活的任务卸载方案。文献[8]构建了一个多无人机辅助移动边缘计算系统,多个无人机充当MEC节点,向放置在地面上的物联网终端节点提供扩展的计算能力;文献[9]基于多无人机辅助的MEC网络模型,提出了双重协同任务卸载机制,充分考虑无人机之间的连通性,有效地进行两级卸载。文献[10]提出了一种基于感知与通信集成的计算分流节能策略,联合考虑卸载决策和动态计算资源分配;文献[11]研究了无人机结合无线能量传输技术为物联网终端设备提供能量供应,提出一种基于Stackelberg博弈的能量交易方案,解决了充电用户与无人机之间自由公平的能量交易问题;文献[12]针对基于雾计算的无人机系统电池和计算资源限制问题进行了研究,通过将计算任务卸载问题转化为双边匹配问题,基于迭代算法,最大限度提高无人机之间协同的效率;文献[13]综合考虑端到端卸载链路和地对空卸载链路,联合优化无人机轨迹、任务卸载率和通信资源分配,利用资源分配(Resource allocation, RA)算法和无人机轨迹选择(UAV trajectory selection, UTS)算法交替求解最优化问题。

上述工作主要集中在离线决策,采用迭代和群集算法求解非凸性优化问题。然而,在高度动态的多无人机环境中,任务的随机分配使得优化问题更加复杂且难以生成可行解。深度强化学习则可以有效解决复杂、动态和非凸性问题。文献[14]基于协作多智能体深度强化学习框架,利用双延迟深度

确定性策略梯度算法,考虑高维连续动作空间,求解非凸性优化问题以获得最小执行延迟和能源消耗的策略;文献[15]和文献[16]通过考虑随时变的无人机信道强度和动态资源请求,分别利用MADDPG、深度Q网络(Deep Q-network, DQN)求解最优的资源分配及计算卸载策略;文献[17]将无人机计算卸载调度问题分解为两层子问题,并通过分层强化学习交替优化子问题,与传统优化算法和强化学习相比,有效提高了算法收敛效率;文献[18]提出了一种分布式的多智能体算法,并引入注意力机制,进一步提升分布式无人机代理的性能;文献[19]将优化问题表述为混合合作-竞争博弈,为解决混合动作空间问题,将离散动作转化为连续动作,基于多智能体深度确定性策略梯度(Multi-agent deep deterministic policy gradient, MADDPG)联合优化无人机轨迹及地面用户接入控制;文献[20]考虑多无人机辅助上行通信的场景,主要侧重于无人机与移动用户之间的协同;文献[21-22]专注于用户级通信覆盖及通信服务的公平性,并通过设计无人机的轨迹来实现加权吞吐量最大化、能耗最小化。

上述都利用深度强化学习来求解多无人机任务卸载的优化问题,然而在高度动态的MEC环境下,多个无人机之间实现协作存在着一些挑战。上述文献忽略了每个无人机任务负载的公平性问题,如部分无人机覆盖范围内的任务计算量过大,导致出现过载情况;而部分无人机所包含区域内需要处理的任务计算量较小,计算和通信资源没有得到充分利用,这将导致多无人机之间无法实现公平、高效的协作。

为了解决上述问题,在由用户设备、无人机构成的两层网络架构下,综合考虑无人机轨迹变化、计算资源分配和多无人机公平协作等因素,提出了一种基于多智能体深度强化学习框架,融合轨迹规划和任务卸载的分布式算法。该算法联合优化多无人机节点的卸载决策和资源分配,以实现高可靠和低时延的应急通信领域应用服务为目标。具体创新点如下:

(1)针对应急通信场景下多无人机任务卸载问题,构建了一个由多用户设备、多无人机组成的两层网络架构,并考虑计算任务双重卸载机制,进一步引入公平性指数作为评估无人机之间的任务负载公平性衡量指标。联合考虑无人机轨迹、计算资源分配以及多无人机协作任务分配,在满足任务最

小延迟的前提下实现长期无人机最大公平负载和最小功率消耗。

(2)针对多无人机协作环境动态变化的问题,提出了融合轨迹规划和任务卸载的分布式算法。该算法基于多智能体深度强化学习框架,通过多无人机之间的信息交互和协同学习,可在动态环境下自适应调整卸载方案,输出每个无人机的最佳协作策略,实现分布式决策。

(3)仿真结果表明,提出的无人机协作卸载方案可以有效均衡无人机负载并节省能耗,同时提高任务成功率。与其他卸载方案和决策算法相比,所提方案更适用于多无人机和多用户场景,且各项性能得到了有效提升。

## 1 系统模型与问题建模

构建的多无人机协作任务卸载模型如图1所示。该地区的基站因为自然灾害而无法使用,无人机作为移动的边缘服务器为地面用户设备提供临时计算服务。区域内部署有 $K$ 个地面用户设备、 $M$ 架无人机,其中每架无人机为其覆盖范围内的多个地面用户设备提供服务。无人机和地面用户设备分别用集合 $\mathcal{M}=\{1,2,\dots,M\},m\in\mathcal{M}$ 、 $\mathcal{K}=\{1,2,\dots,K\},k\in\mathcal{K}$ 表示。假设无人机的服务时间为 $T$ ,将 $T$ 分割为 $L$ 个长度为 $\delta$ 的时隙,时隙 $t$ 的集合定义为 $t\in\{1,2,\dots,L\}$ 。

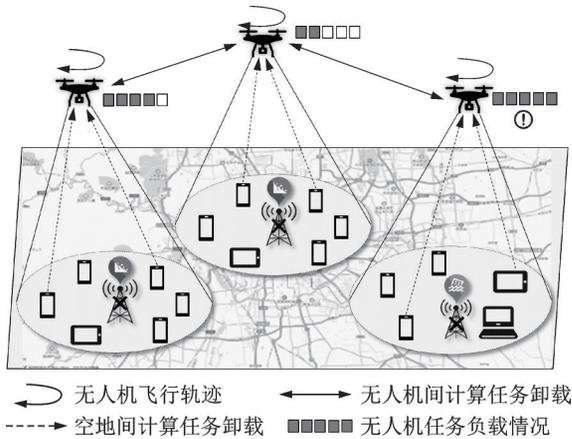


图1 多无人机协同系统模型

Fig.1 Multi-UAV collaborative system model

假设在每个时隙 $t$ 内,地面用户设备 $k$ 产生计算密集型任务 $S_k(t)=\{D_k(t),C_k(t)\}$ ,其中 $D_k(t)$ 表示地面用户设备产生的任务数据量, $C_k(t)$ 表示执行此任务所需的CPU周期总数。考虑到地面用户设备计算能力有限无法进行本地计算,需要将所有任务卸载至无人机。无人机受尺寸、质量、功率等因素的限制,只能提供有限的计算和通信资源,计算任务由多架无人机协同处理。考虑了任务卸

载的两个阶段:(1)地对空卸载,即从地面用户设备卸载到目标无人机,包括地对空传输和目标无人机的计算;(2)空对空卸载,即目标无人机卸载到其他协作无人机,包括空对空传输和各协作无人机间的计算。

### 1.1 无人机移动模型

无人机的轨迹需要规划以节省能量并避免碰撞,因此假设部署每架无人机是为了在一个相应的子区域内为地面用户设备提供服务,并且每个子区域之间没有重叠。

假设所有无人机的飞行高度固定为 $H$ ,在时隙 $t$ 时,无人机 $m$ 的水平坐标为 $L_m(t)=[x_m(t),y_m(t)]$ ,假设无人机 $m$ 飞行的角度为 $\phi_m(t)\in[0,2\pi)$ ,飞行的距离为 $l_m(t)=v_m(t)\delta$ ,则下一个时隙的 $X$ 和 $Y$ 坐标分别为

$$x_m(t+1)=x_m(t)+l_m(t)\cos[\phi_m(t)] \quad (1)$$

$$y_m(t+1)=y_m(t)+l_m(t)\sin[\phi_m(t)] \quad (2)$$

地面用户设备 $k$ 的坐标设置为 $L_k(t)=[x_k(t),y_k(t)]$ ,无人机 $m$ 与地面用户设备 $k$ 之间直线距离定义为 $d_{m,k}(t)$ ,则有

$$d_{m,k}(t)=\sqrt{\|L_m(t)-L_k(t)\|^2+H^2} \quad (3)$$

在每个时隙内,无人机的飞行角度和速度保持不变,无人机 $m$ 的位置相对地面用户设备 $k$ 近似不变。此外,为了保证无人机在服务区域内移动,必须满足移动约束,即

$$0\leq x_m(t)\leq X_{\max} \quad (4)$$

$$0\leq y_m(t)\leq Y_{\max} \quad (5)$$

式中: $X_{\max}$ 和 $Y_{\max}$ 分别为无人机服务区域的长度和宽度。

无人机的水平覆盖半径为 $R$ ,如果地面用户设备位于某个无人机的覆盖范围内,这些设备将由同一架无人机提供服务。任意两架无人机 $m,j$ 之间的距离表示为 $d_{m,j}(t)$ ,为了保证两者服务范围不相互重叠,必须满足以下重叠约束

$$d_{m,j}(t)\geq 2R \quad (6)$$

为了避免任意两架无人机发生碰撞,无人机之间的距离应该不小于安全距离 $d_{\min}$ 。

### 1.2 协作传输模型

设计了两阶段协同计算卸载架构,如图2所示,即地面用户设备卸载到目标无人机和无人机之间协同卸载,接下来将对两阶段中任务卸载的时延和能耗模型进行分析。

(1)当目标无人机 $m$ 接收到地面用户设备产生的计算任务时,二进制变量 $\alpha_{k,m}(t)=1$ ,否则 $\alpha_{k,m}(t)=0$ ,且每个地面用户设备 $k$ 在时隙 $t$ 内最多

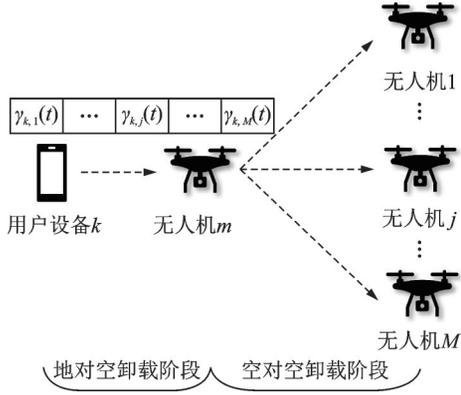


图2 两阶段协同计算卸载架构

Fig.2 Two-stage collaborative computing offloading architecture

与一个目标无人机进行传输。

$$\sum_{m \in \mathcal{M}} \alpha_{k,m}(t) \leq 1 \quad (7)$$

设目标无人机  $m$  在时隙  $t$  处服务的地面用户设备集合为  $\mathcal{N}_m(t) = \{1, 2, \dots, N_m(t)\}$ , 地面用户数量为

$$N_m(t) = \sum_{k \in \mathcal{N}} \alpha_{k,m}(t) \quad (8)$$

系统中的地面用户设备将任务卸载到目标无人机采用正交频分复用方案, 并采用视距(Line of sight, LoS)链路进行通信, 不考虑诸如阴影或小尺度衰落等其他信道损害。路径损耗<sup>[14]</sup>为

$$\rho(t) = \frac{g}{|d_{m,k}(t)|^2} \quad (9)$$

式中:  $g$  表示参考距离为 1 m 的功率增益。从地面用户设备  $k$  到目标无人机  $m$  的上行传输速率为

$$R_{k,m}^{\text{up}}(t) = \frac{B}{N_m(t)} \log_2 \left[ 1 + \frac{\rho(t)P_k}{\sigma_m^2} \right] \quad (10)$$

式中:  $P_k$  为地面用户设备的发射功率,  $\sigma_m^2$  为每架无人机处的加性白高斯噪声功率。在任务卸载过程中, 无人机带宽  $B$  被平均分配给服务的每个地面用户设备。则地面用户设备  $k$  的任务上行传输时间  $T_{k,m}^{\text{up}}(t)$  为

$$T_{k,m}^{\text{up}}(t) = \frac{D_k(t)}{R_{k,m}^{\text{up}}(t)} \quad (11)$$

那么, 目标无人机  $m$  的接收能耗为

$$E_{k,m}^{\text{up}}(t) = P_m^r T_{k,m}^{\text{up}}(t) = \frac{D_k(t)P_m^r}{R_{k,m}(t)} \quad (12)$$

式中  $P_m^r$  为目标无人机  $m$  的接收功率。

(2) 目标无人机  $m$  在接受到地面用户设备卸载任务后, 会将任务分为独立的  $M$  份, 在自己执行计算任务的同时, 将  $M-1$  份任务发送给其余协作无人机协同计算, 在时隙  $t$  目标无人机  $m$  承载的地面用户设备  $k$  产生的任务比例为  $\gamma_{k,m}(t) \in [0, 1]$ , 协

作无人机  $j$  所承担的任务比例为  $\gamma_{m,j}(t) \in [0, 1]$ , 则有

$$\gamma_{k,m}(t) + \sum_{j \in \mathcal{M}, m \neq j} \gamma_{m,j}(t) = 1 \quad m \in \mathcal{M} \quad (13)$$

目标无人机  $m$  和协作无人机  $j$  之间的路径损耗参考文献[23], 即

$$\rho_{m,j}(t) = \xi(t) + \eta_{\text{LoS}} \quad (14)$$

式中:  $\eta_{\text{LoS}}$  为添加到 LoS 链路自由空间传播模型中的附加衰减因子,  $\xi(t)$  根据自由空间路径损失模型<sup>[24]</sup>得出

$$\xi(t) = 20 \lg [d_{m,j}(t)] + 20 \lg f_c - 147.55 \quad (15)$$

式中  $f_c$  为系统载波频率。目标无人机  $m$  对协作无人机  $j$  之间的信道增益为

$$G = 10^{-\frac{L_{m,j}}{10}} \quad (16)$$

记  $P_m^1$  为目标无人机的发送功率, 那么目标无人机  $m$  向协作无人机  $j$  传输数据的速度为

$$R_{m,j}(t) = \frac{B}{M-1} \log_2 \left[ 1 + \frac{P_m^1(t)G}{\sigma_m^2} \right] \quad (17)$$

其中目标无人机  $m$  将带宽平均分配给其余协作无人机。注意, 目标无人机的带宽正交分为两部分: (1) 与用户设备数据传输的带宽; (2) 与协作无人机之间数据传输带宽。因此, 两阶段卸载的数据传输之间没有干扰。

从目标无人机  $m$  到协作无人机  $j$  的传输时间和能耗分别为

$$T_{m,j}^{\text{tran}}(t) = \frac{\gamma_{m,j}(t)D_k(t)}{R_{m,j}(t)} \quad (18)$$

$$E_m^{\text{tran}}(t) = \sum_{j \in \mathcal{M}, m \neq j} P_m^1 T_{m,j}^{\text{tran}}(t) \quad (19)$$

### 1.3 协作计算模型

从地面用户设备  $k$  处接收到任务后, 目标无人机  $m$  决定自身和协作无人机计算多少任务。

(1) 考虑任务比例  $\gamma_{k,m}(t)$ , 目标无人机  $m$  处的计算延迟为

$$T_{k,m}^{\text{comp}}(t) = \frac{\gamma_{k,m}(t)C_k(t)}{f_{k,m}(t)} \quad (20)$$

式中:  $f_{k,m}(t)$  为目标无人机  $m$  分配给地面用户设备  $k$  的计算资源。则目标无人机  $m$  处理计算任务的能耗为

$$E_{k,m}^{\text{comp}}(t) = \sum_{k \in \mathcal{K}} q [f_{k,m}(t)]^2 \gamma_{k,m}(t) C_k(t) \quad (21)$$

式中  $q$  为无人机计算芯片的有效开关电容。

(2) 协作无人机  $j$  处的计算延迟和能耗分别为

$$T_{k,j}^{\text{comp}}(t) = \frac{\gamma_{m,j}(t)C_k(t)}{f_{k,j}(t)} \quad (22)$$

$$E_{k,j}^{\text{comp}}(t) = \sum_{k \in \mathcal{K}} q [f_{k,j}(t)]^2 \gamma_{m,j}(t) C_k(t) \quad (23)$$

协作无人机  $j$  处的总时延由无人机-无人机通信传输时延和计算时延组成,可以表示为

$$T_{k,j}(t) = T_{m,j}^{\text{tran}}(t) + T_{k,j}^{\text{comp}}(t) \quad (24)$$

#### 1.4 问题建模

目标无人机的计算和传输是同步进行的,由于传输、计算的并行性,任务完成延迟取决于任务每部分的最大值,地面用户设备  $k$  在时隙  $t$  中产生的任务的总延迟可以建模为

$$T_k(t) = T_{k,m}^{\text{up}}(t) + \max\{T_{k,m}^{\text{comp}}(t), T_{k,j}(t)\} \quad m, j \in \mathcal{M} \quad (25)$$

目标无人机  $m$  的总能耗包括向协作无人机传输任务的能耗、计算以及协作计算的能耗、接收服务区域内地面用户设备任务的能耗,可以计算为

$$E_m(t) = E_m^{\text{tran}}(t) + E_{k,m}^{\text{comp}} + \sum_{k \in \mathcal{K}_m(t)} E_{k,m}^{\text{up}}(t) \quad (26)$$

在协作模型中,每个无人机可能计算到所有地面用户产生的任务,那么目标无人机  $m$  在时隙  $t$  内的任务负载为

$$D_m(t) = \sum_{k \in \mathcal{K}} D_k(t) \gamma_{k,m}(t) \quad (27)$$

然而,直接最大化所有无人机的总负载可能导致不公平问题,因为一部分无人机可能会过载,而其余无人机的计算资源会闲置,导致服务质量较差。

引入 Jain 公平性指数<sup>[25]</sup>来评估无人机之间的任务负载公平性

$$f(t) = \frac{\left[ \sum_{m=1}^M D_m(t) \right]^2}{M \left[ \sum_{m=1}^M (D_m(t))^2 \right]} \quad (28)$$

本文提出了一个多目标优化问题,该问题联合考虑无人机轨迹、计算资源分配以及多无人机协作任务分配,在满足任务最小延迟的前提下实现长期无人机最大公平负载以及最小功率消耗。

$$\left\{ \begin{array}{l} \text{P}_1: \max_{\substack{\phi_m(t), v_m(t) \\ r_{k,m}(t), f_{k,m}(t)}} \sum_{t=1}^L \frac{f(t)}{\sum_{m=1}^M [E_m(t)]} \\ \text{s.t. } C_1: v_m(t) \leq v^{\max} \quad \forall m \in \mathcal{M} \\ C_2: 0 \leq x_m(t) \leq X_{\max} \quad \forall m \in \mathcal{M} \\ C_3: 0 \leq y_m(t) \leq Y_{\max} \quad \forall m \in \mathcal{M} \\ C_4: d_{m,j}(t) \geq 2R \quad \forall m, j \in \mathcal{M}, n \neq j \\ C_5: d_{m,j}(t) \geq d_{\min} \quad \forall m, j \in \mathcal{M}, n \neq j \\ C_6: \sum_{m \in \mathcal{M}} \alpha_{k,m}(t) \leq 1 \quad \forall k \in \mathcal{K} \\ C_7: \sum_{m \in \mathcal{M}} \gamma_{k,m}(t) = 1 \quad \forall k \in \mathcal{K} \\ C_8: \sum_{k \in \mathcal{K}} f_{k,m}(t) = f \quad \forall m \in \mathcal{M} \\ C_9: T_k(t) \leq T_k^{\max}(t) \quad \forall k \in \mathcal{K} \end{array} \right. \quad (29)$$

式中:  $v^{\max}$  为无人机的最大飞行速度,  $X_{\max}$  和  $Y_{\max}$  分别表示区域的最大范围,  $R$  表示无人机的覆盖范围半径,  $d_{\min}$  为无人机之间的最小安全距离,  $f$  为无人机的计算能力,  $T_k^{\max}(t)$  为地面用户  $k$  产生任务的最大延迟容忍时间,超出该时间表示该任务失败。

约束  $C_1$  表示无人机的速度不能超过最大值;约束  $C_2$ 、 $C_3$ 、 $C_4$ 、 $C_5$  描述了无人机的轨迹约束;约束  $C_6$  意味着地面用户设备同一时隙中最多可以关联一个无人机;约束  $C_7$  表示多个任务总和等于任务的大小;约束  $C_8$  表示无人机分配给各任务的计算能力之和为无人机计算总能力;约束  $C_9$  表示完成任务的时延必须在可承受范围内。

面向多地面用户设备计算任务时变、多无人机高度动态的计算任务卸载场景,某些未知变量(如无人机的位置和信道条件)可以影响能量消耗和执行延迟,特别是在由无人机的移动性引起的动态网络中。因此,优化问题是一个大规模、高维度且具有诸多限制条件的复杂问题,即使能够描述出这种移动边缘计算的卸载模型及问题,求解过程也会相当困难。此外,决策解空间较大,用传统的优化方法很难得到最优策略。为了解决这些挑战,本文面向多无人机协作移动边缘计算任务卸载,设计了一种融合轨迹规划和任务卸载的分布式算法,在较少环境信息的情况下学习接近最优的策略。

## 2 融合轨迹规划和任务卸载的分布式算法

本文提出的一种基于多智能体深度强化学习(Multi-agent deep reinforcement learning, MADRL)框架,融合轨迹规划和任务卸载的分布式算法,使得无人机自主决策。基于 MADRL 的解决方案与传统优化方案相比具有以下优势:(1)集中式的方案需要基站或者云中心来控制所有无人机,而对于基于 MADRL 的解决方案,无人机在训练完成后即可自主执行计算任务;(2)对 MADRL 来说,每架无人机都是独立的计算单元,无人机数量增加不会使算法复杂度显著增加,而传统方案的计算复杂度对无人机数量敏感,容易造成系统崩溃;(3)集中式的控制方式存在中心故障导致整体系统瘫痪的风险,而在基于 MADRL 的系统中,单个无人机的故障不会影响其他无人机执行任务。首先将上述问题建模为一种多智能体决策模型,然后通过采用 MADRL 算法来解决。

### 2.1 多智能体协作决策模型

在系统中,无人机确定其飞行轨迹、计算资源分配和任务分配比以获得最大公平负载和最小功率消耗,系统当前的环境状态由先前的系统状态和

所有无人机的动作决定。在这种情况下,任务卸载优化问题式(29)可以被建模为多智能体马尔可夫决策模型,本文把每个无人机定义为一个智能体,并将第1节中描述的无人机协同计算卸载场景作为一个学习环境,每个智能体独立做出决策,执行轨迹规划,计算资源分配和任务分配。用 $(\mathcal{S}, \{\mathcal{A}_m\}, \{\mathcal{R}_m\}, \lambda)$ 表示决策模型,其中 $\mathcal{S}$ 为 $M$ 个智能体的状态集合, $\{\mathcal{A}_m\}$ 为所有智能体的动作集合, $\mathcal{A}_m$ 为无人机 $m$ 的动作空间, $\{\mathcal{R}_m\}$ 为所有无人机的奖励函数集合, $P$ 为当前所有智能体描述状态转移概率, $\lambda \in [0, 1]$ 为折扣因子。

(1)动作空间 $\mathcal{A}_m$ :无人机的动作包括飞行轨迹(飞行速度和角度)、计算资源的分配以及任务计算量分配,则第 $m$ 个无人机在 $t$ 时刻的动作表示为: $a_m(t) = \{\phi_m(t), \nu_m(t), \gamma_m(t), f_{k,m}(t), \forall k\}$ 。根据约束条件,4个变量的取值范围分别为 $\phi_m(t) \in [0, 2\pi)$ ,  $\nu_m(t) \in [0, \nu^{\max})$ ,  $\gamma_m(t) \in [0, 1]$ ,  $f_{k,m}(t) \in [0, f]$ 。

此外,为了消除变量多样性对系统性能的影响,对变量进行归一化处理: $\bar{\phi}_m(t) = \phi_m(t)/2\pi$ ,  $\bar{\nu}_m(t) = \nu_m(t)/\nu^{\max}$ ,  $\bar{f}_{k,m}(t) = f_{k,m}(t)/f$ 。归一化处理后,动作空间为

$$a_m(t) = \{\bar{\phi}_m(t), \bar{\nu}_m(t), \bar{\gamma}_{k,m}(t), \bar{f}_{k,m}(t), \forall k\} \quad (30)$$

(2)状态空间 $\mathcal{S}$ :每个无人机采取的动作和当前时刻所在的位置,都会影响当前时隙的环境。而每个无人机的观察为

$$o_m(t) = \{x_m(t), y_m(t), \bar{\phi}_m(t-1), \bar{\nu}_m(t-1), \bar{\gamma}_{k,m}(t-1), \bar{f}_{k,m}(t-1), \forall k\} \quad (31)$$

同样对无人机观察到的空间状态进行归一化,则在 $t$ 时刻无人机 $m$ 的观察为

$$o_m(t) = \{\bar{x}_m(t), \bar{y}_m(t), \bar{\phi}_m(t-1), \bar{\nu}_m(t-1), \bar{\gamma}_{k,m}(t-1), \bar{f}_{k,m}(t-1), \forall k\} \quad (32)$$

式中: $\bar{x}_m(t) = x_m(t)/X_{\max}$ ,  $\bar{y}_m(t) = y_m(t)/Y_{\max}$ 。状态空间为所有无人机对环境的观察,无人机之间可以彼此交换信息,以便知道所有无人机和UE的位置,则 $t$ 时刻的状态空间为

$$s(t) = \{o_m(t) | m \in \mathcal{M}\} \quad (33)$$

(3)奖励:为了解决公式化的任务卸载优化问题式(29), $M$ 个无人机应该在满足约束的同时最大化任务负载公平、最小化能耗。奖励设计分为公平性奖励、节能奖励以及惩罚。基于式(28),所有无人机的公平性奖励为 $f(t)$ 。然后,在满足所有约束条件的情况下,将目标无人机 $m$ 的节能奖励定义为能量消耗 $E_m(t)$ 的负值。如果不满足某些约束,则在奖励函数中将存在相应的惩罚,惩罚定义为

$$\eta_m(t) = \zeta_1(t)\eta_1 + \zeta_2(t)\eta_2 + \zeta_3(t)\eta_3 + \zeta_4(t)\eta_4 \quad (34)$$

式中: $\eta_1, \eta_2, \eta_3, \eta_4$ 分别表示与界限约束( $C_2, C_3$ )、重叠约束( $C_4$ )、安全距离约束( $C_5$ )和任务完成( $C_9$ )相关的惩罚。如果无人机飞出了边界,则指示符 $\zeta_1(t) = 1$ ,否则为0;如果无人机的覆盖范围与其他无人机重叠,则指示符 $\zeta_2(t) = 1$ ,否则为0;当无人机与任意无人机之间的距离不满足安全距离约束, $\zeta_3(t)$ 为1,否则为0;当无人机覆盖范围内的任务超时而导致失败时,无人机的任务超时惩罚指示符 $\zeta_4(t) = 1$ ,否则为0。

则在 $t$ 时刻无人机 $m$ 的奖励函数为

$$r_m(t) = f(t) - e_m(t) - \eta_m(t) \quad (35)$$

式中: $e_m(t) = E_m(t) / \sum_{t=1}^L E_m(t)$ 为能耗的归一化处理。

## 2.2 联合优化算法

为了求解上述多智能体协作决策模型,得出每个无人机的最佳协作策略,基于MADDPG的思路,提出了融合轨迹规划和任务卸载的分布式算法。算法采用集中式训练、分散式决策的范式,如图3所示。多个无人机协同工作,每个无人机上由Actor模块、Critic模块和经验存储器组成。Actor模块的输入为每个无人机的局部观测值,输出为动作;Critic模块在集中式训练阶段工作,可以访问所有无人机的观察和动作来评估Actor模块的输出。每个无人机的目标是找到一个策略最大化自己的奖励 $r_m(t)$ 。

所有无人机通过协作以实现总奖励的最大化

$$r(t) = \sum_{m=1}^M r_m(t) \quad (36)$$

追求每个无人机奖励的最大化,容易引起无人机之间的竞争,不利于系统的整体性能提升。通过设计总奖励的最大化,各无人机之间以分布式的方式协作处理计算任务,保证纳什均衡。

Actor模块和Critic模块各有2个神经网络构成,分别为目标网络和评价网络,则Actor模块和Critic模块的评价网络分别为 $\mu_m(o_m; \theta_m^{\mu})$ 和 $Q_m(s, a; \theta_m^Q)$ ,每个模块的评价网络都有对应的目标网络,分别为 $\mu'_m(o_m; \theta_m^{\mu'})$ 和 $Q'_m(s, a; \theta_m^{Q'})$ 。算法1中详细描述了融合轨迹规划和任务卸载的分布式算法的集中训练流程。

首先初始化4个神经网络的参数和经验存储器,在每个训练轮次中,无人机根据自身观察 $o_m(t)$ ,由Actor模块评价网络输出动作 $a_m(t) = \mu_m(o_m(t))$ ,所有无人机都执行飞行角度 $\phi_m(t)$ 、飞

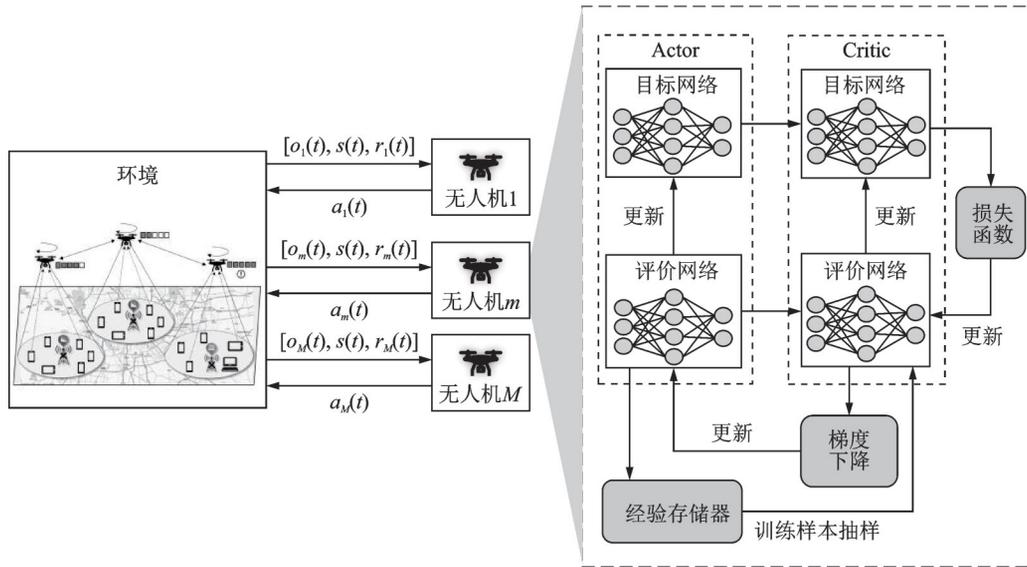


图 3 基于 MADDPG 的协作卸载算法

Fig.3 Collaborative offloading algorithm framework based on MADDPG

行速度  $v_m(t)$ 、计算资源分配  $f_{k,m}(t)$  以及多无人机协作任务分配  $\gamma_{k,m}(t)$ , 所有无人机执行完动作得到共同奖励  $r(t)$  和下一时刻环境状态  $s(t+1)$ 。

在本算法中使用了经验回放机制, 在每个时隙中, 所有无人机可以将自身产生的经验元组  $(a(t), r(t), s(t), s(t+1))$  存储到大小为  $\mathcal{B}$  的经验存储器中, 其中  $a(t) = \{a_1(t), a_2(t), \dots, a_M(t)\}$ 。如果经验存储器已满, 则新生成的经验元组将替换旧的经验元组。Actor 模块和 Critic 模块可以通过在经验存储器中批量采样, 获得训练数据。每个无人机都在经验存储器中随机采样一批大小为  $M_b$  的样本  $(a_i, r_i, s_i, s'_i)$ , 随机采样可以打破样本数据之间的相关性, 并减少训练振荡。

算法通过最小化损失来更新无人机  $m$  的 Critic 评价网络参数  $\theta_m^Q$ , 即

$$L(\theta_m^Q) = \frac{1}{M_b} \sum_{i=1}^{M_b} [y_m(i) - Q_m(s(i), a(i); \theta_m^Q)]^2 \quad (37)$$

$$y_m(i) = r_m(i) + \epsilon Q'_m(s(i+1), (\mu'_1(s(i+1); \theta_1^{Q'}), \dots, \mu'_M(s(i+1); \theta_M^{Q'}); \theta_m^{Q'})) \quad (38)$$

式中  $\epsilon$  为折扣因子。每个无人机可以使用策略梯度更新 Actor 评价网络的参数  $\theta_m^\mu$ , 即

$$\nabla_{\theta_m^\mu} J(\mu_m) = \mathbb{E}_{s, a \sim \mathcal{B}} [\nabla_{\theta_m^\mu} \mu_m(a_m | o_m) \nabla_a Q'_m(s, a) |_{a_m = \mu_m(o_m)}] \quad (39)$$

Actor 模块和 Critic 模块的目标网络参数采用软更新的方式, 分别为

$$\begin{cases} \theta_m^{\mu'} = \tau \theta_m^\mu + (1 - \tau) \theta_m^{\mu'} \\ \theta_m^{Q'} = \tau \theta_m^Q + (1 - \tau) \theta_m^{Q'} \end{cases} \quad (40)$$

式中  $\tau$  为更新率。在分布式执行阶段, 无人机的飞行方向、计算资源和任务分配的决定是基于训练完成的神经网络执行。

**算法 1** 融合轨迹规划和任务卸载的分布式算法

初始化每个无人机的经验存储器空间  $\mathcal{B}$

初始化每个无人机的 Actor 评价网络参数  $\theta_m^\mu$ , Actor 目标网络参数  $\theta_m^{\mu'}$ , Critic 评价网络参数  $\theta_m^Q$ , Critic 目标网络参数  $\theta_m^{Q'}$

- (1) for  $episode = 1, 2, \dots, N$  do
- (2) 初始化状态  $s(t)$ ;
- (3) for  $t=1, L$  do
- (4) for  $m=1, M$  do
- (5) 无人机  $m$  监测环境获得  $o_m(t)$
- (6) 无人机  $m$  执行动作  $a_m(t) = \mu_m(o_m(t))$
- (7) end for
- (8) 所有无人机执行完动作得到共同奖励  $r(t)$  和下一时刻环境状态  $s(t+1)$
- (9) for  $m=1, M$  do
- (10) if 经验存储器内未滿
- (11) 将  $(a(t), r(t), s(t), s(t+1))$  存储到经验存储器
- (12) else
- (13) 使用  $(a(t), r(t), s(t), s(t+1))$  将经验存储器中最早的一组经验替换
- (14) 在所有无人机的经验存储器中进行随机采样获得  $(a(t), r(t), s(t), s(t+1))$
- (15) 通过最小化损失函数来更新 Critic 评价网络参数  $\theta_m^Q$
- (16) 计算策略梯度来更新 Actor 评价网络  $\theta_m^\mu$
- (17) 对 Actor 目标网络和 Critic 目标网络分别进行软更新:
- (18)  $\theta_m^{\mu'} = \tau \theta_m^\mu + (1 - \tau) \theta_m^{\mu'}$
- (19)  $\theta_m^{Q'} = \tau \theta_m^Q + (1 - \tau) \theta_m^{Q'}$
- (20) end if
- (21) end for
- (22) end for
- (23) end for

### 2.3 算法复杂度分析

对提出的分布式算法分别从时间复杂度和空间复杂度两个角度进行分析。

时间复杂度:无人机选择动作的操作时间复杂度为 $O(MLA)$ ,其中 $M$ 为无人机的数量, $L$ 为系统的时隙数, $A$ 为每个无人机输出的动作个数。同时考虑神经网络的前向传播和反向传播,其操作时间复杂度为 $O(MM_bN_p)$ ,其中 $M_b$ 为每个时隙的批次大小, $N_p$ 为神经网络的参数数量。则本算法的时间复杂度为 $O(MLA + MM_bN_p)$ 。

空间复杂度:本算法需要考虑存储神经网络参数和经验。其中,存储神经网络参数的空间复杂度为 $O(N_p)$ 。此外,经验存储器的空间复杂度为 $O(BD)$ ,其中 $B$ 为经验存储器的大小(存储的经验样本数量); $D$ 为每个经验样本的维度, $D=(10+6k)m+1$ ,取决于状态、动作、奖励和下一个状态的维度。则本算法的时间复杂度为 $O(N_p + BD)$ 。

## 3 仿真实验与性能评估

在 Python 3.8.13 和 PyTorch 1.12.1 平台上对所提的多无人机协作架构和决策算法的有效性进行性能验证,计算机环境为 Ubuntu 20.04.5、Intel Xeon Gold 5218R CPU、NVIDIA GeForce RTX 3090。

### 3.1 系统设置

多个地面用户设备分布在面积为 $400\text{ m} \times 400\text{ m}$ 的区域内,多个无人机为地面用户设备提供计算服务,飞行高度固定为 $50\text{ m}$ ,服务半径为 $20\text{ m}$ ,每个时隙长度为 $0.5\text{ s}$ 。具体的系统模型参数和算法设置<sup>[14,26]</sup>分别如表1和表2所示。

表1 系统参数设置

参数	取值
地面用户设备任务数据量 $D_k(t)/\text{Mb}$	[0.2, 1]
任务计算CPU周期数 $C_k(t)/\text{cycle}$	$[2 \times 10^7, 5 \times 10^8]$
任务容忍时延 $T_k^{\max}(t)/\text{s}$	[0.5, 1]
功率增益 $g/\text{dB}$	-50
系统带宽 $B/\text{MHz}$	10
地面用户设备发射功率 $P_k/\text{W}$	0.1
无人机高斯白噪声功率 $\sigma_m^2/\text{dBm}$	-100
无人机 $m$ 的接收功率 $P_m/\text{W}$	0.1
无人机件的安全距离 $d_{\min}/\text{m}$	1
系统载波频率 $f_c/\text{GHz}$	2
LoS链路附加衰减因子 $\eta_{\text{LoS}}/\text{dB}$	5
无人机发射功率 $P_m^s/\text{W}$	5
无人机计算能力 $f/\text{GHz}$	3
计算芯片有效开关电容 $q$	$10^{-28}$

表2 算法参数设置

参数	取值
学习率	0.000 3
批规模/个	64
折扣因子	0.99
经验存储器大小/个	6 000
惩罚 $\eta_1\eta_2\eta_3\eta_4$	1

### 3.2 仿真分析

首先针对多无人机协作架构和算法进行仿真,验证无人机轨迹和任务分配策略的有效性。然后与3种方案或算法进行对比验证,以证明本文方案在系统性能上的提升。

设置3架无人机在固定区域内向20个地面用户设备提供计算服务,地面用户设备随机分布在区域内。无人机飞行轨迹如图4所示,3架无人机同时从区域中心出发,因为无人机服务范围有限,必须移动以尽可能服务更多的地面用户设备。此外,各架无人机以合作的方式覆盖区域,“无人机2”从初始位置移动到右下角以服务更多地面用户设备,而后和“无人机3”一起移动到右上角处理该区域中的地面用户设备任务。

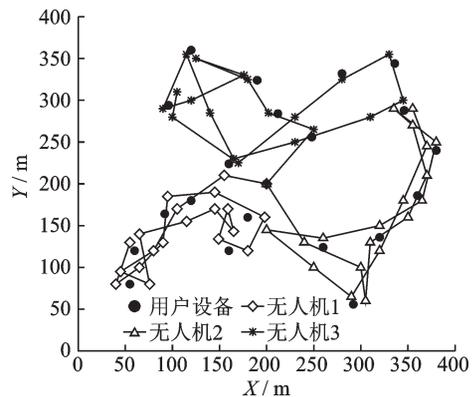


图4 无人机飞行轨迹图

Fig.4 UAV flight trajectory diagram

图5呈现了最优任务协作分配策略。由于3个无人机分别覆盖3个地面用户设备热点区域,因此部分地面用户设备的任务超过50%由其目标无人机进行计算,其余部分由另外两个无人机协作完成。此外,对于计算任务较大的地面用户设备,或者在目标无人机当前计算负载较大时,该任务将综合考虑传输功耗和任务完成时延进行分配。

为了验证协作架构和算法的优越性,将其与以下3种方案进行对比:

(1)无协作方案。所有任务均在单个无人机上进行处理,不再卸载给其他无人机,但仍使用MADDPG决策飞行轨迹和计算资源分配。

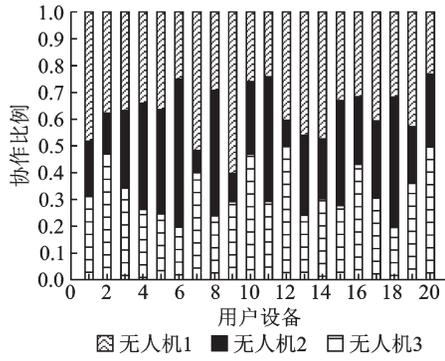


图 5 无人机任务协作分配

Fig.5 Collaborative task assignment of UAVs

(2)深度确定性策略梯度(Deep deterministic policy gradient, DDPG)决策。DDPG 是一种单智能体深度强化学习算法,采用集中式决策输出所有无人机的动作。

(3)随机方案。无人机随机选择自身动作,包括飞行轨迹、计算资源分配和任务分配。

首先在 3 架无人机,20 个地面用户设备的环境下对不同策略的系统能耗进行对比,图 6 给出了不同卸载方案下的系统总能耗。所提算法通过联合优化无人机飞行轨迹和任务分配,在保证任务成功完成的同时,可优化系统总能耗,与基于 DDPG 的方案比较,能耗平均降低了 34%。无协作方案由于无人机之间无任务交互,通信能耗较小。

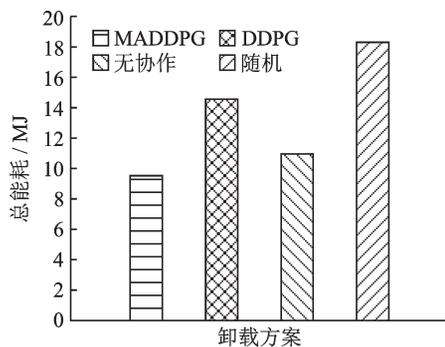


图 6 不同卸载方案下的系统能耗

Fig.6 System energy consumption with different offloading schemes

在地面用户设备数固定为 20 的情况下,研究无人机数量对负载公平指数和任务成功率影响,任务成功率为所有地面用户设备因超时未完成任务量与所有任务的比值。

无人机数量与负载公平指数的关系如图 7 所示。仿真结果表明,随着无人机数量的增加,系统的负载公平指数不断增加,这是因为无人机的数量增加会使得系统总计算能力增加,地面用户设备可以更均匀地接受无人机服务。所提出的基于 MADDPG 算法的协作方案表现最优,负载公平指

数可以维持在 90% 以上,这是因为每个无人机独立决策从而可以实现性能均衡。DDPG 决策依赖于集中式决策,当无人机数量增加时会给求解带来挑战,因此采用该算法的负载公平指数低于 MADDPG 方案。无协作方案由于无人机之间不能互相协助,其性能低于两类协作方案。随机策略因其输出动作的随机性,负载公平指数在 40%~50% 之间波动,表现最差。

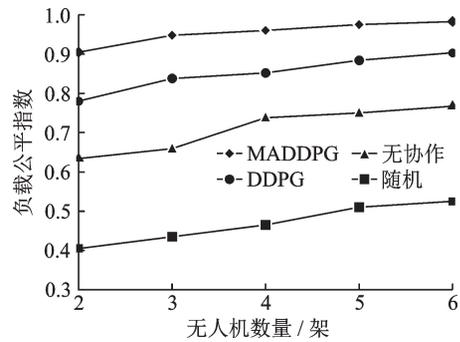


图 7 无人机数量对负载公平指数的影响

Fig.7 Effect of number of UAVs on the load fairness index

无人机数量与任务完成率的关系如图 8 所示。仿真结果表明,随着无人机数量的增加,任务成功率不断增加。在无人机数量为 6 时,基于 MADDPG 的协作卸载方案的任务成功率接近 100%,此时地面用户设备的计算任务都可以被无人机及时处理。基于集中式决策的 DDPG 方案在无人机数量增加时,任务成功率不会明显增加,这是由于无人机数量增加导致动作空间增大,输出最优动作困难。无协作方案在计算任务时,部分无人机因计算能力不足导致任务超时而失败,但其路径规划和计算能力分配仍采用多智能体算法框架,在无人机数量足够多时仍能达到 60% 以上的任务完成率。

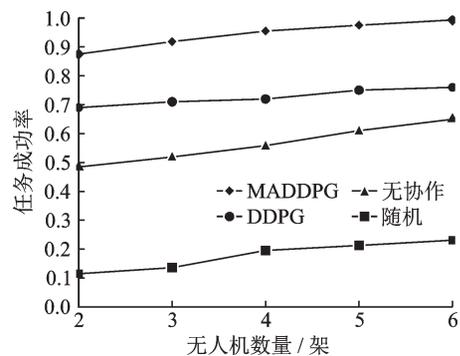


图 8 无人机数量对任务成功率的影响

Fig.8 Effect of number of UAVs on the task success rate

本文研究了无人机固定为 3 架的情况下,地面用户设备数量对负载公平指数和任务成功率的影响。地面用户设备数量与负载公平指数的关系如图 9 所示。仿真结果表明,随着地面用户设备数量的增加,系统的负载公平指数不断降低。所提出的

基于MADDPG算法的协作方案在用户设备数达到30个时,负载公平指数仍在90%左右,对比其他方案和算法,验证了在大规模用户场景下的有效性。DDPG决策在地面用户设备数量在5~20之间时,性能下降较低,但超过20个之后,性能急剧下降,这是由于该算法为集中式决策,动作空间指数式增加导致无人机难以输出最优动作维持负载均衡。

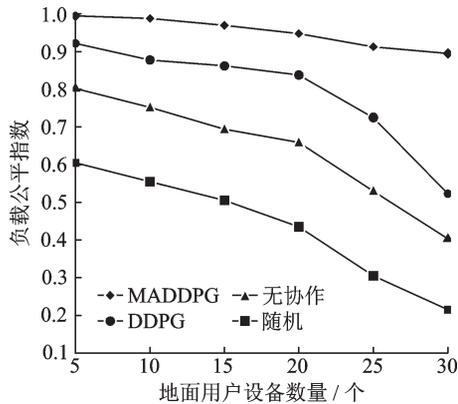


图9 地面用户设备数量对负载公平指数的影响

Fig.9 Effect of number of user devices on the load fairness index

地面用户设备数量与任务完成率的关系如图10所示。仿真结果表明,随着地面用户设备数量的增加,任务成功率不断降低。这是因为计算量不断增加,而无人机计算资源不足导致,同时地面用户设备竞争的系统带宽固定,通信速率降低导致任务失败。在设备数量较少时,本文算法和DDPG算法性能差距较小,随着用户设备增加,DDPG算法方案性能急剧下降,与MADDPG算法的任务成功率相差53%,这是由于大量用户设备导致全局信息获取困难。无协作方案由于无人机之间计算资源无法共享,任务不能及时处理导致失败。随机方案由于不确定性较大,性能最差。

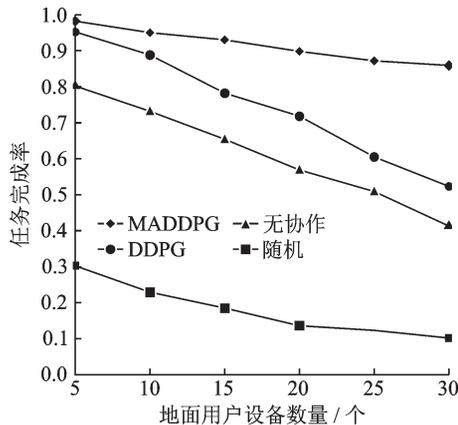


图10 地面用户设备数量对任务完成率的影响

Fig.10 Effect of number of user devices on the task success rate

## 4 结 论

从联合计算和通信的角度考虑动态变化边缘计算环境下的多无人机协作任务卸载问题,提出了多无人机协同计算和分布式卸载决策方案。该方案以在多地面用户设备环境下实现最小化长期能量消耗和最大化负载公平为目标。首先,将该问题建模为分布式多智能体马尔可夫决策过程,然后利用多智能体深度强化学习框架的学习优势,使每架无人机实现分布式决策。通过联合优化无人机飞行轨迹、计算资源分配和多无人机任务协作比例,最终实现降低传输和计算能量消耗,提高任务完成率和负载公平度。仿真结果验证了多无人机协作计算架构和分布式算法对多用户任务多变的边缘计算环境性能提升的有效性,并揭示了不同无人机数量和地面用户设备数量对所提方案性能的影响规律。下一步工作将在本文基础上考虑能量搜集和分配,解决多无人机辅助边缘计算时能量短缺问题。

### 参考文献:

- [1] GUO Yongan, ZHU Hongbo, YANG Longxiang. Smart service system (SSS): A novel architecture enabling coordination of heterogeneous networking technologies and devices for Internet of Things[J]. China Communications, 2017, 14(3):130-144.
- [2] 虞湘宾,王光英,许方铖. 未来移动通信网络中移动边缘计算技术[J]. 南京航空航天大学学报, 2018, 50(5):586-594.  
YU Xiangbin, WANG Guangying, XU Fangcheng. Mobile edge computing technologies in future mobile communication networks[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2018, 50(5): 586-594.
- [3] WANG Qin, CHEN Ye, YIN Shu, et al. Pricing based power control for inter-cell UAV-assisted vehicle-to-vehicle underlay communication[J]. China Communications, 2019, 16(1):57-68.
- [4] 陈谋,吴颖. 具有输入时滞的集群无人机事件触发协同最优控制[J]. 南京航空航天大学学报, 2022, 54(5):958-968.  
CHEN Mou, WU Ying. Event-triggered cooperative optimal control of swarm UAVs with input delays[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2022, 54(5): 958-968.
- [5] WANG D, TIAN J, ZHANG H, et al. Task offloading and trajectory scheduling for UAV-enabled MEC networks: An optimal transport theory perspective[J]. IEEE Wireless Communications Letters, 2021, 11(1): 150-154.

- [6] DING R, GAO F, SHEN X S. 3D UAV trajectory design and frequency band allocation for energy-efficient and fair communication: A deep reinforcement learning approach[J]. *IEEE Transactions on Wireless Communications*, 2020, 19(12): 7796-7809.
- [7] LI M, CHENG N, GAO J, et al. Energy-efficient UAV-assisted mobile edge computing: Resource allocation and trajectory optimization[J]. *IEEE Transactions on Vehicular Technology*, 2020, 69(3): 3424-3438.
- [8] YANG L, YAO H, WANG J, et al. Multi-UAV-enabled load-balance mobile-edge computing for IoT networks[J]. *IEEE Internet of Things Journal*, 2020, 7(8): 6898-6908.
- [9] QI X, CHONG J, ZHANG Q, et al. Collaborative computation offloading in the multi-UAV fleet mobile edge computing network via connected dominating set[J]. *IEEE Transactions on Vehicular Technology*, 2022, 71(10): 10832-10848.
- [10] LIU Q, LIANG H, LUO R, et al. Energy-efficiency computation offloading strategy in UAV aided V2X network with integrated sensing and communication [J]. *IEEE Open Journal of the Communications Society*, 2022, 3: 1337-1346.
- [11] 黄旭民, 张旸, 余荣, 等. 基于 Stackelberg 博弈的无人机辅助无线供能物联网能量优化[J]. *通信学报*, 2022, 43(12): 146-156.  
HUANG Xumin, ZHANG Yang, YU Rong, et al. Stackelberg game based energy optimization for unmanned aerial vehicle assisted wireless-powered Internet of things[J]. *Journal of Communication*, 2022, 43(12): 146-156.
- [12] WANG Y, CHEN W, LUAN T H, et al. Task offloading for post-disaster rescue in unmanned aerial vehicles networks[J]. *IEEE/ACM Transactions on Networking*, 2022, 30(4): 1525-1539.
- [13] HUANG X, YANG X, CHEN Q, et al. Task offloading optimization for UAV-assisted fog-enabled Internet of Things networks[J]. *IEEE Internet of Things Journal*, 2021, 9(2): 1082-1094.
- [14] ZHAO N, YE Z, PEI Y, et al. Multi-agent deep reinforcement learning for task offloading in UAV-assisted mobile edge computing[J]. *IEEE Transactions on Wireless Communications*, 2022, 21(9): 6949-6960.
- [15] SEID A M, BOATENG G O, MARERI B, et al. Multi-agent DRL for task offloading and resource allocation in multi-UAV enabled IoT edge network [J]. *IEEE Transactions on Network and Service Management*, 2021, 18(4): 4531-4547.
- [16] REN T, NIU J, DAI B, et al. Enabling efficient scheduling in large-scale UAV-assisted mobile-edge computing via hierarchical reinforcement learning[J]. *IEEE Internet of Things Journal*, 2021, 9(10): 7095-7109.
- [17] CAI T, YANG Z, CHEN Y, et al. Cooperative data sensing and computation offloading in UAV-assisted crowdsensing with multi-agent deep reinforcement learning [J]. *IEEE Transactions on Network Science and Engineering*, 2021, 9(5): 3197-3211.
- [18] DING R, XU Y, GAO F, et al. Trajectory design and access control for air-ground coordinated communications system with multiagent deep reinforcement learning[J]. *IEEE Internet of Things Journal*, 2021, 9(8): 5785-5798.
- [19] CHEN J, CAO X, YANG P, et al. Deep reinforcement learning based resource allocation in multi-UAV-aided MEC networks[J]. *IEEE Transactions on Communications*, 2023, 71(1): 296-309.
- [20] QIN Z, LIU Z, HAN G, et al. Distributed UAV-BSs trajectory optimization for user-level fair communication service with multi-agent deep reinforcement learning[J]. *IEEE Transactions on Vehicular Technology*, 2021, 70(12): 12290-12301.
- [21] LIU C H, CHEN Z, TANG J, et al. Energy-efficient UAV control for effective and fair communication coverage: A deep reinforcement learning approach[J]. *IEEE Journal on Selected Areas in Communications*, 2018, 36(9): 2059-2070.
- [22] YANG C, LIU B, LI H, et al. Learning based channel allocation and task offloading in temporary UAV-assisted vehicular edge computing networks[J]. *IEEE Transactions on Vehicular Technology*, 2022, 71(9): 9884-9895.
- [23] CHALLITA U, SAAD W. Network formation in the sky: Unmanned aerial vehicles for multi-Hop wireless backhauling [C]//*Proceedings of 2017 IEEE Global Communications Conference (GLOBECOM)*. [S.l.]: IEEE, 2017: 749-754.
- [24] HOURANI A, KANDEEPAN S, JAMALIPOUR A. Modeling air-to-ground path loss for low altitude platforms in urban environments[C]//*Proceedings of GLOBECOM'14, Satellite & Space Communication*. [S.l.]:IEEE, 2014: 2898-2904.
- [25] JAIN R, CHIU D, HAWK W. A Quantitative measure of fairness and discrimination for resource allocation in shared computer systems[J/OL]. 1998: 1-38. <https://arxiv.org/abs/cs/9809099V1>.
- [26] WANG L, WANG K, PAN C, et al. Multi-agent deep reinforcement learning-based trajectory planning for multi-UAV assisted mobile edge computing[J]. *IEEE Transactions on Cognitive Communications and Networking*, 2021, 7(1): 73-84.