ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 38,No. 6,Nov. 2023,pp. 1286－1298
DOI:10. 16337/j. 1004-9037. 2023. 06. 005

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

多无人机网络边缘智能计算卸载算法

王心一，陈志江，雷　磊，宋晓勤

（南京航空航天大学公共实验教学部，南京 211106）

摘　要：为了解决大规模部署固定边缘计算节点成本高、机动性差和难以应对突发事件等问题，针对计算密集型和延迟敏感型移动业务需求，提出了一种基于深度强化学习的计算任务卸载算法。考虑多架无人机飞行范围、飞行速度和系统公平效益等约束条件，最小化网络平均计算延时与无人机能耗的加权和。将该非凸性、NP(Non-deterministic polynomial)难问题转化为部分观测马尔可夫决策过程，利用多智能体深度确定性策略梯度算法进行移动用户卸载决策和无人机飞行轨迹优化。仿真结果表明，所提算法在移动服务终端的公平性、系统平均时延和多无人机的总能耗等方面的性能均优于基线算法。其中，所提算法能够得到不同计算性能下的最佳功耗控制，当CPU频率为12.5 GHz时，能耗相比基线降低29.16%，相比随机策略梯度算法降低8.67%。

关键词：移动边缘计算；计算卸载策略；无人机轨迹优化；深度确定性策略梯度；用户公平

中图分类号：TP393　　**文献标志码**：A

Computation Offloading Algorithm for Multi-UAV Network Based on Edge Intelligence

WANG Xinyi, CHEN Zhijiang, LEI Lei, SONG Xiaoqin

(Public Experimental Teaching Department, Nanjing University of Aeronautics & Astronautics, Nanjing 211106, China)

Abstract: In order to solve the problems of high cost, poor mobility and difficulty in coping with emergency in large-scale deployment of fixed edge computing nodes, a computing task offloading algorithm based on deep reinforcement learning is proposed to meet the needs of computing-intensive and delay-sensitive mobile services. Considering constraints such as the flight range, flight speed and system fairness benefits of multiple unmanned aerial vehicles (UAVs), the method aims to minimize the weighted sum of the average computing delay of the network and the UAV energy consumption. This non-convex and non-deterministic polynomial (NP)-hard problem is transformed into a partially observed Markov decision process, and a multi-agent deep deterministic policy gradient algorithm is used for mobile user offloading decision and UAV flight trajectory optimization. Simulation results show that the proposed algorithm outperforms the baseline algorithm in terms of fairness of mobile service terminals, average system delay and total energy consumption of multiple UAVs. Especially, the proposed algorithm can obtain the optimal power consumption control under different computing performance. When the CPU frequency is 12.5 GHz, the energy consumption is 29.16% lower than the Cruise algorithm, and 8.67% lower than the advantage actor-critic(A2C) algorithm.

基金项目：国家自然科学基金（62371232）；江苏省教育厅及未来网络创新研究院“未来网络”科研基金(FNSRFP-2021-ZD-4)。
收稿日期：2023-07-30；**修订日期**：2023-09-12

Key words: mobile edge computing; computation offloading policy; UAV trajectory optimization; deep deterministic policy gradient; user fairness

引 言

随着移动网络技术和无线通信技术的快速发展,智能设备规模急剧增长,随之而来的是一系列创新应用,例如增强现实(Augmented reality, AR)、人脸识别及自动驾驶等[1]。这些应用都有延时敏感、计算复杂的特点,使得移动用户对设备的计算需求和服务质量(Quality of service, QoS)的要求不断提高[2]。由于云计算架构难以满足移动设备的低延时和隐私需求,人们在其基础上提出了移动边缘计算(Mobile edge computing, MEC)以缓解用户计算资源受限的问题[3]。通过将MEC服务器布置在移动网络边缘,用户设备可以通过无线通信将计算任务卸载到MEC服务器从而降低计算延时[4]。

传统边缘基础设施由于其位置固定而受到高部署成本的限制,无人机技术与MEC的结合可以比传统MEC系统在特定的场景上更具优势和灵活。当受到自然灾害导致网络基础设施不可用或移动设备的突然增多超出了网络服务能力,无人机就可以作为临时的通信中继站或边缘计算平台在通信中断或流量热点地区增强无线覆盖,提供计算支持[5]。为提供高质量通信链路,无人机经过航迹规划,可以通过飞行调整自身位置,轻松建立与地面用户的视距(Line-of-sight, LoS)链路[6]。此外,由于单无人机的计算和覆盖能力有限,多无人机能够让更多任务在网络边缘计算,以减少系统延时,提升可靠性[7]。

但是无人机的计算资源与电量受限,为提高MEC系统的性能,有许多关键问题还需解决,包括安全性[8]、任务卸载、能量消耗、资源分配和各种信道情况下的用户延迟性能等,使得无人机辅助MEC模式下的计算卸载研究受到国内外学者的广泛关注。在无人机MEC网络中,可以优化多种类型的变量以实现期望的调度目标,相关的研究工作按照系统模型可分为单无人机和多无人机计算卸载,调度方案可分为集中式和分布式两种,其中集中式常用的有凸优化算法和群智能算法,分布式多基于博弈论方法。对于单无人机卸载模型,文献[9]设计了一种单无人机-边缘云系统,无人机作为移动边缘计算服务器与远程中心云交互,为地面终端提供计算服务,通过块坐标下降算法(Block coordinate decent, BCD)对资源分配和无人机三维轨迹进行迭代优化来最小化无人机的整体能耗;文献[10]研究了物联网(Internet of things, IoT)中的无人机辅助MEC系统,作者分别采用拉格朗日对偶法和逐次凸逼近算法(Successive convex approximation, SCA)来处理非凸问题,着重通过时隙调度、无人机路径规划和功率分配来降低整体能耗,进一步扩大了计算资源的规模。对于多无人机计算卸载,文献[11]提出了一种两层联合优化方法,外层利用粒子群结合遗传算法(Genetic algorithm, GA)来优化无人机的部署,内层采用贪心算法获得合理的卸载决策,以最小化平均任务响应时间;文献[12]提出了一种在城市场景下多无人机的车辆辅助计算卸载架构,将无人机和车辆之间计算数据的交易过程建模为一个交易博弈,通过分析交易过程,可以得到最优的交易策略。

文献[9-12]使用的传统优化方法由于需要大量迭代和先验知识来获得一个近似最优解,因此不适用于动态环境中的实时MEC应用。随着机器学习在研究中的广泛应用,许多研究人员也在探索基于学习的MEC调度算法[13],鉴于机器学习的最新进展,深度强化学习(Deep reinforcement learning, DRL)现已成为研究热点。对于单无人机卸载模型,文献[14]提出了一种基于深度Q网络(Deep Q-network, DQN)的端到端模型来联合优化计算卸载和无人机轨迹控制,以降低整个系统的时延和能耗,借助深度神经网络(Deep neural network, DNN)的拟合能力,DRL可以有效地解决具有高维状态空间的复杂决策问题;为了缓解DQN中典型的高估问题,文献[15]采用双深度Q网络优化无人机的飞行轨迹和关联用户,在用户实时移动环境下实现最大化系统吞吐量;文献[16]提出了另一种改进型的深

度确定策略梯度算法来寻求能够提高用户体验的最佳策略，通过重新设计Critic网络结构提升了算法的稳定性和收敛速度。文献[14-16]虽然采用深度强化学习方法实现计算卸载，但算法用的是单智能体，不适合具有复杂任务划分的系统。对于多无人机卸载模型，文献[17]使用多架无人机作为边缘服务器，为物联网设备提供计算卸载的机会，采用随机优先重放机制的DQN算法加速训练时间，提高了收敛的稳定性；文献[18]设计了一种基于多无人机的IoT边缘网络模型，提出了一种基于多智能体深度强化学习(Multi-agent deep reinforcement learning，MADRL)的方法用于协同计算卸载和资源分配，通过集中训练和分散执行的方式降低计算成本，提高分配效率。文献[17-18]进一步证明了深度强化学习在无线通信网络中的有效性，但在综合能耗节省和延迟性能保障等方面仍存在一定的不足。

尽管现有的工作取得了很大的进展，但基于学习的方法仍需要进一步研究，以使其更适用于复杂动态环境下的无人机辅助MEC系统。因此，本文试图通过多智能体深度强化学习的方法来求解无人机轨迹和卸载优化问题，从而获得可扩展和有效的调度策略，主要的研究工作与创新点如下：

(1) 研究了一个多无人机辅助MEC模型，其中多无人机部署在三维空间，充当网络边缘设施。在该系统模型的基础上，基于用户的位置和任务信息，联合优化多无人机的飞行轨迹和计算卸载策略以最小化系统时延和无人机能耗，同时保证用户的服务公平。

(2) 将上述非凸计算卸载优化问题表示为一个部分观测马尔可夫决策过程(Partially observable Markov decision process，POMDP)，将模型环境中的变量都转化为POMDP中的元素，并提出了一种端到端的基于多智能体深度确定策略梯度(Multi-agent deep deterministic policy gradient，MADDPG)的轨迹优化卸载算法来解决该优化问题。

(3) 设置不同的模型参数进行仿真试验，结果验证了该算法的有效性。在相同的仿真条件下，本文提出的算法与其他基线算法相比，在降低系统时延和无人机能耗方面也有显著的优势。

1 无人机辅助计算卸载模型

1.1 系统模型

无人机辅助用户卸载的移动边缘计算系统，如图1所示。该系统有M个移动用户设备(Mobile device，MD)随机分布在一块方形区域，区域边长设为$l_{\max}$，将移动设备的集合记为$m\in\mathcal{M}\triangleq\{1,2,\cdots,M\}$。同时有$U$架搭载MEC服务器的无人机，在目标区域上空以固定的高度H_u飞行，用于给移动设备提供卸载服务，无人机集合记为$u\in\mathcal{U}\triangleq\{1,2,\cdots,U\}$。设无人机执行一次飞行任务的总时长为$T$，总时长可被分为$N$个等长的时隙，时隙的集合记为$\tau\in\mathcal{T}\triangleq\{1,2,\cdots,N\}$。每个MD在每个时隙$\tau$有一个计算密集型任

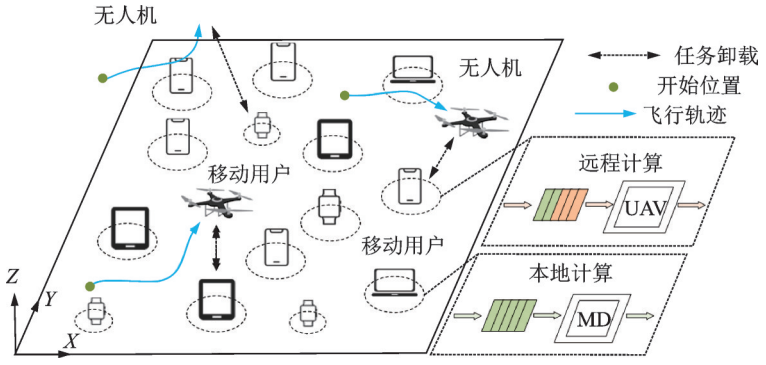

图1 无人机辅助计算卸载模型

Fig.1 Assisted computation offloading model of UAV

务,任务记为$S_m(\tau)=\{D_m(\tau),F_m(\tau)\}$,其中$D_m(\tau)$表示数据比特量,$F_m(\tau)$表示每比特所需CPU周期。

每架无人机在每个时隙τ只为一个终端设备提供计算卸载服务,用户只需在本地计算任务的一小部分,其余卸载到无人机辅助计算,以减少计算的延时和能耗,卸载计算量的比例记为$\Delta_{m,u}(\tau)\in[0,1]$。无人机和用户设备之间的卸载决策变量可表示为

$$D=\{\alpha_{m,u}(\tau)|u\in\mathbb{U},m\in\mathbb{M},\tau\in\mathbb{T}\} \tag{1}$$

式中$\alpha_{m,u}(\tau)\in\{0,1\}$。当$\alpha_{m,u}(\tau)=1$时表示设备$\mathrm{MD}_m$在时隙$\tau$的计算任务由无人机$\mathrm{UAV}_u$辅助计算,$\Delta_{m,u}(\tau)>0$,当$\alpha_{m,u}(\tau)=0$时则表示只在本地执行计算任务,$\Delta_{m,u}(\tau)=0$。决策变量需要满足

$$\sum_{u\in\mathbb{U}}\alpha_{m,u}(\tau)\leqslant 1\quad\forall m\in\mathbb{M},\forall\tau\in\mathbb{T} \tag{2}$$

1.2 移动模型

与之前的研究类似,移动设备会在每个时隙内随机移动到新的位置,每个设备的移动与其当前的速度和角度有关。假设MD_m在时隙τ的坐标记为$c_m(\tau)=[x_m(\tau),y_m(\tau)]$,则其下一时隙$\tau+1$的坐标可表示为

$$\begin{cases}x_m(\tau+1)=x_m(\tau)+\cos(2\pi\rho_{1,m})d_{\max}\rho_{2,m}\\ y_m(\tau+1)=y_m(\tau)+\sin(2\pi\rho_{1,m})d_{\max}\rho_{2,m}\end{cases} \tag{3}$$

式中:$d_{\max}$代表设备移动的最大距离;移动方向和距离概率均服从均匀分布,$\rho_{1,m},\rho_{2,m}\sim U(0,1)$。为了简化模型,无人机服务终端时仅考虑其在该时隙的起始位置。

同样地,每架无人机在高度H_u的水平面轨迹也可以用无人机在每个时隙的离散位置$c_u(\tau)$来表示,假设UAV_u在时隙τ选择飞去服务MD_m,则其飞行方向记为$\beta_u(\tau)\in[0,2\pi]$,飞行速度为$v_u(\tau)\in[0,V_{\max}]$,其中$V_{\max}$为飞行最大速度,飞行时间为t_{fly},无人机飞行消耗的能量为[19]

$$E_u^{\mathrm{fly}}(\tau)=\mu\left(\frac{\|c_u(\tau)-c_u(\tau-1)\|}{t_{\mathrm{fly}}}\right)=\mu\|v_u(\tau)\|^2 \tag{4}$$

式中$\mu=0.5M_ut_{\mathrm{fly}}$,$M_u$为无人机总质量。

1.3 通信与计算模型

在本次设计的MEC系统中,计算卸载采用可部分卸载策略,MD_m在时隙τ的本地计算延时可表示为

$$T_m^{\mathrm{local}}(\tau)=\frac{\left(1-\sum_{u\in\mathbb{U}}\Delta_{m,u}(\tau)\right)D_m(\tau)F_m(\tau)}{f_m} \tag{5}$$

式中f_m表示MD_m的本地计算能力(每秒CPU周期数)。

本次实验采用视距链路(LoS link)模型模拟实际的无人机对地通信[20],无人机和用户之间的信道增益$h_{m,u}(\tau)$遵循自由空间路径损失模型,可表示为

$$h_{m,u}(\tau)=\frac{g_0}{\|c_u(\tau)-c_m(\tau)\|^2+H_u^2} \tag{6}$$

式中:g_0为每米信道功率增益;H_u为飞行高度。

由于每架无人机在每个时隙只服务一个用户,因此本次研究忽略信道间的通信干扰,则无人机和地面设备之间的瞬时传输速率$r_{m,u}(\tau)$定义为

$$r_{m,u}(\tau)=B\log_2\left(1+\frac{\hat{p}_mh_{m,u}(\tau)}{\sigma^2}\right) \tag{7}$$

式中:B代表信道带宽;$\hat{p}_m$为移动设备上传链路的发射功率;σ^2代表无人机端的高斯白噪声。关联用户

MD_m的传输数据延时为

$$T_{m,u}^{\mathrm{Trans}}(\tau)=\frac{\alpha_{m,u}(\tau)D_m(\tau)\Delta_{m,u}(\tau)}{r_{m,u}(\tau)} \tag{8}$$

在传输完计算任务后,无人机执行卸载计算任务,卸载计算的延时和能耗分别为

$$T_{m,u}^{\mathrm{off}}(\tau)=\frac{\alpha_{m,u}(\tau)D_m(\tau)F_m(\tau)\Delta_{m,u}(\tau)}{f_u} \tag{9}$$

$$E_{m,u}^{\mathrm{off}}(\tau)=\bar{p}_u T_{m,u}^{\mathrm{off}}(\tau) \tag{10}$$

式中:f_u表示无人机计算能力;$\bar{p}_u=\kappa_u f_u^3$表示无人机执行计算时的CPU功率,$\kappa_u=10^{-27}$为芯片常数[21]。

由于各种计算密集型任务的结果输出数据量都远小于输入,因此可以忽略下行链路传输所花费的延时。基于以上通信与计算模型,MD_m在时隙τ完成任务$S_m(\tau)$的时延$T_m(\tau)$可以计算为

$$T_m(\tau)=\max\left(T_m^{\mathrm{local}}(\tau),T_{m,u}^{\mathrm{trans}}(\tau)+T_{m,u}^{\mathrm{off}}(\tau)\right) \tag{11}$$

无人机UAV_u辅助计算卸载在时隙τ的总能耗包括卸载计算能耗和飞行能耗,记为

$$E_u(\tau)=E_{m,u}^{\mathrm{off}}(\tau)+E_u^{\mathrm{fly}}(\tau) \tag{12}$$

1.4 优化问题

在本文提出的无人机辅助MEC系统中,目标是通过优化无人机卸载决策和飞行轨迹最大限度的减少无人机的能耗和系统平均计算延时。用户MD_m的平均$m\in\mathbb{M}\triangleq\{1,2,\cdots,M\}$延时可以表示为

$$T_m^{\mathrm{avg}}=\frac{1}{N}\sum_{\tau=1}^{N}T_m(\tau) \tag{13}$$

则系统平均计算延时可计算为

$$T^{\mathrm{avg}}=\frac{1}{N}\sum_{\tau=1}^{N}\frac{1}{M}\sum_{m\in\mathbb{M}}T_m(\tau)=\frac{1}{N}\sum_{\tau=1}^{N}T^{\mathrm{mean}}(\tau) \tag{14}$$

同时为了保证服务的公平性,避免无人机在任务期间只服务某几个移动设备以减少能耗,而不服务其他用户,可以使用公平指数ξ_τ来衡量这一情况,定义如下[22]

$$\xi_\tau=\frac{\left(\sum\limits_{m\in\mathbb{M}}\sum\limits_{\tau'=1}^{\tau}\sum\limits_{u\in\mathbb{U}}\alpha_{m,u}(\tau')\right)^2}{M\sum\limits_{m\in\mathbb{M}}\left(\sum\limits_{\tau'=1}^{\tau}\sum\limits_{u\in\mathbb{U}}\alpha_{m,u}(\tau')\right)^2} \tag{15}$$

从任务初始到时隙τ,如果所有用户的被服务累积次数相近,ξ_τ的值就接近1。将优化问题总结如下

$$\begin{cases}\min\limits_{\{P,Z\}}\sum\limits_{\tau=1}^{N}\mathbb{E}\left[\phi_t T^{\mathrm{mean}}(\tau)+\phi_e E_u(\tau)\right]\\ \text{s.t. C1:}\sum\limits_{m\in\mathbb{M}}\alpha_{m,u}(\tau)\leqslant 1\quad\forall u\in\mathbb{U},\forall\tau\in\mathbb{T}\\ \quad\text{C2:}\ 0\leqslant x_u(\tau),y_u(\tau)\leqslant l_{\max}\quad\forall u\in\mathbb{U},\forall\tau\in\mathbb{T}\\ \quad\text{C3:}\ v_u(\tau)\leqslant V_{\max}\quad\forall u\in\mathbb{U},\forall\tau\in\mathbb{T}\\ \quad\text{C4:}\ \beta_u(\tau)\in[0,2\pi]\quad\forall u\in\mathbb{U},\forall\tau\in\mathbb{T}\\ \quad\text{C5:}\ \Delta_{m,u}(\tau)\in[0,1]\quad\forall u\in\mathbb{U},\forall\tau\in\mathbb{T}\\ \quad\text{C6:}\ \|c_u(\tau)-c_{u'}(\tau)\|\leqslant d_{\mathrm{safe}}\quad\forall u,u'\in\mathbb{U},\forall\tau\in\mathbb{T}\\ \quad\text{C7:}\ \xi_N\geqslant\xi_{\min}\end{cases} \tag{16}$$

式中:$P=\{\beta_u(\tau),v_u(\tau)\}$;$Z=\{\alpha_{m,u}(\tau),\Delta_{m,u}(\tau)\}$;$\phi_t$和$\phi_e$为权重参数;C1限制无人机每个时隙只服务一个用户;C2和C6限制无人机的飞行范围;C3和C4限制无人机的飞行速度和角度;C5表示计算任务可以被部分卸载;C7保证系统的公平效益;d_{safe}和$\xi_{\min}$为预先设定的无人机之间最小安全距离和最低公平指数。

2 无人机轨迹优化计算卸载算法

上述提出的优化问题既包括了连续变量也包含了离散变量,由于系统需要在每个时隙做出决策,很难用传统的方法求解这类混合整数非线性规划问题(Mixed integer nonlinear programming problem, MINLP)。由此本文提出了一种基于MADDPG的深度强化学习算法。

2.1 POMDP建模

在强化学习的方法中,可以将多无人机辅助计算卸载问题看作是一个部分观测马尔可夫决策过程,由元组$\{S, A, O, Pr, R\}$构成[7],如图2所示。通常有多个智能体与环境交互,在当前状态$s_\tau\in S$,每个智能体基于s_τ得到自身观察$o_\tau\in O$并做出动作$a_\tau\in A$,环境对动作产生即时奖励$r_\tau\in R$以评估当前动作的好坏,并以概率$Pr(S_{\tau+1}|S_\tau,A_\tau)$进入下一状态,新状态只取决于当前的状态和各个智能体的动作。智能体的动作基于策略$\pi(a_\tau|o_\tau)$执行,其目标为学习到最优策略以最大化长期累积奖励,可表示为

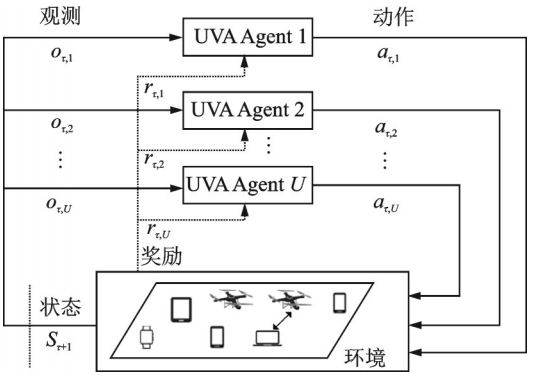

图2 POMDP决策过程示意图

Fig.2 Diagram of POMDP decision process

$$\pi^*=\arg\max_{\pi}E_{a_\tau\sim\pi(a_\tau|o_\tau)}\left[\sum_{\tau'=\tau}^{\infty}\gamma^{\tau'-\tau}r_{\tau'}\right] \tag{17}$$

式中γ为奖励折扣。在给定状态s_τ下,策略的状态动作价值函数用来评判每一个动作的表现,可以表示为

$$Q_\pi(s_\tau,a_\tau)=E_\pi\left[\sum_{\tau'=\tau}^{\infty}\gamma^{\tau'-\tau}r_{\tau'}|_{s_\tau,a_\tau}\right] \tag{18}$$

由于学习的目标是找到最优策略,可以由状态动作价值函数的最大值确定,即通过最大状态动作价值就可以找到对应的最优策略,最优动作价值函数的贝尔曼方程可以表示为

$$Q^*(s_\tau,a_\tau)=E\left[r_\tau+\gamma\max_{a_{\tau+1}}Q^*(s_{\tau+1},a_{\tau+1})\right] \tag{19}$$

状态动作价值可采用时序差分(Temporal difference, TD)的方法不断迭代更新。

把每架无人机当作一个智能体,对本次模型的观测、动作、状态和奖励函数定义如下:

(1)观测空间。每架无人机都只有有限的观测范围,观测范围的半径设为r_{obs},因此只能观测到部分状态信息,而全局的状态信息和其他无人机的动作都是未知的。单架无人机UAV_u在时隙τ能观测到的信息有自身的位置信息$c_u(\tau)$和观测范围内K个移动用户当前的位置信息、任务信息以及服务次数$k_u(\tau)=\left\{c_m(\tau),S_m(\tau),\sum_{\tau'}^{\tau}\sum_{u\in\mathbb{U}}\alpha_{m,u}(\tau')|m=1,2,\cdots,K\right\}$,则智能体的观测值可记为

$$o_u(\tau)=\{c_u(\tau),k_u(\tau)\} \tag{20}$$

(2)动作空间。基于观测到的信息,无人机会选择相应的动作,首先需要确定在当前时隙τ服务哪位用户以及卸载比例$\Delta_{m,u}(\tau)$,再决定自身的飞行角度$\beta_u(\tau)$和飞行速度$v_u(\tau)$,因此动作可记为

$$a_u(\tau)=\{m(\tau),\Delta_{m,u}(\tau),\beta_u(\tau),v_u(\tau)\} \tag{21}$$

(3)状态空间。系统的状态可看作所有无人机观测结果的集合,包含所有无人机的位置,所有移动设备的位置、任务以及其他信息,状态记为

$$s(\tau)=\{o_u(\tau)|u\in\mathbb{U}\} \tag{22}$$

(4)奖励。智能体执行动作后得到的反馈称之为奖励,用于判定动作的好坏,指导智能体更新其策略。设计合适的奖励函数对智能体的学习起着重要作用,一般来说,奖励函数都与优化目标相对应,本次优化的目标是最小化无人机的能耗和系统平均计算延时,与最大奖励回报正好呈负相关,因此将无人机执行动作后的奖励定义为

$$r_u(\tau)=D_m(\tau)\cdot\left(-T^{\mathrm{mean}}(\tau)-\psi E_u(\tau)-P_u(\tau)\right) \tag{23}$$

式中:ψ用来对无人机能耗和用户平均时延进行数值对齐;$P_u(\tau)$为额外的惩罚项,如果无人机执行动作后飞出场地或和其余无人机的距离小于安全距离,就需要增加惩罚;$D_m(\tau)\in[0,1]$为衰减系数,定义为无人机处理移动终端卸载任务后得到的效益,具体计算如下

$$D_m(\tau)=1-\exp\left(-\frac{\left(\sum\limits_{\tau'=1}^{\tau}\sum\limits_{u\in\mathbb{U}}\alpha_{m,u}(\tau')\right)^{\eta}}{\sum\limits_{\tau'=1}^{\tau}\sum\limits_{u\in\mathbb{U}}\alpha_{m,u}(\tau')+\beta}\right) \tag{24}$$

式中η和β为相关常数,其函数图像为广义线性模型(sigmoid),输入为当前用户的累积服务次数,次数越多,其值越大,奖励越小,效益越低。

2.2 基于MADDPG计算卸载算法

在多智能体的环境中,由于环境是动态变化的,单个智能体可能无法仅靠自身来适应动态环境,策略梯度中的方差也会随着智能体数量的增多而变大。基于MADDPG的深度强化学习算法适用于具有连续动作空间的多智能体策略学习,智能体可以协同学习,提高系统性能[23]。

MADDPG基于Actor-Critic框架,每个智能体都有自己的Actor网络和Critic网络,以及各自的目标网络,如图3所示。Actor网络负责为智能体制定策略,θ_u代表其网络参数;Critic网络输出对最优状态-动作价值函数的估计,用于评估训练阶段的Actor网络的策略性能,记为$Q(s(\tau),a_1(\tau),\cdots,a_U(\tau)|w_u)$,$w_u$代表其网络参数。MADDPG算法采用的是集中训练与分散执行的模式,Critic网络的输入包含一个时隙内所有智能体的观测值和动作,网络参数在集中训练模式下更新,

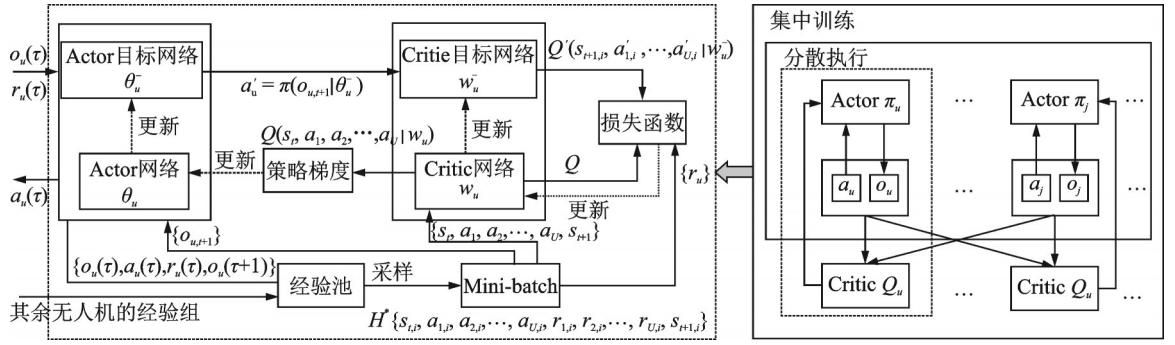

图3 本文算法训练模型

Fig.3 Training model of the proposed algorithm

但在分散执行时,Actor网络的输入仅有自身的观测值。这样做的好处是Critic网络可以在训练过程中参考其他智能体的行为,从而更好地评估Actor网络的性能,提高策略的稳定性。

在每轮和环境的交互中,无人机基于当前环境状态$s(\tau)$和观测范围,将观测值$o_u(\tau)$输入到Actor网络中得到即时动作$a_u(\tau)$。为了能到达更好的探索效果,通常会在动作上人为增加一个服从正态分布的噪声$n_e \sim N(0,\sigma_e^2)$,并随着训练以0.999 5的速率慢慢衰减。等所有无人机执行完动作后,环境给每个无人机返回奖励$r_u(\tau)$并进入下一个时隙的状态$s(\tau+1)$,无人机又得到下一时隙的观测值$o_u(\tau+1)$。$\{o_u(\tau),a_u(\tau),r_u(\tau),o_u(\tau+1)\}$称为一条经验组,无人机将其存放在经验池$B_u$中,用以网络参数训练更新。

当每架无人机的经验池里的记录达到足够数量的时候,就开始训练神经网络,在每轮的训练过程中,随机从每个智能体的经验池中抽取H组记录,将同样时刻的每组拼接得到H条新记录,记为:$\left\{s_{t,i},a_{1,i},a_{2,i},\cdots,a_{U,i},r_{1,i},r_{2,i},\cdots,r_{U,i},s_{t+1,i}|i=1,2,\cdots,H\right\}$,此时为off-policy训练,与当前时隙$\tau$无关。算法同时对$Q$函数以及最优策略进行学习,首先使用时序差分集中训练每一个智能体的Critic网络,训练Q值函数的损失函数定义为

$$L(w_u)=\frac{1}{H}\sum_{i=1}^{H}\left(y_{u,i}-Q(s_{t,i},a_{1,i},a_{2,i},\cdots,a_{U,i}|w_u)\right)^2 \tag{25}$$

$$y_{u,i}=r_{u,i}+\gamma Q'(s_{t+1,i},a'_{1,i},a'_{2,i},\cdots,a'_{U,i}|w_u^-)|_{a'_{j,i}=\pi(o_{j,t+1,i}|\theta_j^-)} \tag{26}$$

式中:$Q'(\cdot|w_u^-)$代表Critic目标网络;$\pi(\cdot|\theta_u^-)$代表Actor目标网络,它们都具有滞后更新的网络参数,使训练变得更稳定。

Critic网络需要尽量降低损失以逼近真实的Q^*值,Actor网络则用Q值的确定策略梯度作梯度上升更新网络参数以最大化动作价值

$$\nabla_{\theta_u}J\approx\frac{1}{H}\sum_{i=1}^{H}\nabla_{\theta_u}\pi(o_{u,t,i}|\theta_u)\nabla_{a_{u,i}}Q(s_{t,i},a_{1,i},a_{2,i},\cdots,a_{U,i}|w_u) \tag{27}$$

最后在固定的间隔以更新率φ更新目标网络

$$\begin{cases}\theta_u^-\leftarrow\varphi\theta_u+(1-\varphi)\theta_u^-\\ w_u^-\leftarrow\varphi w_u+(1-\varphi)w_u^-\end{cases} \tag{28}$$

基于MADDPG的无人机辅助计算卸载训练算法代码详细描述如下:

输入:用户的位置集合、任务集合和服务次数,无人机的位置集合以及信道参数

输出:到达当前最优策略的动作向量

初始化:各智能体Actor网络和Critic网络及其各自目标网络参数,经验池容量经验池容量B_u,动作噪声方差σ_e^2,软更新率φ,奖励折扣γ

(1) FOR Episode=1, 2,…, MAX_E DO:

(2) 重置环境,获得初始观测值$(o_1(1),o_2(1),\cdots,o_U(1))$

(3) FOR τ=1,2,…,N DO:

(4) FOR u=1,2,…,U DO:

(5) 生成动作$a_u(\tau)=\pi(o_u(\tau)|\theta_u)+n_e$

(6) 执行所有动作,如果飞出区域或距离太近,无人机将暂留在当前位置

(7) 根据式(19)获得即时奖励$(r_1(\tau),r_2(\tau),\cdots,r_U(\tau))$

(8) 获取新的观测值$(o_1(\tau+1),o_2(\tau+1),\cdots,o_U(\tau+1))$

(9) END FOR

(10) FOR u=1,2,…,U DO:

(11) $\{o_u(\tau), a_u(\tau), r_u(\tau), o_u(\tau+1)\}$存储到$B_u$中

(12) IF 存储数量足够大 DO:

(13) 随机抽取H组记录,并与其余无人机的记录组合

(14) 根据式(25)最小化损失更新Critic网络

(15) 根据式(27)计算策略梯度更新Actor网络

(16) 根据式(28)软更新目标网络

(17) END FOR

(18) END FOR

(19) END FOR

3 仿真结果与分析

本节将通过代码仿真结果来评估所提出的基于MADDPG的计算卸载方案。首先介绍实验环境设置,包括训练环境和神经网络超参数。然后分析模型中的参数设置对其收敛性能和系统性能的影响。之后在各种环境设置下将所提出的算法和其他几种基线算法进行比较。

3.1 仿真环境

无人机辅助MEC系统由各种实体构成,包括无人机、用户和网络环境。本文将目标区域设置为边长$l_{\max}=200$ m的方形区域,各架UAV的初始位置设为(50,50),(150,50),(50,150),(150,150),飞行高度设置100 m,最大飞行速度为30 m/s,初始能量500 kJ。MD被随机放置在区域内,每次移动的最大距离设为5 m。对于数据传输,信道带宽为1 MHz,为了提高信噪比,MD都以最大发送功率$\hat{p}_m=0.1$ W传输。其余主要参数如表1所示。

表1 环境参数设置

Table 1 Environment parameter setting

参数	描述	取值
M	终端设备数量	[8,20]
U	无人机数量	4
N	总时隙数量	80
T/s	任务周期	480
$D_m(\tau)$/ Mb	任务数据量	[2,4]
$F_m(\tau)$/ (cycle•bit^{-1})	任务计算复杂度	1 000
M_u/ kg	无人机总质量	9.65
f_m/ GHz	移动端计算能力	0.6
f_u/ GHz	服务端计算能力	2.5
g_0/ dB	信道功率增益	−50
σ^2/ dBm	白噪声功率	−90
d_{safe}/ m	最小安全距离	40
$\xi_{\min}$	最小公平指数	0.7

对于模型训练,使用tensorflow1.10.0实现MADDPG算法,Actor网络结构设计为[300,100,4]的3层全连接神经网络,Critic网络设计为[400,300,100,1]的4层全连接神经网络,每个智能体都是相同的DNN结构。Actor网络的隐藏层使用ReLU作为激活函数,输出层使用tanh作为激活函数,可以使模型的动作输出固定在[−1,1]。在训练阶段,batch-size为H=64,最大回合数为3 000,Actor和Critic的学习率分别为0.001和0.002,神经网络都采用Adam优化器,经验池的大小为10 000。对未来奖励的折扣γ设为0.8,目标网络的软更新率φ设为0.01。 能耗与时延的比率设为$\phi_e:\phi_t=5:6$。

3.2 参数分析

图4显示了不同的探索噪声参数σ_e^2对模型收敛性能的影响,纵坐标为其中一架无人机的累积奖励,横坐标为训练周期。噪声的影响主要集中在前200个Episode,初始探索率较大可以尽力搜索动作空间中的优秀策略,随着训练的继续,探索率逐渐变小,模型的最终性能并无特殊变化表示智能体学习到了最佳策略。当噪声方差较小时($\sigma_e^2=5$),模型收敛的速度较快,但容易对动作空间探索不完全,所以会在后面收敛阶段出现异常波动;当噪声方差较大时($\sigma_e^2=20$),智能体则需要花更多的时间来探索环境,

收敛速度较慢。综上,本文可以选择适中的方差$\sigma_e^2=10$获得较为理想的效果。

图5显示了无人机不同的探测范围r_{obs}对用户服务公平的影响,纵坐标为系统公平指数,横坐标为训练周期。使用衰减系数后,无人机倾向于选择服务次数少的用户以获得更多奖励。可以看到无人机探测范围越大,能考虑的用户的数量越多,用户的服务次数越平均;但无人机探测范围越大,它们探测重叠的范围就越多,也越容易发生决策冲突,如果同时选择同一用户反而降低系统性能。综上,$r_{\text{obs}}=40$为较合适的探测范围。

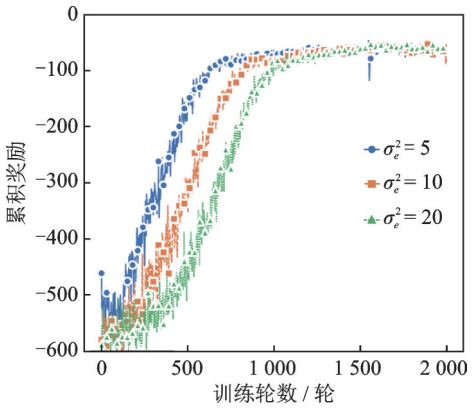

图4 不同噪声参数对模型收敛性能的影响($M=20$)

Fig.4 Convergence performance of different exploration noises ($M=20$)

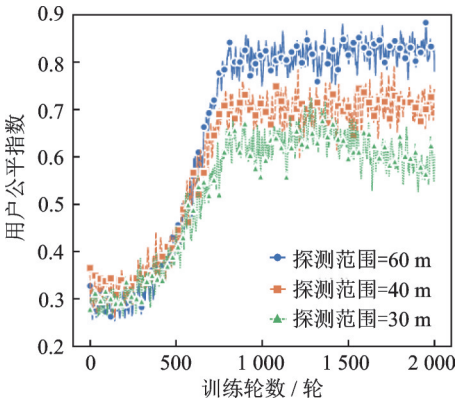

图5 无人机不同探测范围对用户公平的影响($M=20$)

Fig.5 Fairness under different detection ranges of UAV ($M=20$)

图6显示了奖励衰减系数$D_m(\tau)$对用户服务公平的影响。如果不使用衰减系数,无人机从节省能耗的角度最后只会一直服务最近的一两个用户来获取更高的奖励,从图中可以看到ξ_N的值在探索一段时间后开始快速下降,不能够满足约束条件,因此利用本次设计的衰减系数能够有效分配计算资源给各个用户,保证服务的公平。

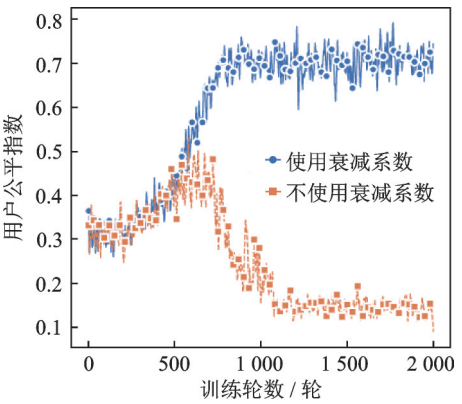

图6 衰减系数对用户服务公平的影响($M=20$)

Fig.6 Effect of attenuation coefficient index on fairness performance ($M=20$)

3.3 性能比较

为便于比较,本文采用如下两种算法:(1)巡航模式(Cruise),基线算法,无人机以区域某中心为圆心,50 m为半径,绕圈定点飞行,在每个时隙无人机随机选择用户卸载,卸载比率为$\Delta_{m,u}(\tau)=0.8$;(2)基于随机策略梯度的强化学习算法(Advantage actor-critic, A2C)[24],为了实现连续的动作空间,策略由不相关的正态分布随机变量组成的随机向量表示,Actor网络输出2个四维向量,第1个向量是动作分布平均值$\boldsymbol{\mu}_a$,第2个向量是动作方差$\boldsymbol{\sigma}_a^2$,Critic网络则输出状态价值并由此更新网络参数。

图7比较了模型训练时不同算法对无人机能耗的收敛情况,纵坐标为无人机执行一次任务周期剩余的电池能量,横坐标为训练周期。巡航模式下由于没有做任何其余的优化,每次都有固定的飞行消耗,因此能量消耗的较多。A2C由于是on-policy算法,相邻的输入之间存在相关性会导致网络收敛不

稳定，使智能体输出动作的优化空间变小，而本次所提出的基于MADDPG的算法则采用经验回放的方式打破相关性，Critic网络更容易从复杂的环境中获取信息，因此有很好的收敛性。通过探索整个动作空间，并采用确定动作，MADDPG比其他模型消耗的能量更少，从而以端到端的方式优化无人机轨迹和卸载任务，其最终效果也是最好的。

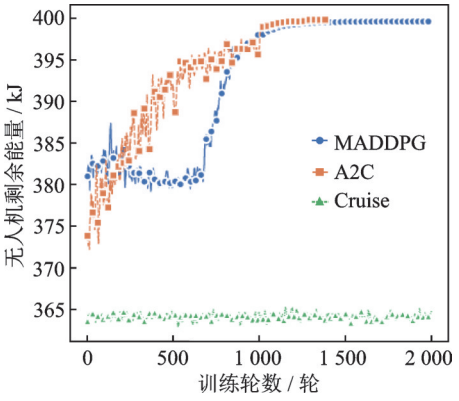

图7 不同算法在能耗方面的收敛情况(M=10)

Fig.7 Convergence of different algorithms in energy consumption (M=10)

图8显示了无人机的计算性能对能耗的影响。固定用户数量不变，随着无人机计算性能的提升，无人机处理计算任务的时间减少，用户可以将更多的计算任务卸载上传，因此在同一时隙内无人机的计算功耗增加，计算任务的处理时延会适当降低。由图8所示，基于MADDPG算法能够得到不同计算性能下的最佳功耗控制，当CPU频率为12.5 GHz时，能耗相比基线降低29.16%，相比随机策略梯度算法降低8.67%。

图9比较了不同用户数量下各方法之间的系统平均计算延时T^{avg}。可以看到，随着用户数量的上升，总计算任务量增加，但无人机的计算资源有限，平均处理延时随之增加，相比之下MADDPG可以实现最低延时，它可以在连续动作空间中找到确定的最优卸载比例，从而减少任务处理时延。当用户数量小于12时，用户的状态维度较少，MADDPG算法可以获得最佳性能，系统时延比基线算法可以减少360 ms，比基于随机策略梯度的强化学习算法减少了276 ms；当用户数量变多，环境的状态信息变多，决策的复杂度上升，MADDPG的性能会降低，延时靠近随机策略梯度算法A2C。

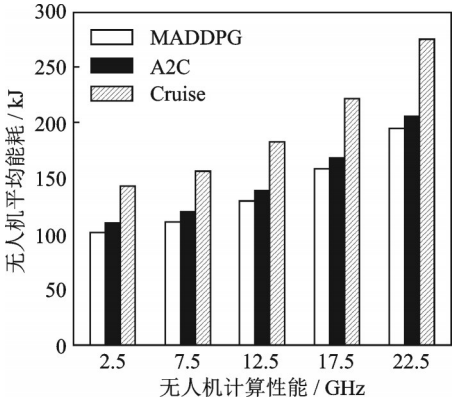

图8 无人机计算能力对能耗的影响(M=10)

Fig.8 Effect of computing capabilities of UAV on energy consumption (M=10)

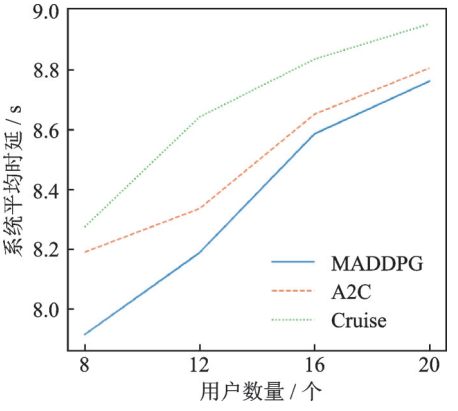

图9 不同用户数量下平均处理延时对比

Fig.9 Comparison of processing delay under different number of UDs

在用户数量不变的情况下，随着终端计算任务复杂度的增加，每种方法的系统平均时延也都在增加，如图10所示。应用深度强化学习有利于在动态网络环境中做出最优策略，随着计算复杂度上升，时延占奖励的比重越大，无人机会更优先考虑给计算任务分配更多资源，从而减小计算任务的时延。在计算复杂度最高时，本文算法的计算任务平均时延与基线相比减少1.14 s，与随机策略梯度算法相比减少244 ms。

4 结束语

本文研究了一个大规模多无人机辅助MEC网络计算卸载问题,其中的移动设备MD可以将计算任务部分卸载到无人机或在本地执行。在满足约束条件的情况下,通过联合优化无人机卸载决策和飞行轨迹最大限度地减少无人机的能耗和系统平均计算延时。对此,本文提出了一种基于MADDPG的深度强化学习算法,无人机之间可以相互合作并分享经验,以集中训练和分散执行的方式得到近似最优策略。然后通过仿真实验分析了动作噪声和探测范围对基于MADDPG模型的影响。最后与基线算法比较,仿真结果表明,在能耗和平均任务延时方面,所提算法均具有更好的性能。未来可以加入对计算和信道资源分配的研究,让无人机可以服务多个用户,使场景更加细化。在深度神经网络结构上可以利用神经语言程序学中的注意力机制将注意力单元嵌入到网络中,使智能体能够自行筛选输入中的重要信息,提高分布式算法的收敛性能。

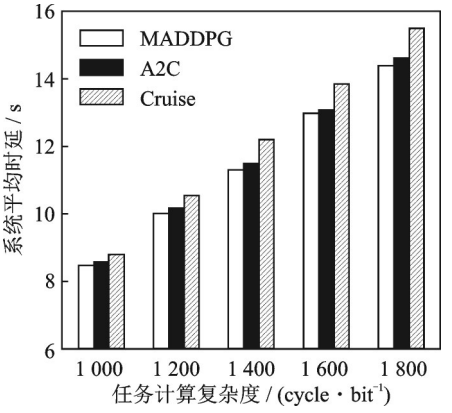

图10 系统平均时延与任务计算复杂度的关系(M=10)

Fig.10 Relation of processing delay and computational complexities (M=10)

参考文献:

[1] 宁兆龙, 张凯源, 王小洁, 等. 基于多智能体元强化学习的车联网协同服务缓存和计算卸载[J]. 通信学报, 2021, 42(6): 118-130.
NING Zhaolong, ZHANG Kaiyuan, WANG Xiaojie, et al. Cooperative service caching and peer offloading in Internet of vehicles based on multi-agent meta-reinforcement learning[J]. Journal on Communications, 2021, 42(6): 118-130.

[2] SABELLA D, VAILLANT A, KUURE P, et al. Mobile-edge computing architecture: The role of MEC in the Internet of things[J]. IEEE Consumer Electronics Magazine, 2016, 5(4): 84-91.

[3] SHI W, ZHANG X, WANG Y, et al. Edge computing: State-of-the-art and future directions[J]. Journal of Computer Research and Development, 2019, 56(1): 69-89.

[4] WANG C, YU F, LIANG C, et al. Joint computation offloading and interference management in wireless cellular networks with mobile edge computing[J]. IEEE Transactions on Vehicular Technology, 2017, 66(8): 7432-7445.

[5] GAN Y, HE Y. Trajectory optimization and computing offloading strategy in UAV-assisted MEC system[C]//Proceedings of 2021 Computing, Communications and IoT Applications. Shenzhen, China: IEEE, 2021: 132-137.

[6] WU Q, ZHANG R. Common throughput maximization in UAV-enabled OFDMA systems with delay consideration[J]. IEEE Transactions on Communications, 2018, 66(12): 6614-6627.

[7] PENG H, SHEN X. Multi-agent reinforcement learning based resource management in MEC- and UAV-assisted vehicular networks[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(1): 131-141.

[8] MEHTA P, GUPTA R, TANWAR S. Blockchain envisioned UAV networks: Challenges, solutions, and comparisons[J]. Computer Communications, 2020, 151: 518-538.

[9] MEI H, YANG K, LIU Q, et al. Joint trajectory-resource optimization in UAV-enabled edge-cloud system with virtualized mobile clone[J]. IEEE Internet of Things Journal, 2019, 7(7): 5906-5921.

[10] ZHANG T, XU Y, LOO J, et al. Joint computation and communication design for UAV-assisted mobile edge computing in IoT[J]. IEEE Transactions on Industrial Informatics, 2019, 16(8): 5505-5516.

[11] CHEN Z, ZHENG H, ZHANG J, et al. Joint computation offloading and deployment optimization in multi-UAV-enabled MEC systems[J]. Peer-to-Peer Networking and Applications, 2022, 15(1): 194-205.

[12] DAI M, SU Z, XU Q, et al. Vehicle assisted computing offloading for unmanned aerial vehicles in smart city[J]. IEEE

Transactions on Intelligent Transportation Systems, 2021, 22(3): 1932-1944.

[13] 刘雷, 陈晨, 冯杰, 等. 车载边缘计算中任务卸载和服务缓存的联合智能优化[J]. 通信学报, 2021, 42(1): 18-26.
LIU Lei, CHEN Chen, FENG Jie, et al. Joint intelligent optimization of task offloading and service caching for vehicular edge computing[J]. Journal on Communications, 2021, 42(1): 18-26.

[14] ZHANG L, ZHANG Z Y, MIN L, et al. Task offloading and trajectory control for UAV-assisted mobile edge computing using deep reinforcement learning[J]. IEEE Access, 2021, 9: 53708-53719.

[15] LIU Q, SHI L, SUN L, et al. Path planning for UAV-mounted mobile edge computing with deep reinforcement learning[J]. IEEE Transactions on Vehicular Technology, 2020, 69(5): 5723-5728.

[16] LU H, HE X, DU M, et al. Edge QoE: Computation offloading with deep reinforcement learning for Internet of Things[J]. IEEE Internet of Things Journal, 2020, 7(10): 9255-9265.

[17] SHI S, WANG M, GU S, et al. Energy-efficient UAV-enabled computation offloading for industrial internet of things: A deep reinforcement learning approach[J]. Wireless Networks, 2021. DOI: 10.1007/S11276-021-02789-7.

[18] 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.

[19] 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.

[20] ISLAM S K, HAIDER M R. Sensors and low power signal processing[M]. [S.l.]: Springer Science & Business Media, 2009.

[21] LIU Y, XIONG K, NI Q, et al. UAV-assisted wireless powered cooperative mobile edge computing: Joint offloading, CPU control, and trajectory optimization[J]. IEEE Internet of Things Journal, 2019, 7(4): 2777-2790.

[22] CHANG H, CHEN Y, ZHANG B, et al. Multi-UAV mobile edge computing and path planning platform based on reinforcement learning[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2021. DOI: 10.1109/TETCI.2021.3083410.

[23] 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.

[24] CHENG N, LYU F, QUAN W, et al. Space/aerial-assisted computing offloading for IoT applications: A learning-based approach[J]. IEEE Journal on Selected Areas in Communications, 2019, 37(5): 1117-1129.

作者简介:

王心一(1987-),**通信作者**,女,工程师,研究方向:计算机网络、智能无人机集群技术,E-mail: wangxinyi@nuaa.edu.cn。

陈志江(1997-),男,硕士研究生,研究方向:智能无人机集群技术。

雷磊(1981-),男,教授,博士生导师,研究方向:航空平台组网技术、智能无人机集群技术、无线泛在网络技术等。

宋晓勤(1973-),女,副教授,研究方向:无线泛在网络技术等。

(编辑:张黄群)