引用格式：巨涛, 李林娟, 张文金, 等. 多无人机辅助的移动边缘计算任务卸载及路径优化方法 [J]. 电子科技大学学报, 2025, 54(1): 72-83.
JU T, LI L J, ZHANG W J, et al. MATOPO: A multi-UAV assisted task offloading and path optimization method for moving edge computing[J]. Journal of University of Electronic Science and Technology of China, 2025, 54(1): 72-83.

多无人机辅助的移动边缘计算任务卸载及路径优化方法

巨 涛*，李林娟，张文金，张宇斐，火久元

（兰州交通大学 电子与信息工程学院，兰州 730070）

摘要：针对多无人机辅助移动边缘计算中的任务卸载决策和路径优化问题，提出了一种基于多智能体深度强化学习的计算任务卸载与路径优化方法，以降低系统总能耗，提升计算性能。首先，设计了多无人机辅助移动边缘计算系统模型，通过软件定义网络技术对无人机网络进行集中管理；然后，在考虑无人机负载及用户设备关联服务公平性的基础上，以系统总能耗为优化目标，通过设计多智能体深度确定性策略梯度算法完成任务卸载与无人机路径管理优化，以实现负载均衡、降低整个系统总能耗。仿真实验结果表明，与其他基准算法相比，所提方法在充分利用无人机辅助移动边缘计算系统计算资源的基础上，可在一定程度上降低系统能耗和计算延迟，保证整个系统的高效、稳定和可靠性，较好地满足移动边缘用户的服务请求。

关键词：移动边缘计算；多无人机网络；任务卸载；路径优化；多智能体深度强化学习
中图分类号：TP391 **文献标志码：**A DOI: 10.12178/1001-0548.2023276

MATOPO: A multi-UAV assisted task offloading and path optimization method for moving edge computing

JU Tao*, LI Linjuan, ZHANG Wenjing, ZHANG Yufei, and HUO Jiuyuan

(School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: Aiming at solving the task offloading and path planning challenge of multi-UAV assisted mobile edge computing, a multi-agent deep reinforcement learning method for task offloading and path optimization is proposed to reduce the total energy consumption of the system and improve computing performance. Firstly, the model of multi-UAV assisted mobile edge computing system is designed, and the UAV network is centrally managed by software-defined network technology. Then, on the basis of considering the load of the UAV and the fairness of the associated service of the user equipment, taking the total energy consumption of the system as the optimization goal, the multi-agent depth deterministic strategy gradient algorithm is designed to complete the task unloading and the path management optimization of the UAV, so as to achieve load balancing and reduce the total energy consumption of the whole system. Simulation results show that compared with other benchmark algorithms, the proposed method can reduce system energy consumption and computing delay to a certain extent, ensure the efficiency, stability and reliability of the whole system, and better meet the service requests of mobile edge users on the basis of making full use of the computing resources of UAV-assisted mobile edge computing systems.

Key words: mobile edge computing; multi-UAV network; task offloading; path optimization; multi-agent deep reinforcement learning

随着人工智能、物联网和 5G 通信的发展，许多延迟敏感且计算密集型的应用正在逐步扩展。移动终端设备因受限于自身计算能力，无法支持这类计算需求高且延迟敏感的应用[1]。移动边缘计算

收稿日期：2023−11−08
基金项目：国家自然科学基金（61862037, 62262038）；甘肃省科技计划（23CXGA0028）；兰州市人才创新创业项目（2021-RC-40）
作者简介：巨涛，博士，副教授，主要从事边缘计算、并行计算、深度学习及并行优化等方面的研究。
*通信作者 E-mail: jutao@mail.lzjtu.cn

（mobile edge computing, MEC）可用于解决这类计算任务与资源受限的用户设备之间的冲突。MEC 将服务器放置在移动网络边缘（蜂窝基站或 Wi-Fi 接入点），并提供计算和存储资源，从而可以更方便地提供计算服务来处理终端的密集计算任务，降低服务延迟，提高服务质量[2]。但是，具有计算服务器的固定基站可能会因为用户设备（user equpiments,UEs）移动，不能有效处理来自物联网设备的大量通信任务而出现暂时故障，或由于不可抗的自然灾害而损坏。因此，传统的 MEC 技术不能有效保证在上述特殊场景下所需的用户服务质量（quality of services, QoS）、网络可用性和可扩展性。

由于无人机（unmanned aerial vehicles, UAVs）具有自适应能力强、移动速度快、部署成本低等特点，一种新的 UAV 空中计算模式受到越来越多的关注。通过在 UAVs 上搭载移动边缘计算服务器或利用 UAVs 作为移动中继，为地面设备提供灵活的通信、计算和缓存服务[3]。与传统 MEC 相比，UAV 辅助 MEC 具有高移动性，当地面通信基础设施受到自然灾害的严重破坏时，UAVs 可以作为临时基站，将用户连接到骨干网，满足用户对计算资源的需求[4]。其次，利用空地链路的视距特性，其可以提供更高的数据速率，显著降低能耗和任务处理延迟，从而有效地保证用户服务质量（QoS）[5]。

1　相关工作

UAV 作为辅助移动边缘计算的移动边缘节点，已在学术界和工业界得到了大量研究。文献 [6-8] 在支持单 UAV 的 MEC 系统中采用传统凸优化或深度强化学习方法来完成任务卸载或 UAV 路径优化以降低时延或能耗。但是，由于单 UAV 资源和覆盖范围的限制，任务数量和计算需求的增加，单个 UAV 无法满足用户设备的资源需求。因此，具有更多资源和更大服务范围的多 UAV 辅助 MEC 系统引起了广泛关注。针对无人机辅助的应急通信网络，文献 [9] 提出了一种基于 Q-learning 和卷积神经网络的深度强化学习资源调度方法，以最大限度地提高频谱效率，然而最大限度地提高通信效率会限制搭载了边缘服务器的无人机服务范围，不能较好地满足服务公平性的需要。文献 [10] 针对无人机计算和通信资源有限，很难服务多种质量等级计算任务的问题，提出了一种可扩展的空中计算解决方案和基于协同计算的空中视频流解决方案。文献 [11] 将一组 UAVs 作为移动边缘服务器，提出一种基于李雅普诺夫的动态资源分配方案对 UAV 的资源进行动态分配，目的是最大限度地降低 UAV 辅助 MEC 系统的系统成本和最大化系统效用，但仅考虑了完整的计算卸载，即需要卸载的任务不能分区。文献 [12] 部署了多 UAVs 集群网络，提出了基于无模型深度强化学习的协同计算卸载与资源分配方案，每个智能体在网络中独立学习高效的计算卸载策略，目标是最小化任务执行延迟和能量消耗，但没有考虑对 UAV 飞行轨迹的优化。文献 [13] 通过对各 UAV 飞行轨迹的独立管理，共同优化了各 UEs 的地理公平性、各 UAV 的负载公平性以及无人机的总能耗。文献 [14] 研究了一种多 UAVs 的边缘云协同移动边缘计算系统，通过联合设计 UAV 的飞行轨迹、计算任务分配和通信资源管理，解决任务卸载问题，最大限度地减少执行延迟和能耗，但其未考虑无人机的负载公平性。文献 [15] 针对 MEC 网络中部署多架 UVAs 进行计算卸载时，缺乏灵活的学习方案来根据无人机的动态移动模式和无人机故障有效调整计算卸载策略的问题，提出了一种基于分布式深度强化学习的协同探索和优先经验重放方法。

上述多 UAVs 辅助移动边缘计算仍面临着许多亟待解决的问题。一方面，在多 UAVs 辅助移动边缘计算场景中，需要通过制定合理的任务卸载策略，以相对均衡的方式将用户的任务合理卸载到特定的 UAV，并为其分配所需的计算资源，实现 UAV 资源的高效利用，以提高系统的整体效率[16]。另一方面，由于地面用户设备的非均匀分布，会使得部分 UAV 承担大量任务，而其他 UAV 处于空闲状态，从而出现负载不均衡，任务最繁重的边缘节点会导致非常长的计算延迟[17]，使实时任务需求得不到满足，或部分 UAV 因长期高强度运行而损坏，因此必须对这些 UAV 的负载公平性进行有效衡量，以实现负载均衡。同时，UAVs 辅助 MEC 系统面临着能源限制的挑战，如由于尺寸和重量的限制，导致 UAV 的机载能量、续航力和计算性能有限，在为地面用户设备提供服务过程中，不合理的飞行轨迹会产生更多的能量消耗。因此，通过优化 UAV 飞行轨迹来降低能量消耗，是 UAV 辅助 MEC 系统的关键问题。

针对以上问题，本文设计了多 UAVs 辅助的 MEC 框架，通过软件定义网络（software defined network, SDN）技术对多 UAVs 网络进行集中管理。在该框架下，通过制定合理的无人机用户关联策略和卸载策略，使多架 UAVs 协同为地面 UE 提供计算任务卸载服务，实现了协作环境下的 UAVs 路径优化及动态任务卸载。主要通过优化每个 UAV 的飞行轨迹和任务卸载决策，最小化 UAV 网络和用户设备所产生的系统总能耗，通过 Jain 公平指数最大化每个 UAV 的负载公平性，保证整个计算系统中的负载均衡，提升整个边缘计算系统的整体计算效能。在以上框架的基础上，提出了一种多 UAV 协同任务卸载和路径优化方案 MATOPO，采用多智能体深度确定性策略梯度（multi-agent deep deterministic policy gradient, MADDPG）方法，在动态空到地（air-to-ground, A2G）网络环境下为多智能体寻找最优的任务卸载策略，通过集中训练和分散执行技术降低总体训练成本，从而实现长期奖励最大化[18]。

2 系统模型

2.1 网络模型

图 1 为本文提出的多无人机辅助移动边缘计算系统模型，该模型是由一个 SDN 控制器、多 UAV 网络和多个 UE 构成的多 UAV 辅助 MEC 系统。将具有多单元和多计算节点的 UAV 网络作为空中 MEC 服务器，代替受损基站为边缘用户设备提供计算卸载服务。

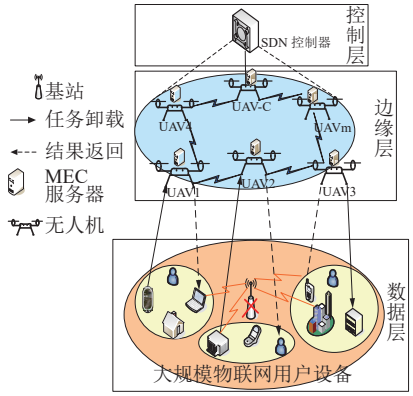

图 1 多无人机辅助移动边缘计算系统模型

多无人机的 MEC 架构因其高性能和可扩展性，可以提供更全面复杂的数据服务，在部署过程中通常将多架无人机集成为一个统一网络，使得无人机可以按照预先定义的通信协议相互通信，提供网络协同运行的能力。但这同时也给系统控制和数据传输组件带来了额外的负担。首先，基于多无人机的 MEC 网络是一种移动自组织网络，网络拓扑结构动态变化，难以监控。其次，基于传统网络架构的多无人机 MEC 网络具有分布式控制平面，从而导致不能提供整个系统的抽象视图，无法部署精确的网络功能。SDN 已成为一种高效的网络管理方法[19]，可以将 MEC 集成到软件定义的无人机网络框架中，实现多无人机 MEC 网络的协同运行。其中，SDN 将网络控制平台与网络功能组件分离开来，网络功能设备只负责执行数据转发等网络功能，所有的网络策略都在统一的网络控制平台上确定和编程，这样可以对 UAV 辅助的 MEC 网络进行集中的网络控制，使得网络管理员可以通过软件编程的方式快速配置和管理网络资源，让 MEC 的部署更加灵活和简便，能够更快地满足移动应用的需求。同时 SDN 可以实现网络流量优化和智能路径控制，从而提高网络性能和响应速度，这对于 UAV 的应用非常重要，因为 UAV 需要实时获取数据和控制信号，以确保其稳定和安全运行[20]。

为了实现多无人机 MEC 网络的协同网络功能，采用 SDN 对多无人机 MEC 网络的架构进行优化，将基于多无人机的 MEC 网络分为控制层、边缘层和数据层 3 个功能层。控制层包含 SDN 控制器，用于对数据层设备和边缘计算服务器进行集中管理。该层使用 OpenFlow 协议，通过部署流表来部署和执行网络功能（即数据传输功能），具体基于 OpenFlow 的流表机制，引入数据传递表，实现基于多 UAVs 的 MEC 网络的精确控制和协同功能。同时 SDN 控制器还可以操纵资源分配、计算卸载决策和其他决策。边缘层由 SDN 控制器控制，包括交换机/路由器、UAV 管理器（UAV-C）和计算 UAV。计算 UAV 主要负责处理任务，具有较好的计算性能，可视为计算服务器。UAV 管理器负责信息管理和控制，在计算能力上不如计算型 UAV，其主要用来收集地面用户的信息，并能够从 SDN 控制器转发任务卸载和资源分配相关决策。数据层具有超密集的异构物联网设备，由基本的终端用户智能设备组成。通过将 MEC 集成到软件定义的 UAV 网络中，采用 SDN 技术提高 UAV 辅助 MEC 网络的可扩展性和可编程性。本文使用的主要符号定义如表 1 所示。

表 1　主要符号定义列表

符号	定义
$\mathcal{N}, n, N$	UEs集合，UE索引，UE数量
$\mathcal{M}, m, M$	UAVs集合，UAV索引，UAV数量
$\mathcal{T}, t, T$	时隙集，时隙索引，时隙数
$\theta_h^m(t), \theta_v^m(t)$	UAV的水平偏转角、垂直偏转角
$\theta_c, v_m(t)$	UAV最大仰角，飞行速度
$\{X_m(t), Y_m(t), Z_m(t)\}$	UAV在时隙t的位置坐标
$\{x_m(t), y_m(t), 0\}$	UE在时隙t的位置坐标
$R_{\max}^m(t)$	UAV最大水平飞行半径
$R_{mm'}(t), R_u$	UAV间距离和最小保持距离
$L_{m,n}(t), d_{m,n}(t), p_{m,n}^{\mathrm{LOS}}(Z_m(t), d_{m,n}(t))$	路径损耗，UAV与UE之间视距链路距离和概率
α和β, c, f_c	环境类型参数，光速，载波频率
$\eta_{\mathrm{LOS}}, \eta_{\mathrm{NLOS}}$	LoS和非LoS连接对应的损耗
$B, p^{\mathrm{Tr}}, \sigma^2$	带宽，传输功率，噪声功率
$r_{m,n}(t), D_n(t)$	传输速率，任务数据量
$\varphi_{m,n}(t)$	任务卸载比率
$f_n(t)$	UE的计算资源
C_n	UE端的CPU电容系数
$F_n(t)$	UE处理1位数据所需CPU周期数
$f_{m,n}(t)$	UAV分配给UE的计算资源
C_m	UAV边缘服务器的CPU电容系数
$F_m(t)$	UAV处理1位数据所需CPU周期数
$c_m(t)$	UAV平均工作负载
$a_{m,n}(t)$	UAV与UE间关联变量
δ	UAV叶片局部截面阻力系数
ρ, m	空气密度，无人机有效载荷
g	重力加速度
c_T	UAV基于阀瓣面积的推力系数
A、c_s	UAV各转子盘面积和转子固度
c_f	UAV感应功率增量修正系数
τ_c, d_0	UAV爬升和各旋翼的机身阻力比
$s_m, a_m, R(t)$	状态空间，动作空间，奖励函数
$\eta 1$和$\eta 2, \gamma$	惩罚因子，奖励折扣因子
D	经验缓冲池
done, r	终止条件，奖励集合

2.2　无人机移动模型

假设有 N 个 UE 随机分布在边长为 $l_{\max}$ 的方形区域内，将 UE 集合记为$\mathcal{N} \triangleq \{n=1,2,\cdots,N\}$，在目标区域上空以高度 $Z_m(t)$ 飞行的多架无人机为地面 UEs 服务，UAVs 的集合记为$\mathcal{M} \triangleq \{m=1,2,\cdots,M\}$，其中每个 UAV 配备一个 MEC 服务器用来计算和通信，所有 UAVs 在 T 个时隙$\mathcal{T} \triangleq \{t=1,2,\cdots,T\}$内完成一次飞行任务，在连续的 T 个时隙中，UE 在每个时隙都有一个计算任务要执行，每个任务可以由 UE 执行，也可以卸载到其中一架 UAV 上。

在时隙 t 中，每个 UAV 的飞行方向由水平偏转角$\theta_h^m(t)$和垂直偏转角$\theta_v^m(t)$控制，飞行速度为$v_m(t)$，如图 2 所示，并且不能超出目标区域的边界。

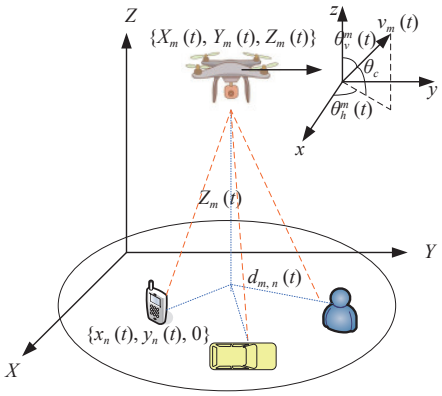

图 2　无人机移动性示意图

用三维笛卡尔坐标描述 UAV 和 UE 的位置，设无人机 m 的初始坐标为$q_{\mathrm{uav}}^m = \{X_m(0), Y_m(0), Z_m(0)\}$，则在时隙 t 无人机 m 的位置为$q_{\mathrm{uav}}^m(t) = \{X_m(t), Y_m(t), Z_m(t)\}$，UEs 在地面上匀速移动，$t$ 时隙用户 n 的位置为$q_{\mathrm{ue}}^n(t) = \{x_n(t), y_n(t), 0\}$。

其中：

$$X_m(t) = X_m(0) + v_m(t)\sin\theta_v^m(t)\cos\theta_h^m(t) \quad \forall m \in \mathcal{M}, t \in \mathcal{T} \tag{1}$$

$$Y_m(t) = Y_m(0) + v_m(t)\sin\theta_v^m(t)\sin\theta_A^m(t) \quad \forall m \in \mathcal{M}, t \in \mathcal{T} \tag{2}$$

$$Z_m(t) = Z_m(0) + v_m(t)\cos\theta_v^m(t) \quad \forall m \in \mathcal{M}, t \in \mathcal{T} \tag{3}$$

假设 UAV 有最大仰角 θ_c，在 t 时隙第 m 个无人机的最大水平飞行半径为：$R_{\max}^m(t) = Z_m(t)\tan\theta_c$。

时隙 t 无人机之间的距离为$R_{m,m'}(t)$，表示如下：

$$R_{m,m'}(t) = \sqrt{(X_m(t) - X_{m'}(t))^2 - (Y_m(t) - Y_{m'}(t))^2} \quad \forall m \in \mathcal{M}, t \in \mathcal{T} \tag{4}$$

UAV 在每个时隙 t 中应保持最小距离R_u(R_u= 1 m) 以避免碰撞，无人机之间的距离应满足如下条件：$R_{m,m'}(t) \geqslant R^u, \forall m, m' \in M, m \neq m'$。

2.3　通信传输延迟和能耗模型

空对地（A2G）信道具有更高的视距连通性，考虑了不同传播环境下视距链路和非视距链路出现的概率，采用文献 [21] 中的概率路径损失模型进行通信的传输延迟和能耗建模。在时隙 t 中，第 n 个 UE 与第 m 个 UAV 之间的距离为：

$$d_{m,n}(t) = \sqrt{(X_m(t) - x_n(t))^2 + (Y_m(t) - y_n(t))^2 + Z_m^2(t)} \quad \forall m \in \mathcal{M}, n \in \mathcal{N}, t \in \mathcal{T} \tag{5}$$

则第 t 个时隙 UAV 和 UE 之间的视距链路概率为：

$$p_{m,n}^{\mathrm{LOS}}(Z_m(t),d_{m,n}(t))=\frac{1}{1+\alpha\exp\left(-\beta\left(\arctan\left(\frac{H}{d_{m,n}(t)}\right)-\alpha\right)\right)} \tag{6}$$

式中，α 和 β 取决于传播环境的类型，在此设置中，忽略用户设备的高度以及用户设备和无人机的天线高度，路径损耗表达式为：

$$L_{m,n}(t)=20\log\left(\frac{4\pi f_c}{c}\right)+20\log(d_{m,n}(t))+p_{m,n}^{\mathrm{LOS}}(t)\eta_{\mathrm{LOS}}+\left(1-p_{m,n}^{\mathrm{LOS}}(t)\right)\eta_{\mathrm{NLOS}} \tag{7}$$

式中，c 为光速（m/s）；f_c 为载波频率（HZ）；η_{LOS} 和 η_{NLOS}（dB）分别为 LOS 和非 LOS 连接对应的损耗。

因此，用户设备与无人机之间的传输速率为：

$$r_{m,n}(t)=B\log_2\left(1+\frac{p^{\mathrm{Tr}}}{\sigma^2}10^{-\frac{L_{m,n}(t)}{10}}\right) \tag{8}$$

式中，B 表示分配给 UE 的带宽；p^{Tr} 表示用户设备的传输功率；σ^2 表示噪声功率。

在时隙 t 中，假设第 n 个 UE 有一个计算密集型任务 $T_n(t)$ 要执行，定义为：

$$T_n(t)=\{D_n(t),F_n(t)\}\quad \forall n\in\mathcal{N},t\in\mathcal{T} \tag{9}$$

式中，$D_n(t)$ 表示需要处理的数据量；$F_n(t)$ 表示第 n 个 UE 处理 1 位数据所需 CPU 周期数。

定义$\varphi_{m,n}(t)\in[0,1]$ 为第 m 个 UE 将任务卸载至无人机时的比率，$\varphi_{m,n}(t)$=0 表示任务在本地执行，即所有任务在用户设备上处理；$\varphi_{m,n}(t)$=1 表示全部卸载，即任务全部卸载至无人机端处理；$\varphi_{m,n}(t)\in$(0,1) 表示部分卸载，即根据卸载比例，将一部分任务卸载到无人机端执行，剩余任务在本地执行。

因此，第 n 个 UE 与第 m 个 UAV 通信的传输延迟和能耗为：

$$T_{m,n}^{\mathrm{Tr}}(t)=\frac{\varphi_{m,n}(t)D_n(t)}{r_{m,n}(t)} \tag{10}$$

$$E_{m,n}^{\mathrm{Tr}}(t)=p^{\mathrm{Tr}}T_{m,n}^{\mathrm{Tr}}(t)=p^{\mathrm{Tr}}\frac{\varphi_{m,n}(t)D_n(t)}{r_{m,n}(t)} \tag{11}$$

2.4 计算模型

计算模型由卸载策略决定，当$\varphi_{m,n}(t)$=0 时，任务可全部卸载到用户设备上进行处理，期间所产生的时延和能耗如下：

$$T_n^{\mathrm{Com}}(t)=\frac{(1-\varphi_{m,n}(t))D_n(t)F_n(t)}{f_n(t)} \tag{12}$$

$$E_n^{\mathrm{com}}(t)=C_n(f_n(t))^3T_n^{\mathrm{Com}}(t)=C_n(1-\varphi_{m,n}(t))D_n(t)F_n(t)(f_n(t))^2 \tag{13}$$

式中，$f_n(t)$ 为第 n 个 UE 的计算资源；C_n 表示 UE 端的 CPU 电容系数；$F_n(t)$ 表示第 n 个 UE 处理 1 位数据所需的 CPU 周期数。

当任务卸载至无人机端处理时，产生的时延和能耗如下：

$$T_{m,n}^{\mathrm{Com}}(t)=\frac{\varphi_{m,n}(t)D_n(t)F_m(t)}{f_{m,n}(t)} \tag{14}$$

$$E_{m,n}^{\mathrm{Com}}(t)=C_m(f_{m,n}(f))^3T_n^{\mathrm{Com}}(t)=C_m(1-\varphi_{m,n}(t))D_n(t)F_n(t)(f_{m,n}(t))^2 \tag{15}$$

式中，$F_m(t)$ 表示第 m 个 UAV 处理 1 位数据所需的 CPU 周期数；$f_{m,n}(t)$ 表示第 m 个 UAV 分配给第 n 个 UE 的计算资源；C_m 表示 UAV 边缘服务器的 CPU 电容系数。

在系统运行过程中，可能会导致服务的不公平性，一些无人机可能会比其他无人机为更多的 UE 提供服务。为了解决这个问题，根据 Jain 公平方程[22]，提出了一个公平指数$f^u(t)$：

$$f^u(t)=\frac{\left(\sum_{m=-1}^{M}\sum_{t=1}^{T}c_m(t)\right)^2}{M\sum_{m=1}^{M}\left(\sum_{t=1}^{T}c_m(t)\right)^2} \tag{16}$$

式中，$c_m(t)$ 为时隙 t 连接 UEs 的第 m 个 UAV 的平均工作负载，具体计算公式如下：

$$C_m(t)=\frac{\sum_{n=1}^{N}\varphi_{m,n}(t)}{N}\quad \forall m\in\mathcal{M},t\in\mathcal{T} \tag{17}$$

公平指数$f^u(t)$反映了无人机之间的物理公平性，如果所有无人机平均工作负载相等，则$f^u(t)$值接近 1。

同时，定义关联变量$\alpha_{m,n}(t)\in\{0,1\}$表示 UE 是否与 UAV 建立连接产生关联。当$\alpha_{m,n}(t)$=1 时表示在时隙 t 第 n 个 UE 与第 m 个无人机产生关联，当$\alpha_{m,n}(t)$=0 时表示任务在本地执行，即与无人机未产生关联。

在给定的时隙，由于不同的因素，UAVs 可能无法访问所有联网设备。为了避免某些 UE 占用过

多时隙获得服务，而其他 UE 得不到服务的情况，定义 UE 之间的公平性系数$f^e(t)$如下：

$$f^e(t)=\frac{\left(\sum_{n=1}^{N}\sum_{t=1}^{T}a_{m,n}(t)\right)^2}{\sum_{n=1}^{N}\left(\sum_{t=1}^{T}a_{m,n}(t)\right)^2} \tag{18}$$

公平性系数$f^u(t)$显式地反映了 UE 之间的公平性，如果所有 UEs 与 UAV 产生关联被 UAV 服务的次数相似，则$f^u(t)$的值更接近于 1。

2.5　四旋翼无人机能量消耗模型

无人飞行器的能量消耗由两部分组成：一部分用于通信，另一部分用于产生推力以帮助无人飞行器克服阻力和重力。在实践中，通信的能量通常比飞行的能量小两个数量级[23]。因此，本文忽略了通信能耗。其中，UAV 推进功率[7] 可以表示为：

$$P_m^{\mathrm{fly}}(t)=n\left[\frac{\delta}{8}\left(\frac{T}{c_T\rho A}+3v^2\right)\sqrt{\frac{T_h\rho c_s^2 A}{c_T}}+(1+c_f)F\left(\sqrt{\frac{T_h^2}{4\rho^2A^2}+\frac{v^2}{4}}\right)^{\frac{1}{2}}+\frac{mgv}{n}\sin\tau_c+\frac{1}{2}d_0v^3\rho c_sA\right] \tag{19}$$

式中，δ 为叶片局部截面阻力系数；c_T 为基于阀瓣面积的推力系数；ρ 为空气密度；A、c_s 分别为各转子盘面积和转子固度；c_f 为感应功率增量修正系数；τ_c 表示爬升角；d_0 为各旋翼的机身阻力比；m 为无人机有效载荷；g 为重力加速度。

则第 m 架无人飞机在时隙 t 的飞行能耗为：

$$E_m^{\mathrm{fly}}(t)=P_m^{\mathrm{fly}}(t)T_m^{\mathrm{fly}}(t)=P_m^{\mathrm{fly}}(t)\max\left\{\left[\max\left\{T_{m,n}^{\mathrm{Tr}}(t)\right\}+T_{m,n}^{\mathrm{Com}}(t)\right],\max\left\{T_n^{\mathrm{Com}}(t)\right\}\right. \tag{20}$$

$T_m^{\mathrm{fly}}(t)$为第 m 个 UAV 完成时隙 t 内所有连接的 UE 任务所需的时间，其为 UE 侧处理任务的最大时间与传输任务的最大时间加上无人机侧处理任务的总时间之间的较大值。

2.6　优化目标

本文优化目标是最大限度地提高每个 UAV 的 UE 负载的公平性，以及每个 UE 在所有时隙上被 UAV 服务的公平性，通过联合优化卸载策略Ψ和无人机轨迹 P 最大限度地降低系统的总体能耗。

首先，无人机和 UEs 在时隙 t 的系统总能耗如下：

$$E(t)=\frac{1}{f^e(t)f^u(t)}\left[\sum_{m=1}^{M}\sum_{n=1}^{N}(E_{m,n}^{\mathrm{tr}}(t)+E_n^{\mathrm{com}}(t)+E_{m,n}^{\mathrm{com}}(t))+\sum_{m=1}^{M}E_m^{\mathrm{fly}}(t)\right]$$
$$\forall m\in\mathcal{M},n\in\mathcal{N},t\in\mathcal{T} \tag{21}$$

然后，将优化问题表述为：

$$\mathcal{P}1:\min_{\mathcal{P},\Psi}\sum_{t=1}^{T}E(t) \tag{22}$$

$$\mathrm{C1}:0\leqslant\varphi_{m,n}(t)\leqslant1\quad\forall m\in\mathcal{M},n\in\mathcal{N},t\in\mathcal{T} \tag{22a}$$

$$\mathrm{C2}:\sum_{m=0}^{M}\alpha_{m,n}(t)=1\quad\forall n\in\mathcal{N},t\in\mathcal{T} \tag{22b}$$

$$\mathrm{C3}:0\leqslant X_m(t),x_n(t)\leqslant l_{\max}\quad\forall m\in\mathcal{M},n\in\mathcal{N},t\in\mathcal{T} \tag{22c}$$

$$\mathrm{C4}:0\leqslant Y_m(t),y_n(t)\leqslant l_{\max}\quad\forall m\in\mathcal{M},n\in\mathcal{N},t\in\mathcal{T} \tag{22d}$$

$$\mathrm{C5}:Z_{\min}\leqslant Z_m(t)\leqslant Z_{\max}\quad\forall m\in\mathcal{M},t\in\mathcal{T} \tag{22e}$$

$$\mathrm{C6}:0\leqslant\theta_\mu^m(t),\theta_v^m(t)\leqslant\pi\quad\forall m\in\mathcal{M},t\in\mathcal{T} \tag{22f}$$

$$\mathrm{C7}:v_{\min}\leqslant v_m(t)\leqslant v_{\max}\quad\forall m\in\mathcal{M},t\in\mathcal{T} \tag{22g}$$

$$\mathrm{C8}:R_{m,m'}(t)\geqslant R^u\quad\forall m,m'\in\mathcal{M},m\neq m' \tag{22h}$$

$$\mathrm{C9}:0\leqslant f^u(t),f^e(t)\leqslant1\quad\forall t\in\mathcal{T} \tag{22i}$$

式中，$\mathcal{P}=\left\{\theta_\mu^m(t),\theta_v^m(t),v_m(t)\ \forall m\in\mathcal{M},\mathcal{T}\in T\right\}$，$\Psi=\{\varphi_{m,n}(t)\ \forall m\in\mathcal{M},n\in\mathcal{N},t\in\mathcal{T}\}$。目标函数 (22) 保证 UAV 完成一次飞行的系统总能耗最小。约束 (22a) 是卸载策略的可选择范围，明确任务是否被卸载，卸载多少任务量。约束 (22b) 假设每个 UE 最多只能与一架无人机产生关联，约束 (22c)、(22d)、(22e) 为无人机和用户设备位置的取值范围，该约束使无人机和用户设备在一定范围内移动。约束条件 (22f) 为无人机飞行角度取值范围。约束条件 (22g) 为速度标量的有效取值范围。约束 (22h) 表示无人机之间应保持最小间隔距离避免发生碰撞。(22i) 是公平性指数范围，越接近 1 表示越公平。

3　算法设计与实现

由于多 UAVs 辅助 MEC 系统具有高维的动作和状态空间，传统算法很难得到最优解，而多智能体深度确定性策略梯度（multi-agent deep deterministic policy gradient, MADDPG）算法是解决多智能体连续动作的有效方法。同时由于系统总

成本由当前系统环境状态和所有 UAVs 的联合行动决定，而且前一个状态和动作会共同触发系统环境进入一个新的随机状态[24]。所以，本文首先将优化目标表述为一个多智能体马尔可夫决策过程（markov decision process, MDP），然后，提出了一种基于 MADDPG 算法的任务卸载和轨迹控制优化策略 MATOPO 方法求解问题。

3.1 马尔可夫决策过程构建

将每个 UAV 视为一个智能体，定义一个 MDP，将环境模型描述为（S, A, P, R）。

状态空间 S：由 UAV 和 UE 的状态组成。在不同的时间段，上述状态都在变化，这意味着 UE 在移动并产生新的任务。则第 m 个智能体状态空间可以表示为：

$$s_m=\{q^m_{\text{uav}}(t), q^n_{\text{ue}}(t), D_n(t), F_n(t), \forall n\in\mathcal{N}, t\in\mathcal{T}, m\in\mathcal{M}\} \quad (23)$$

动作空间 A：将 UAV 的飞行速度、水平偏转角、垂直偏转角和任务卸载率定义为第 m 个智能体的动作，记为：

$$a_m=\left\{v_m(t), \theta^m_h(t), \theta^m_v(t), \varphi_{m,n}(t), \forall t\in\mathcal{T}, m\in\mathcal{M}, n\in\mathcal{N}\right\} \quad (24)$$

状态转移概率 P：$T=\{p(s'\mid s,a), \forall s,s'\in S, a\in A\}$ 表示通过动作 $a=[a_1,a_2,\cdots,a_M]$ 从当前状态 $s=[s_1,s_2,\cdots,s_M]$ 到下一状态 $s'=\left[s'_1,s'_2,\cdots,s'_M\right]$ 的转移概率。

奖励函数 R：为了解决拟定的任务卸载和路径优化问题，M 个智能体在满足一定的约束条件（如范围约束和碰撞约束）的情况下，协同最小化系统总成本。奖励函数 $R(t)$ 定义为系统成本 $E(t)$ 的负值。如果不满足相应的约束，就在奖励函数 $R(t)$ 中给予其相应的惩罚，如下式所示：

$$R(t)=-\sum_{t=1}^{T}E(t)-\eta 1-\eta 2 \quad (25)$$

式中，$\eta 1$ 和 $\eta 2$ 分别表示与范围约束和碰撞约束相关的惩罚。如果任意 UAVs 飞出规定边界范围，违反范围约束 (22c)、(22d)、(22e)，则在奖励函数中给予其一个惩罚 $\eta 1$。此外，如果任意两个 UAVs 之间的距离不满足碰撞约束 (23h)，则两个 UAVs 的奖励函数中给予惩罚 $\eta 2$。

3.2 多智能体深度强化学习算法

MADDPG 算法是多智能体深度强化学习方案中的一种[25]，算法分为两个阶段，即集中式学习（训练）阶段和分布式执行阶段。在集中式的离线训练阶段，除了智能体 m 通过局部观察得到的状态信息 s_m 和当前执行动作 α_m 外，需要引入其他智能体的动作信息和状态信息，联合存储到当前智能体的经验池中用于集中式训练 Critic 网络。在分布式执行阶段，由于 Actor 网络只需要局部观察，所以每个智能体都可以在不了解其他智能体环境信息的情况下获得其动作[26]。

用 $\mu=[\mu_1,\mu_2,\cdots,\mu_M]$ 表示 Actor 策略网络中 M 个 agent 的确定性策略，$\mu'=[\mu'_1,\mu'_2,\cdots,\mu'_M]$ 表示目标策略网络中 M 个 agent 的确定性策略。用 $\theta=[\theta_1,\theta_2,\cdots,\theta_M]$ 表示 Actor 策略网络中确定性策略的参数。

第 m 个智能体的累积期望奖励为：

$$J(\theta_m)=\mathrm{E}_{s,a\sim D}\left[\sum_{t=1}^{T}\gamma r_{m,t}\right] \quad (26)$$

式中，D 表示经验缓冲区，存放着所有智能体的经验 $\{s, a, r, s', \text{done}\}$，$r=[r_1,r_2,\cdots,r_M]$ 为所有 agent 的奖励集合，done 为终止条件（UAVs 到达终点或飞出边界）；γ 为奖励折扣因子。

为了稳定训练过程，提高样本效率，每个智能体将当前经验 $\{s_m, a_m, r_m, s'_m, \text{done}_m\}$ 存储在经验缓冲区 D 中。在训练过程中从 D 中随机抽取小批量数据，然后将 s_m 输入 Actor 策略网络中生成策略 $\mu_m(s_m)$，使用策略梯度更新 Actor 策略网络的权值。对于确定性策略 μ，策略梯度为：

$$\nabla_{\theta_m}J(\theta_m)=\mathrm{E}_{s,a\sim D}\left[\nabla_{\theta_m}\mu_m(a_m|s_m)\nabla_{a_m}Q^{\mu}_m(s,a)|_{a_m=\mu_m(s_m)}\right] \quad (27)$$

式中，Q^{μ}_m 表示 Critic 网络基于所有代理的状态和动作作为输入的集中动作值函数输出，用于评估 Actor 网络的输出策略。通过最小化损失函数来更新 Critic 策略网络 Q^{μ}_m，如式 (28) 所示：

$$L(\theta_m)=\mathrm{E}_{s,a\sim D}\left[(Q^{\mu}_m(s_m,a_m)-y_m)^2\right] \quad (28)$$

目标值 y_m 为：

$$y_m=\sum_{t=1}^{T}r_{m,t}+(1-\text{done}_m)\gamma Q^{\mu'}_m(s',a')|_{a'_m=\mu'_m(s'_m)} \quad (29)$$

式中，$\alpha'=[\mu'_1(s'_1), \mu'_2(s'_2), \cdots, \mu'_M(s'_M)]$ 为 M 个 agent 的动作集合；$Q^{\mu'}_m$ 表示具有延迟更新参数 $\theta'=[\theta'_1, \theta'_2, \cdots, \theta'_M]$ 的确定性策略集合 μ' 的目标网络。

延迟参数 θ' 可通过式 (30) 更新，其中 τ 为软更新参数：

$$\theta'_m\leftarrow\tau\theta_m+(1-\tau)\theta'_m \quad (30)$$

3.3　MATOPO 算法框架设计

为了解决上述多智能体的 MDP 问题，考虑到任务卸载和路径优化问题的高维连续动作空间，基于 MADDPG 算法，设计了本文所提系统场景下的任务卸载与路径优化算法框架 MATOPO，算法整体框架如图 3 所示。其中假设智能体与环境交互，SDN 控制中心掌握系统全局信息。Actor 网络用于生成动作，其中包括每架 UAV 的飞行方向和速度。Critic 网络用于获取动作的 Q 值，对 Actor 网络生成的动作进行评估。在每个时隙中，智能体将 Actor 网络生成的动作发送给每个无人机。然后，每个 UE 尝试在其覆盖范围内关联一架 UAV，如果最小卸载能耗大于本地执行的能耗，则终端将决定在本地执行任务。

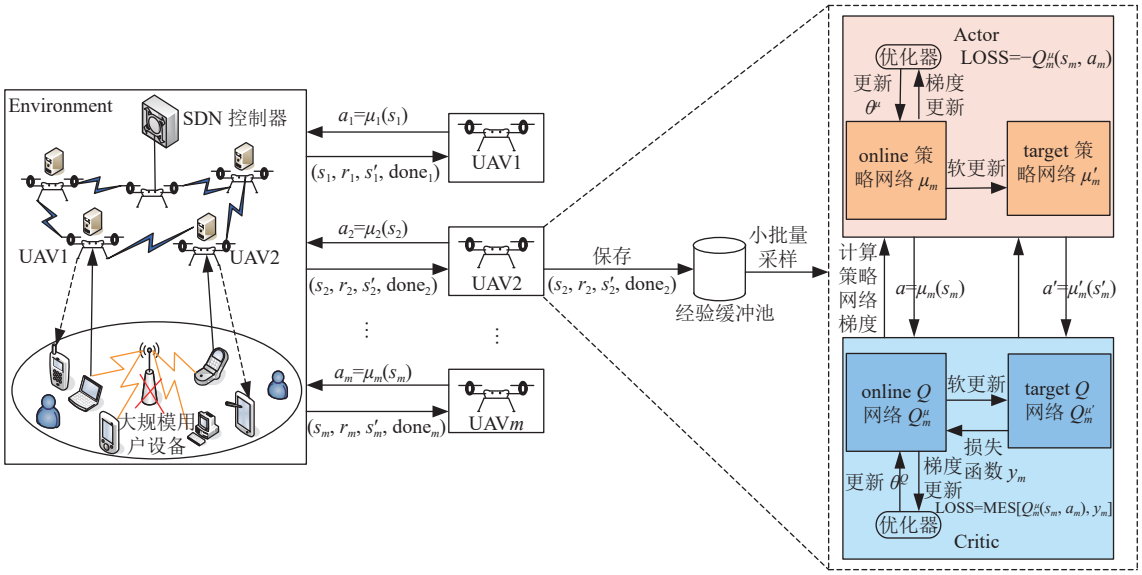

图 3　MATOPO 算法框架

在以上算法框架的基础上，设计了基于 MADDPG 的多 UAVs 辅助 MEC 任务卸载与路径优化算法 MATOPO，具体如算法 1 所示。

算法 1　基于 MADDPG 的多 UAVs 辅助 MEC 任务卸载与路径优化算法 MATOPO

初始化 Actor 策略网络 μ 及目标策略网络 μ' 的权重 θ 和 θ'、经验缓冲池 D、随机噪声 N_i，step=0

for episode=1 to MaxEpisode do:

初始化观测状态 s；

while done=False do:

step=step+1;

选择动作 $a=\mu_m(s_m;\theta_m)+N_i$；

for UAV $m=1, 2, \cdots, M$ do:

输入状态 s 和动作 a 到环境中；

执行算法 2 生成 UE 与 UAV 关联策略；

end for

执行动作 a；

获得奖励 r 和下一时刻状态 s'；

if 经验缓冲池 D 未满：

在 D 中存储状态转移信息 $(s, a, r, s', \text{done})$；

else

更新经验缓存池 D；

从 D 中随机抽取小批量训练数据；

根据式 (29) 获取目标值 y_m；

根据式 (28) 最小化损失函数值更新 Critic 网络参数；

用式 (27) 中梯度下降法更新 Actor 网络参数；

使用式 (30) 更新目标网络参数更新速率 τ；

end if

for UAV $m=1, 2, \cdots, M$ do:

if UAV m 完成飞行周期 then

done_m=True

奖励值为 $R(t)=-\sum_{t=1}^{\text{step}} E(t)$

end if

else if UAV m 飞出边界或发生碰撞 then

done_m=True

给予相应惩罚 $R(t)=-\sum_{t=1}^{\text{step}} E(t)-\eta 1-\eta 2$

end if
end for
end
end for

3.4 用户与无人机选择关联算法

设计 UE 与 UAV 间的选择关联算法，以决定用户与无人机之间的选择关联[27]，具体如算法 2 所示。

算法 2 用户无人机关联算法

初始化 A 和 E_m
for UAV $m=1,2,\cdots,M$ do
  for UE $n=1,2,\cdots,N$ do
    计算传输能耗$E_{m,n}^{\mathrm{Tr}}(t)$和本地计算能耗$E_n^{\mathrm{Com}}(t)$
    if $E_n^{\mathrm{Com}}(t)>E_{m,n}^{\mathrm{Tr}}(t)$ then
      将第 n 个 UE 存储在 E_m 中
    end if
  end for
  对 E_m 中$E_n^{\mathrm{Com}}(t)$与$E_{m,n}^{\mathrm{Tr}}(t)$的差值进行降序排序
end for
repeat
for UAV $m=1,2,\cdots,M$ do
  n=GetTopItem(E_m)
  if $E_{m,n}^{\mathrm{Tr}}(t)<E_{m,A(n)}^{\mathrm{Tr}}(t)$或 $A(n)=0$ then
    $A(n)=m$
  end if
  RemoveTopItem(E_m)
end for
直到列表 E_m 中的所有 UE 均被检查
return A

4 仿真实验与结果分析

4.1 参数设置及评测方法

1）仿真参数设置

10 个 UEs 和 3 架 UAVs 在 [1 000 m, 1 000 m] 的水平范围内移动，UAVs 的飞行高度范围为 [0 m, 500 m]，UAVs 的飞行速度 $v_m(t)$ 范围为 [30 m/s, 50 m/s]，垂直偏角$\theta_v^m(t)$被限制在 [0, π]，水平偏转角$\theta_h^m(t)$范围分别为 [0, π/2]、[0, π] 和 [π/2, π]。所有 UAVs 都会根据与 UEs 通信的 LoS 和 NLoS 等因素适当调整飞行高度。UEs 的位置在每个时隙内移动±50 m，并且每个 UEs 的任务也同步更新。

MATOPO 算法网络超参数为：Actor-Critic 网络学习率 lr=0.000 1，奖励折扣因子 γ=0.9，软更新系数 τ=0.01，随机抽取的数据批量大小 batch size = 512。参数设置主要参考文献 [5, 7, 25, 28] 中的参数进行配置，通信、计算和无人机飞行相关仿真参数见表 2。

表 2 仿真参数

参数	值	参数	值
$D_n(t)$/Mbits	[1,10]	C_n	10^{-27}
$F_n(t)$/cycles·bit^{-1}	800	$f_{m,n}(t)$/GHz	5
α，β	12.08, 0.11	C_m	10^{-28}
c/m·s^{-1}	3×10^8	$F_m(t)$/cycles·bit^{-1}	1 000
f_c/GHz	2.5	δ	0.012
η_{LOS}/dB	1.6	c_T	0.302
η_{NLOS}/dB	23	A/m^2	0.031 4
B/MHZ	1	c_s	0.095 5
p^{Tr}/W	0.1	ρ/Kg·m^{-3}	1.225
σ^2/dBm·Hz^{-1}	−70	m/kg	2.0
$f_n(t)$/GHz	1	g/m·s^{-2}	9.8

2）评测方法

以系统总能量消耗为评价指标，将所提方法与平均卸载和随机卸载两种基线方法进行对比，以验证本文方法的有效性。

平均卸载：将计算任务平均分配给 UEs 和 UAVs，一半任务在本地执行，一半任务卸载至 UAV 边缘端执行，此时卸载率Φ=0.5。

随机卸载：任务卸载率为一个随机数，在 [0,1] 范围内随机生成。

4.2 实验结果分析

1）MATOPO 收敛性能分析

图 4 比较了不同折扣因子 γ 对于算法性能的影响，γ 分别取 0.70、0.90、0.95 和 0.99，适当的 γ 值将提高训练策略的最终性能。如图 4 所示，当 γ=0.99、0.95 和 0.70 时曲线的振荡幅度较大且难以稳定收敛至最佳值，而当 γ=0.90 时，曲线的变化趋势较为平缓，且经过训练的计算卸载策略具有最佳性能，可以获得最优奖励收敛值，所以本文算法将 γ 值设定为 0.90。

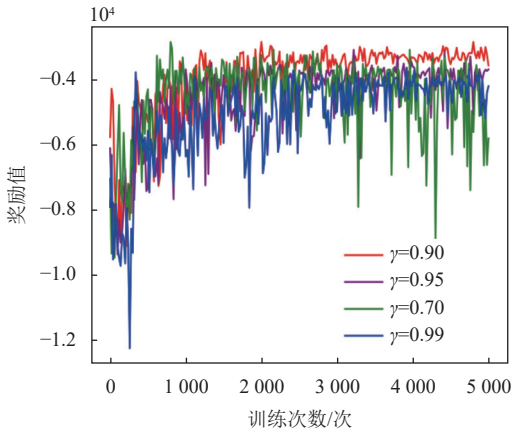

图 4 不同折扣因子下 MATOPO 收敛性能

图 5 展示了本文所提 MATOPO 算法的收敛性，如图所示，奖励值随着训练次数的增加而增加，MATOPO 算法在 2 000 次训练后开始逐步收敛，由于奖励函数设置为 UAVs 总能量消耗，所以此时 UAVs 所消耗的能量最少。

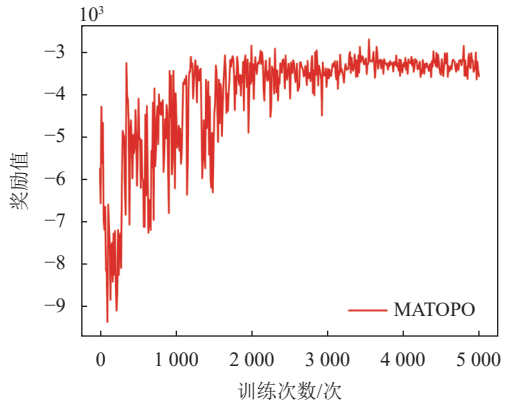

图 5　MATOPO 的收敛性能

2）负载公平性分析

图 6 展示了 UEs 数量对 UAVs 工作负载公平性指数的影响。由图可以观察到随着 UEs 数量的增加，平均卸载方式和随机卸载方式下，UAVs 的公平指数单调下降，而 MATOPO 算法则保持较高的公平指数。这是因为 MATOPO 可以针对 UEs 和 UAVs 的不同状态，选择合理的卸载策略，从而保证 UAVs 较高的公平性和较小的能耗，因此，随着 UEs 数量的增加，MATOPO 的公平指数不会明显下降。

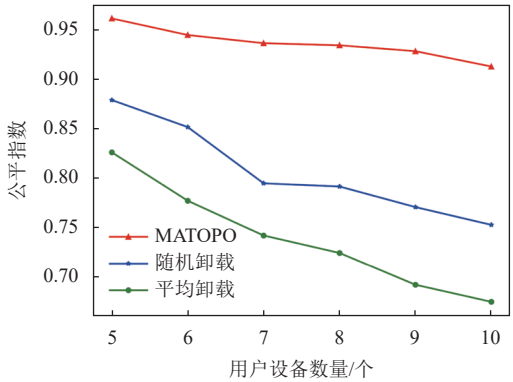

图 6　无人机负载公平性

以 UE 间的公平性指数为标准，将本文使用的用户无人机选择关联算法与距离最近选择关联和随机选择关联两种关联方式进行对比。最近关联：UEs 均选择与它们之间距离最近的 UAV 进行关联。随机关联：所有 UEs 都在 3 个 UAVs 中随机选择一个 UAV 进行任务卸载。

图 7 显示了 UE 在不同选择关联方式下的公平性比较，可以发现基于 MATOPO 的 UEs 公平性收敛到 0.92 左右，说明 UEs 对 UAVs 的选择关联已经达到了较为公平的状态。同时，由于使用均匀随机分布，当 UEs 随机选择 UAVs 时，UEs 被关联服务的次数可以保持稳定状态，从而可以保持稳定的公平性，但此时公平性不是最优的。当采用最近选择关联时，可能会出现 UAVs 集中在某个 UE 附近，从而导致某个 UE 与 UAV 多次产生关联被多次服务，而其他 UE 未被关联服务，UEs 的公平性较差，处于较低水平。

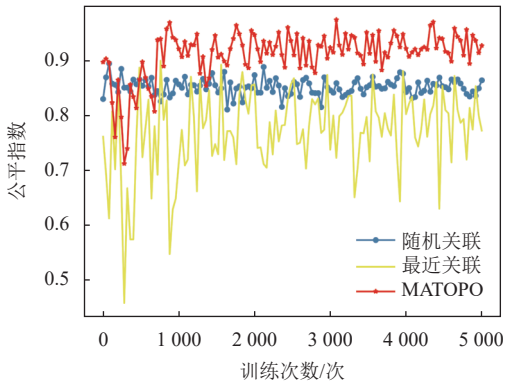

图 7　用户设备的公平性指数

3）系统总能量消耗结果分析

图 8 中给出了每个智能体（每个 UAV）单独产生的能量消耗以及系统总能耗。如图所示，UAV1 在达到平衡后产生的能量消耗较小，因为其主要是管理 UAV，不过多地参与任务的计算处理过程。而 UAV0 和 UAV2 能耗相似，因为由于 UAV0 和 UAV2 飞行距离相似，所以它们产生的能耗也大致相似。经过大约 2 000 次的训练，所有智能体的奖励趋于收敛，此时所有智能体所产生的能耗之和为系统总能耗。

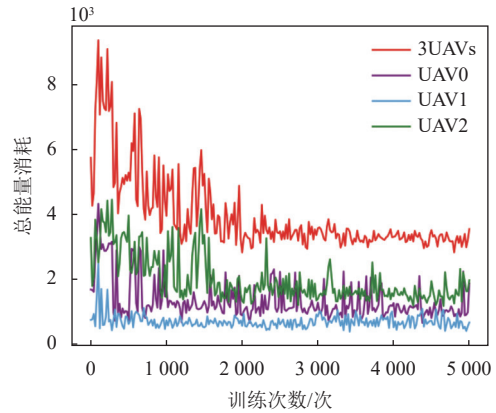

图 8　3 个无人机的能量消耗

图 9 分析了不同 UAV 数量下 MATOPO、随机卸载和平均卸载的能量消耗。为了公平起见，3 种方法中 UEs 选择关联 UAVs 的方式一致。从图 9 中可以看出，随着无人机数量的增加，系统复杂度上升，这是因为部署了更多的无人机，产生了更多的飞行和通信开销，从而使 3 种方法下的系统能耗均呈上升趋势。但相较于其他两种方式，由于 MATOPO 在执行过程中会学习获取合理的卸载策略，因而产生较低的能耗，获得最佳性能。

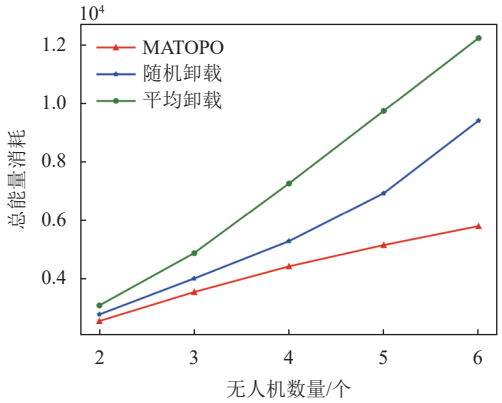

图 9 不同无人机数量下的系统总能耗

4）UAV 飞行轨迹分析

图 10 和图 11 展示了不同 UE 数量下 UAV 的飞行轨迹，UAV 会根据地面 UE 的分布状态调整自身飞行轨迹。如图 10 所示，地面 UE 数量为 10 且各 UE 位置比较分散，此时 3 架 UAV 均以较高的高度飞行，这样保持较好的视距链路，从而更好地为地面分散的 UE 服务。图 11 中 UE 数量为 30 且各 UE 位置相对来说比较集中，此时无人机以中等或较低的高度飞行，这样可在为更多地面 UE 服务的同时降低飞行能耗。在本文所提方法中，UAV 可以通过调整合理的飞行轨迹来降低整体的能量消耗。

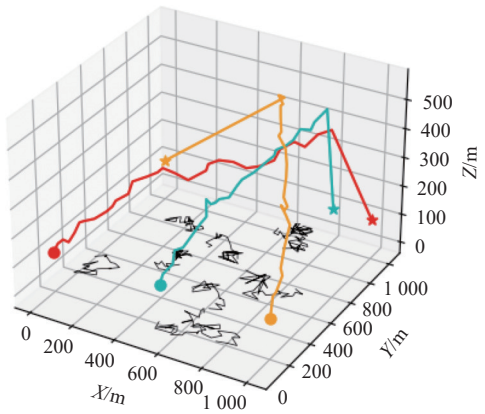

图 10 10 个用户设备时无人机的飞行轨迹

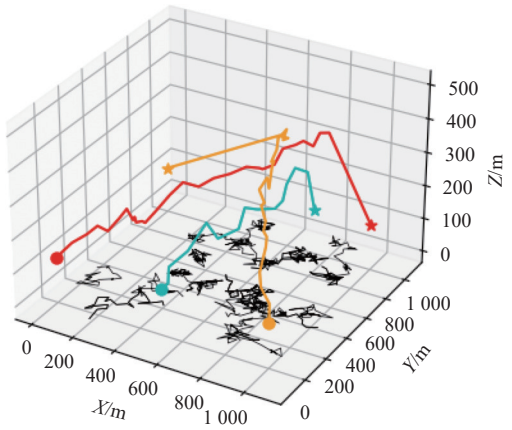

图 11 30 个用户设备时无人机的飞行轨迹

5 结束语

针对多无人机辅助移动边缘计算下的任务卸载与路径优化问题，提出一种多智能体深度强化学习算法。该算法可以联合优化任务卸载策略与无人机飞行轨迹，使系统能量消耗降至最低。此外，为了实现负载均衡，引入用户无人机关联匹配算法，对无人机负载公平性与用户设备被关联服务的公平性进行衡量以实现负载均衡。实验评测结果表明，所提出的多智能体深度强化学习方法相较于基线方法可以获取合理的任务卸载策略并完成无人机路径优化，从而提升任务处理效率、降低能量消耗。

参 考 文 献

[1] ZABIHI Z, MOGHADAM A M E, REZVANI M H. Reinforcement learning methods for computing offloading: A systematic review[J]. ACM Computing Surveys, 2023, 56(1): 1-41.

[2] ISLAM A, DEBNATH A, GHOSE M, et al. A survey on task offloading in multi-access edge computing[J]. Journal of Systems Architecture, 2021, 118: 102225.

[3] 董超, 沈赟, 屈毓锛. 基于无人机的边缘智能计算研究综述[J]. 智能科学与技术学报, 2020, 2(3): 227-239.
DONG C, SHEN Y, QU Y B. A survey of UAV-based edge intelligent computing[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(3): 227-239.

[4] ZHAO N, LU W, SHENG M, et al. UAV-assisted emergency networks in disasters[J]. IEEE Wireless Communications, 2019, 26(1): 45-51.

[5] BOR-YALINIZ R I, El-KEYI A, YANIKOMEROGLU H. Efficient 3-D placement of an aerial base station in next generation cellular networks[C]//2016 IEEE International Conference on Communications (ICC). [S.l.]: IEEE, 2016: 1-5.

[6] JEONG S, SIMEONE O, KANG J. Mobile edge computing via a UAV-mounted cloudlet: Optimization of bit allocation and path planning[J]. IEEE Transactions on Vehicular

Technology, 2017, 67(3): 2049-2063.
[7] 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.
[8] WANG H, KE H, SUN W. Unmanned-aerial-vehicle-assisted computation offloading for mobile edge computing based on deep reinforcement learning[J]. IEEE Access, 2020, 8: 180784-180798.
[9] WANG C, DENG D, XU L, et al. Resource scheduling based on deep reinforcement learning in UAV assisted emergency communication networks[J]. IEEE Transactions on Communications, 2022, 70(6): 3834-3848.
[10] LIU Z, ZHAN C, CUI Y, et al. Robust edge computing in UAV systems via scalable computing and cooperative computing[J]. IEEE Wireless Communications, 2021, 28(5): 36-42.
[11] LIN J, HUANG L, ZHANG H, et al. A novel lyapunov based dynamic resource allocation for UAVs-assisted edge computing[J]. Computer Networks, 2022, 205: 108710.
[12] SEID A M, BOATENG G O, ANOKYE S, et al. Collaborative computation offloading and resource allocation in multi-UAV-assisted IoT networks: A deep reinforcement learning approach[J]. IEEE Internet of Things Journal, 2021, 8(15): 12203-12218.
[13] 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, 2020, 7(1): 73-84.
[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] WEI D, MA J, LUO L, et al. Computation offloading over multi-UAV MEC network: A distributed deep reinforcement learning approach[J]. Computer Networks, 2021, 199: 108439.
[16] GUO H, WANG Y, LIU J, et al. Multi-UAV cooperative task offloading and resource allocation in 5G advanced and beyond[J]. IEEE Transactions on Wireless Communications, 2023, 23(1): 347-359.
[17] 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.
[18] ABOUELENEEN N, ALWARAFY A, ABDALLAH M. Deep reinforcement learning for Internet of drones networks: Issues and research directions[J]. IEEE Open Journal of the Communications Society, 2023, 4: 671-683.
[19] CHEN M, HAO Y. Task offloading for mobile edge computing in software defined ultra-dense network[J]. IEEE Journal on Selected Areas in Communications, 2018, 36(3): 587-597.
[20] LIN C, HAN G, SHAH S B H, et al. Integrating mobile edge computing into unmanned aerial vehicle networks: An SDN-enabled architecture[J]. IEEE Internet of Things Magazine, 2021, 4(4): 18-23.
[21] Al-HOURANI A, KANDEEPAN S, LARDNER S. Optimal LAP altitude for maximum coverage[J]. IEEE Wireless Communications Letters, 2014, 3(6): 569-572.
[22] 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.
[23] ZENG Y, ZHANG R. Energy-efficient UAV communication with trajectory optimization[J]. IEEE Transactions on Wireless Communications, 2017, 16(6): 3747-3760.
[24] DING F, ZHANG X, XU L. The innovation algorithms for multivariable state - space models[J]. International Journal of Adaptive Control and Signal Processing, 2019, 33(11): 1601-1618.
[25] LOWE R, WU Y I, TAMAR A, et al. Multi-agent actor-critic for mixed cooperative-competitive environments [C]//Procoedings of the 31st International Conference on Neural Information Processing Systems. Lone Beach: MIT Press, 2017: 6382-6393.
[26] 方维维, 王云鹏, 张昊, 等. 基于多智能体深度强化学习的车联网通信资源分配优化[J]. 北京交通大学学报, 2022, 46(2): 64-72.
FANG W W, WANG Y P, ZHANG H, et al. Optimized communication resource allocation in vehicular networks based on multi-agent deep reinforcement learning[J]. Journal of Beijing Jiaotong University, 2022, 46(2): 64-72.
[27] WANG L, WANG K, PAN C, et al. Deep reinforcement learning based dynamic trajectory control for UAV-assisted mobile edge computing[J]. IEEE Transactions on Mobile Computing, 2021, 21(10): 3536-3550.
[28] HE Y, GAN Y, CUI H, et al. Fairness-based 3D multi-UAV trajectory optimization in multi-UAV-assisted MEC system[J]. IEEE Internet of Things Journal, 2023, 10(13): 11383-11395.

编　辑　刘飞阳