

空地网络资源分配与无人机基站动态部署算法

张尚伟, 和思梦

(西北工业大学网络空间安全学院, 710072, 西安)

摘要: 为解决空地网络中地面设备数量变化引起的用户体验质量无法满足的问题,提出了一种智能网络资源分配与多无人机基站动态部署方案。首先,考虑用户体验质量和无人机、地面设备能量约束,以最小化系统总能耗为目标进行问题建模;其次,将多无人机的动态部署问题转化成具有连续动作集的马尔可夫决策过程,并根据优化目标设计了基于能耗惩罚的奖励函数;然后,采用基于确定性策略梯度的深度强化学习算法求解此问题;最后,通过仿真和对比实验验证所提方案的有效性和优越性。实验结果表明:对于海量用户场景,所提算法比深度强化学习和演员-评论家算法有更好的收敛性和更高的累积奖励,与单无人机和传统地面基站部署方案相比,所提方案系统的能耗降低约30%~40%,用户服务质量满意度提升约50%~60%。

关键词: 无人机;资源分配;动态部署;强化学习算法

中图分类号: TN929.5 **文献标志码:** A

DOI: 10.7652/xjtuxb202403016 **文章编号:** 0253-987X(2024)03-0172-11

Resource Allocation and Dynamic Deployment Algorithm for Unmanned Aerial Vehicle Enabled Base Stations in Air-Ground Networks

ZHANG Shangwei, HE Simeng

(School of Cybersecurity, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: To address the problem of unsatisfactory user experience quality caused by the fluctuation of ground device quantity in an air-ground network, a solution for intelligent network resource allocation and dynamic deployment of base stations with multiple unmanned aerial vehicles (UAVs) is proposed. Firstly, considering user experience quality and the energy constraints of UAVs and ground devices, the problem is modeled with the objective of minimizing the total system energy consumption. Secondly, the dynamic deployment of multiple UAVs is transformed into a Markov decision process (MDP) with a continuous action set, and a reward function based on energy penalty is designed according to the optimization objective. Thirdly, a deep reinforcement learning algorithm based on deep deterministic policy gradient (DDPG) is used to solve this problem. Lastly, the effectiveness and superiority of the proposed solution are verified through simulation and comparative experiment. Experimental results show that, for scenarios with a massive number of users, the proposed algorithm exhibits better convergence and higher cumulative rewards compared to deep reinforcement learning and actor-critic algorithms. In comparison to single UAV and traditional ground base station deployment solutions, the proposed solution reduces energy consumption by approximately 30%

to 40%, and improves user satisfaction with service quality by around 50% to 60%.

Keywords: unmanned aerial vehicle; resource allocation; dynamic deployment; reinforcement learning algorithm

近年来,无人机(UAV)由于其高机动性、高灵活性、易维护性和易操纵性受到各行业的广泛关注。具有高算力的服务器可被安装在无人机上,作为通信基站、中继节点或边缘计算节点为目标区域内的设备提供通信、存储和计算等服务。与传统地面方案相比,无人机的优势主要体现在部署速度快、成本低、可扩展性强等方面,在未来空地一体化网络、工业互联网、车联网、智慧城市、智慧农业等新兴物联网应用及行业领域具有广阔的应用前景。虽然无人机辅助移动边缘计算在优化网络性能、降低传输时延、增加网络容量等方面有着巨大的优势和潜力,但存在设备能量受限、空地链路特征复杂、视距链路干扰强、信道质量差异大等问题^[1]。这些因素使得利用无人机建立大覆盖范围、高速率、低能耗的移动边缘计算网络仍面临着很大的技术挑战。

针对上述问题,研究人员近年来从无人机轨迹规划及动态部署角度开展了深入研究,并针对用户服务质量、系统性能及能耗等方面给出了众多高效的解决方案。根据应用场景和无人机在其中扮演的角色,相关研究可归为两类:一类借助地面网络,无人机被部署为空中用户设备来执行相应任务^[2];另一类无人机携带移动边缘计算服务设备,作为空地通信系统中的空中基站或者移动中继,为地面设备提供相应的通信、计算服务^[3]。其中,第二类应用场景中无人机轨迹规划和动态部署优化研究主要以能耗、时延、用户服务质量(QoS)及用户体验质量为目标展开,也是本文研究重点。

在以能耗最小化为目标的工作中,Wang等^[4]为实现最小化系统能耗,并保证所有无人机之间的地理公平性、负载公平性,提出多无人机轨迹规划问题,并给出了低复杂度的卸载决策方案求解此问题。Shen等^[5]在大规模机器类型通信场景中,由无人机负责高效采集数据,由于设备能量有限,提出了以最大化能量效率为目标的无人机轨迹规划问题,采用基于贪心算法的动态规划算法解决此问题。Zhu等^[6]提出了一种多无人机辅助的无线传感器网络,其中集群无人机负责从各个节点采集数据并将数据传输给头无人机;为了最小化系统能耗,提出无人机轨迹规划问题,并采用基于深度神经网络的强化学习算法解决此问题。

在以时延最小化为目标的研究中,Tan等^[7]考虑5G蜂窝中的延迟敏感社交AR服务,建立基于多无人机的数据采集模型,并提出以最小化处理和传输时延为目标的多无人机巡航轨迹控制问题。随后,提出自组织协同方案并利用连续凸优化算法解决此问题。Zhan等^[8]提出无人机辅助物联网通信框架,在物联网设备任务和资源约束下,提出以最小化任务完成时间为目标优化计算卸载、资源分配、无人机轨迹的问题,并将路径规划问题转化为离散的非凸等价问题,利用交替优化算法求解此问题。

在以满足用户服务质量为目标的研究中,Huang等^[9]考虑应急信息快速收集和可靠传输的现实需求,提出了以最大化用户服务质量为目标的无人机路径规划问题;通过联合优化无人机和用户设备的位置、传输功率,最大程度地提高了上行和下行数据传输的可靠性;通过最大限度地提高每个无人机的加速度,在最短的时间内遍历所有的路径点。

综上所述,现有研究大多针对无人机移动边缘计算下的数据采集、应急通信、物联网计算等场景,给出了路径规划和动态部署的解决方案,能够有效减少能量消耗、降低服务延迟、提高用户服务质量和用户体验质量、扩大网络覆盖范围。然而,现有研究部分采用贪心策略求解路径规划问题,此类策略可能获得局部最优解;部分采用凸优化方法,该方法具有一定的局限性,难以解决非凸优化问题。另外,凸优化算法的收敛结果与初始值设定密切相关,不同的初始值可能会求得不同的解。此外,在海量地面设备多无人机的场景中,现有的研究工作较少考虑物联网设备动态变化这一实际场景。因此,本文研究热点区域地面设备数量可变场景中的多无人机动态部署问题,考虑用户体验质量、设备能量、无人机能量等因素,以最小化系统总能耗为目标进行问题建模;将多无人机动态部署问题转化为马尔可夫决策过程,并采用基于确定性策略梯度的深度强化学习(DDPG)算法求解问题。

1 系统模型

本文考虑部署无人机为海量地面用户设备提供通信和计算服务的场景。无人机可以灵活且快速移动到需要覆盖的区域,尤其适用于城市中的高密度

用户区域。在紧急情况或大型活动等场景中,借助无人机可以随时应对人群聚集和网络拥堵,解决了传统基站部署成本高、建设周期长的问题。其次,无人机可以部署到较高的空中位置,向高层建筑物提供更广泛的信号覆盖范围,能够从一定程度上克服信号遮挡问题,弥补传统地面网络在部署时受地形因素影响的局限性^[10]。此外,无人机可以搭载各种传感器,在城市管理规划、环境保护监测以及交通管理等应用中,为大量的高密度物联网终端提供实时的数据采集、处理以及计算等服务^[11]。

如图1所示,假设场景中存在 N 个地面用户设备,其计算能力无法满足当前任务计算需求,需将任务卸载到无人机上的边缘服务器中。在系统中的基站端部署一个虚拟管理系统与无人机进行信息交换以控制网络资源分配,并进一步协调无人机基站的动态部署^[12]。当设备数量增加时,通过新增无人机(飞行高度 H 保持不变,飞行速度为 V)来满足设备需求;当联网设备数量降低时,可在满足服务质量的条件下适当减少部署无人机的数量。

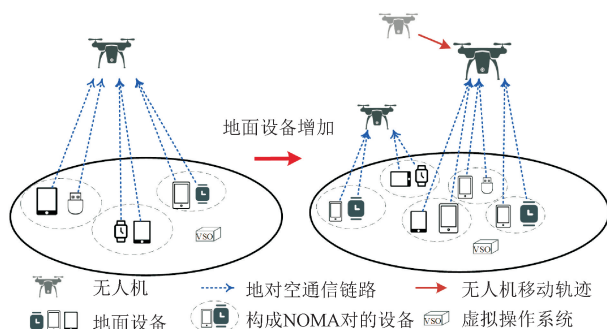


图1 系统模型

Fig. 1 System model

将整个系统服务提供时间分成多个时隙,假设在第 t 个时隙,第 m 架($m \leq \chi$)无人机被启动加入到该系统中,无人机的水平位置可以表示为

$$\begin{cases} l_m^{\text{UAV}}(t) = (x_m^{\text{UAV}}(t), y_m^{\text{UAV}}(t)) \\ m \in \{0, 1, 2, \dots, M\}, t \in \{0, 1, 2, \dots, T\} \end{cases} \quad (1)$$

式中: $x_m^{\text{UAV}}(t)$ 表示无人机的横坐标; $y_m^{\text{UAV}}(t)$ 表示无人机的纵坐标; χ 为无人机最多部署数量。由于覆盖范围有限,无人机在当前位置服务完所有的设备后,需要重新调整位置继续服务其他用户。给定无人机的运动方向 θ 、飞行距离 d_{fly} 及第 t 时隙的位置,其在第 $t+1$ 时隙的悬停位置为

$$\begin{cases} x_m^{\text{UAV}}(t+1) = x_m^{\text{UAV}}(t) + \cos\theta(t)d_{\text{fly}}(t) \\ y_m^{\text{UAV}}(t+1) = y_m^{\text{UAV}}(t) + \sin\theta(t)d_{\text{fly}}(t) \end{cases} \quad (2)$$

式中: $d_{\text{fly}}(t)$ 为无人机移动距离, $d_{\text{fly}}(t) \in [d_{\min}, d_{\max}]$; θ 为运动方向, $\theta \in [0, 2\pi]$; $d_{\max}$ 为一个时隙最大飞行距离。

对于动态增加的地面设备数量,设定每过一段时间加入一定数量设备,那么在服务时间为 Γ 时,设备数量为 N_r 。地面设备一旦进入到目标服务区域,其位置保持不变,那么任一设备 n 的位置可表示为

$$l_n^{\text{D}} = (x_n^{\text{D}}, y_n^{\text{D}}), \quad n \in \{0, 1, 2, \dots, N\} \quad (3)$$

对于任务卸载策略,本文考虑地面设备采用全卸载的策略将数据卸载至无人机服务器上^[13]。

1.1 非正交多址接入(NOMA)通信模型

在任一时隙 t ,假设无人机 m 悬停在位置 $l_m^{\text{UAV}}(t)$ 与地面设备 n 建立连接,则它们之间的距离为

$$d_{m,n}(t) = (H^2 + (x_m^{\text{UAV}}(t) - x_n^{\text{D}})^2 + (y_m^{\text{UAV}}(t) - y_n^{\text{D}})^2)^{-2} \quad (4)$$

对于空地信道(A2G)模型,本文同时考虑大尺度路径损耗和小尺度的瑞利衰落因素^[14]。在第 t 个时隙,无人机 m 与设备 n 之间平均路径损耗表示为

$$L_{m,n}(t) = P_{m,n}^{\text{LoS}} L^{\text{LoS}}(t) + (1 - P_{m,n}^{\text{LoS}}) L^{\text{NLoS}}(t) \quad (5)$$

式中: L^{LoS} 和 L^{NLoS} 分别是视距和非视距链路的路径损耗; $P_{m,n}^{\text{LoS}}$ 是无人机与设备之间的空地视距链路的概率。 L^{LoS} 、 L^{NLoS} 、 $P_{m,n}^{\text{LoS}}$ 分别表示为

$$\begin{cases} L^{\text{LoS}}(t) = 20 \lg \frac{4\pi f_d d_{m,n}(t)}{c} + \eta_{\text{LoS}} \\ L^{\text{NLoS}}(t) = 20 \lg \frac{4\pi f_d d_{m,n}(t)}{c} + \eta_{\text{NLoS}} \\ P_{m,n}^{\text{LoS}} = \frac{1}{1 + \alpha \exp(-\beta(90 - \theta_{m,n} - \alpha))} \end{cases} \quad (6)$$

其中, f_d 是载波频率, c 是光速, η_{LoS} 和 η_{NLoS} 是平均附加损耗,此值取决于环境因素, $(90 - \theta_{m,n})$ 是设备到无人机的仰角, α 和 β 取决于环境的常数值(如城市、郊区、农村等^[15])。

结合式(5),在第 t 个时隙,无人机 m 与地面设备 n 之间的信道增益为

$$g_{m,n}(t) = 10^{\frac{-L_{m,n}(t)}{10}} = \frac{10^{\frac{-P_{m,n}^{\text{LoS}} L^{\text{LoS}}(t) - (1 - P_{m,n}^{\text{LoS}}) L^{\text{NLoS}}(t)}{10}}}{10^{\frac{-L_{m,n}(t)}{10}}} \quad (7)$$

为提高频谱利用率和系统覆盖率,采用功率域NOMA技术。在任一时刻,将信道条件差异较大的两个设备构成一个NOMA组^[16]。在任一NOMA组中,信道条件较好的用户 n 以及信道条件较差的用户 n' 的上行数据传输速率可分别表示为^[17]

$$\left\{ \begin{aligned} r_{m,n}(t) &= B_n \text{lb} \left(1 + \frac{P_{\text{tr}} g_{m,n}(t)}{P_{\text{tr}} g_{m,n'}(t) + \sum_{k \in N, k \neq n} r P_{\text{tr}} g_{m,k}(t) + \sigma^2} \right) \\ r_{m,n'}(t) &= B_n \text{lb} \left(1 + \frac{P_{\text{tr}} g_{m,n'}(t)}{\sum_{k \in N, k \neq n} r P_{\text{tr}} g_{m,k}(t) + \sigma^2} \right) \end{aligned} \right. \quad (8)$$

式中: B_n 表示无人机给地面设备划分的子信道带宽; P_{tr} 是地面设备的发射功率; σ^2 为零加性高斯白噪声功率; $r \in \{0, 1\}$ 是上行链路频谱复用指标, 当 $r = 1$ 时, 表示设备 k 复用当前设备通信信道而产生同频干扰。

1.2 时延与能耗模型

将任务到达至计算完成的时长定义为用户等待延迟, 若超过阈值 τ , 则认为需求未被满足。设无人机 m 在第 $t-1$ 时隙完成对当前设备服务后, 需在第 t 时隙重新调整位置以服务其余设备, 其飞行时间为

$$D_{\text{fly},m}(t) = \frac{\|l_m^{\text{UAV}}(t), l_m^{\text{UAV}}(t-1)\|}{V} \quad (9)$$

式中: $\|\cdot\|$ 表示两点之间的欧几里得距离, 表示无人机 m 在第 t 个时隙的飞行距离。考虑到信道质量, 假设在第 t 个时隙地面设备尝试和距离最近的无人机通信, 关联的设备集合记为 Ω 。

无人机 m 与所关联用户设备 n 之间的任务传输时间可表示为

$$d_{m,n}(t) = \frac{a_{m,n}}{r_{m,n}(t)} \quad (10)$$

式中: $a_{m,n}$ 是地面设备 n 卸载到无人机 m 的任务数据量。在时隙 t , 地面设备 n 上传任务耗费时长为

$$D_{\text{com},n}(t) = \max_{n \in \Omega, a_{m,n} > 0} \{d_{m,n}(t)\} \quad (11)$$

在该时隙, 无人机 m 计算与之关联的地面设备上的卸载任务的时长为

$$D_{\text{comm},m}(t) = \sum_{n \in \Omega} \frac{a_{m,n}}{f_c} \quad (12)$$

考虑到无人机移动性, 未被服务设备的等待时延为无人机飞行时长和服务其他设备的时长总和。因此, 在时隙 t , 地面设备 n 的等待时间可以表示为

$$D_{\text{wait},n}(t) = D_{\text{fly}}(t) + D_{\text{comm},k}(t) + \sum_{k \in N, k \neq n} D_{\text{com},k}(t) \quad (13)$$

在整个工作周期中, 设备总等待时延表示为

$$D_{\text{delay},n} = \sum_{t=0}^T D_{\text{wait},n}(t) \quad (14)$$

则无人机的一个工作周期总时长可以表示为

$$D_{\text{time}} = \sum_{t=0}^T D_{\text{fly}}(t) + \sum_{t=0}^T D_{\text{comm},n}(t) + \sum_{t=0}^T \sum_{n=0}^N D_{\text{com},n}(t) \quad (15)$$

在能耗方面, 本文考虑系统总能耗包括无人机飞行能耗、任务计算能耗、所有地面设备的发射能耗。在任意时隙 t , 相应的能耗分别表示为

$$E_{\text{fly}}(t) = \sum_{m \in M} P_f D_{\text{fly},m}(t) \quad (16)$$

$$E_{\text{comm}}(t) = \sum_{m \in M} \sum_{n \in N} \gamma_c C(f_c)^2 a_{m,n} \quad (17)$$

$$E_{\text{up}}(t) = \sum_{n \in N} P_{\text{tr}} D_{\text{comm},n}(t) \quad (18)$$

式中: P_f 为无人机飞行功率; C 为边缘计算服务器计算 1 bit 需要的 CPU 转数; γ_c 为有效电容开关; f_c 为边缘计算服务器 CPU 频率。对于上行通信, 设定所有设备的发射功率相同, 且保持恒定。在第 t 个时隙, 无人机的剩余电量可以表示为

$$b(t) = b(t-1) - E_{\text{fly}}(t) - E_{\text{comm}}(t) \quad (19)$$

式中: 无人机的初始电量设置为 $b(0) = B$ 。

1.3 问题定义

一般地, 基于地面基站的组网方式在通信、冷却、数据存储、计算、监控管理以及检查维护等方面需要大量地消耗电力^[18], 而基于无人机组网方案的能耗主要用于通信、计算和保持飞行方面, 通常来说要低于基站的能耗。在实际应用中, 无人机由于受到自身能源限制导致其在空中飞行时间非常有限。为此, 本文的目标是在用户体验质量的约束下, 进行多无人机动态部署以最小化系统总能耗。基于此目标以及上述对系统的通信建模、时延与能耗建模, 将问题描述为

$$\begin{aligned} & \max_{\Lambda} \sum_{t=0}^T E_{\text{fly}}(t) + E_{\text{comm}}(t) + E_{\text{up}}(t) \\ & \text{s. t. } \left\{ \begin{aligned} & C_1: 0 \leq x_m^{\text{UAV}} \leq X, \forall m \in \{0, 1, \dots, M\} \\ & C_2: 0 \leq y_m^{\text{UAV}} \leq Y, \forall m \in \{0, 1, \dots, M\} \\ & C_3: x_{m,n}(t) \in \{0, 1\}, \forall t \in \{0, 1, \dots, T\} \\ & C_4: S_m(t) = \{n \in N, x_{m,n}(t) = 1\} \\ & C_5: \sum_{m=0}^M 1 \left\{ \sum_{t=0}^T |S_m(t)| > 0 \right\} \leq \chi \\ & C_6: Q_t = \sum_{n=0}^N 1 \{D_{\text{delay},n} \leq \tau\} \\ & C_7: \frac{\sum_{t \leq T} Q_t}{N_r} \geq \phi, \forall t \in \{0, 1, \dots, T\} \\ & C_8: b(t) \geq 0, \forall t \in \{0, 1, \dots, T\} \end{aligned} \right. \end{aligned} \quad (20)$$

式中: Λ 为无人机部署方案; X 和 Y 表示目标区域边

长; $x_{m,n}(t)$ 表示第 t 时隙地面设备 n 与无人机 m 之间是否存在关联, 若 $x_{m,n}(t) = 0$ 表示在此时隙设备 n 与无人机 m 没有关联; $x_{m,n}(t) = 1$ 则表示存在关联; $S_m(t)$ 表示在第 t 时隙与无人机 m 关联的地面设备数量; χ 表示可部署无人机最大数量; τ 表示用户最大等待时间, 若等待时间超过此值, 表示未能满足用户需求(不好的体验质量); Q_t 表示在第 t 时隙满足用户需求的设备数量; ψ 是满足用户需求的最小设备比例; C_1 和 C_2 为防止无人机飞出目标区域边界的约束条件; C_5 表示启动无人机的数量不能超过预期数量; C_7 表示满足用户体验质量需要符合的条件; C_8 表示不能超过无人机总电量。

2 算法设计

2.1 多无人机动态部署方案

当目标区域联网设备数量增加后, 无人机将动态调整悬停位置, 具体的部署方案如算法 1 所示。

算法 1 多无人机动态部署算法

输入 已激活用户设备总数量 N_r

输出 无人机最优动态部署策略 Δ

```

1  初始化  $i = 0$ ;
2  while  $i < 5$  do
3      激活  $i/5$  个用户设备;
4      无人机服务时间  $t_s = 0$ ;
5      根据信道质量进行无人机与地面设备的关联
      以及用户 NOMA 分组;
6      根据最优部署策略  $\Delta$  进行位置部署;
7       $t_s = t_s + D_{\text{time}}$ ;
8      while  $\sum_{t \leq T} Q_t / N_r < \psi$  do
9          启动新无人机服务用户设备;
10         无人机服务时间  $t_s = 0$ ;
11         根据信道质量进行无人机与地面设备的关
            联以及用户 NOMA 分组;
12         根据算法 2 求得最优部署策略  $\Delta$ , 并对无
            人机进行部署;
13          $t_s = t_s + D_{\text{time}}$ ;
14     end while
15      $i = i + 1$ ;
16 end while

```

2.2 用户关联和 NOMA 分组方法

本文根据信道质量进行无人机与地面设备的通信关联。对于任意无人机 m 和设备 n , 考虑到通信双方的距离是刻画空地信道质量的关键因素, 可根据式(4)计算它们之间的距离 $d_{m,n}(t)$, 如果满足条件 $d_{m,n}(t) < r$ 和 $d_{m,n}(t) < d_{m,k}(t)$, 则无人机 m 与设备

n 可以建立通信关联。对于空地 NOMA 用户分组问题, 在第 t 个时隙, 给定与无人机 m 关联的设备集合 Ω , 首先根据式(7)计算集合 Ω 中 $2K$ 个设备信道增益 $c_{m,n}(t)$, 并对设备信道增益排序: $c_{m,1}(t) \cdots \leq c_{m,K}(t) \leq \cdots \leq c_{m,2K}(t)$; 随后, 为了使得配对设备之间有着相对均匀的信道增益差值, 采取均匀信道增益差异配对算法, 即将集合 Ω 中 $2K$ 个设备分为强用户集和弱用户集, 弱用户集为 Ω' , $\Omega' = \{1, 2, \dots, K\}$; 强用户集为 Ω'' , 其中 $\Omega'' = \{K+1, K+2, \dots, 2K\}$; 将强用户集中最大信道增益的用户设备与弱用户集中最大信道增益的用户设备配对, 依此类推, 最终分配方案可以表示为

$$\begin{pmatrix} \text{用户对 1} \\ \text{用户对 2} \\ \vdots \\ \text{用户对 } N \end{pmatrix} = \begin{pmatrix} 1, K+1 \\ 2, K+2 \\ \vdots \\ K, 2K \end{pmatrix} \quad (21)$$

2.3 基于确定性策略梯度的深度强化学习算法

考虑到无人机在当前时隙的状态只与上一个时隙的状态有关, 可认为无人机的状态集具有马尔可夫特性。将控制多个无人机的虚拟管理系统视为一个智能体, 多无人机的动态部署问题转化成一个连续状态空间、连续动作空间的马尔可夫决策过程^[19]。首先, 无人机向虚拟管理系统发送各自覆盖范围内的网络信息(如信道状态信息、地面激活用户数量及服务请求、无人机计算能力和剩余能量等)。随后, 将此初始状态输入神经网络中进行训练。训练结束后, 虚拟管理系统将结果传递给无人机, 以协调多个无人机之间的动态部署。下面定义马尔可夫过程四元组 (S, A, P, R) , 其中 S 表示状态集, A 表示动作集, P 表示转移概率, R 表示奖励函数。四元组 (S, A, P, R) : 无人机在每个时隙 t 选择飞向另一个位置与其当前位置、电量、剩余等待服务的设备数量 N_r 以及剩余等待服务设备的位置集 L 有关, 针对该问题的马尔可夫决策过程的状态集合定义为

$$S = \{s_t \mid s_t = \{l_m^{\text{UAV}}(t), b(t), N_r, L\}\} \\ m \in \{0, 1, \dots, M\}, L \in N_t, \forall t \in \{1, 2, \dots, T\} \quad (22)$$

在 $t+1$ 时隙的位置部署由此时隙无人机的运动方向 θ 与运动方向 d_{fly} 决定, 因而马尔可夫决策过程的动作集合描述^[20]为

$$A = \{a_t \mid a_t = \{\theta(t), d_{\text{fly}}(t)\}\} \\ \theta(t) \in [0, 2\pi], d_{\text{fly}}(t) \in [d_{\min}, d_{\max}] \quad (23)$$

将无人机与地面设备总能耗最小化转化为马尔可夫决策决策过程的最大化累积奖励。系统的奖励函数需要结合无人机处理用户设备的正奖励与无人

机能耗带来的负奖励。若正奖励与负奖励的数量级相差太多,会导致训练结果出现局部最优,因而采用 Sigmoid 函数定义系统的正向奖励,使得正奖励的值处于 0~1 之间。正向奖励可表示为

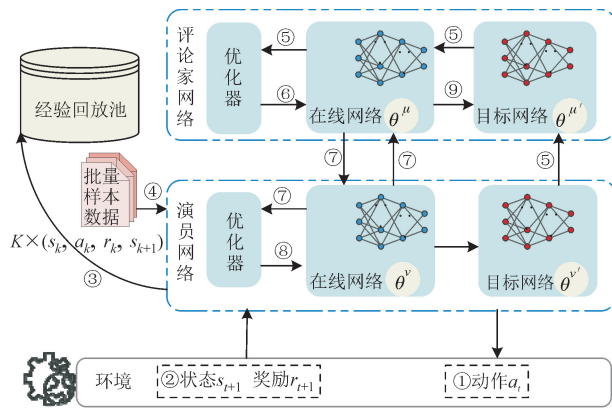
$$\xi_t = 1 - \exp\left[\frac{(N_{\text{server}}(t))^{\eta}}{N_{\text{server}}(t) + \beta}\right] \quad (24)$$

为达到最小化系统总能耗目标,负奖励被定义为无人机和设备总能耗,结合式(16)~式(18),在第 t 个时隙奖励函数被定义为

$$r_t = \xi_t - \varphi_1(E_{\text{comm}}(t) + E_{\text{up}}(t) + E_{\text{fly}}(t)) \quad (25)$$

式中: φ_1 是系统能量消耗权重因子; η 和 β 是调节正向奖励函数系数,不同值对应不同的系统正向奖励影响; $N_{\text{server}}(t)$ 表示多架无人机服务设备总数量。针对本研究场景中无人机连续动作空间的动态部署,拟采用基于策略梯度的强化学习算法解决此问题。

如图 2 所示,DDPG 算法在传统的强化学习算法中加入经验回放池,将马尔可夫决策过程中出现的四元组 (S, A, P, R) 放到一个先入先出的数据结构中存储起来。在之后的学习过程中,随机从经验回放池中抽取数据更新深度神经网络的参数^[21]。同时,DDPG 算法用卷积神经网络(CNN)来逼近价值函数,预测网络与目标网络使用相同的卷积神经网络,将当前状态 s_t 输入到预测网络中,并根据当前预测网络的参数输出预测值 $Q(s_t, a_t; \theta)$,将下一个状态 s_t 输入到目标网络中,找到下一个状态的状态动作对应的最大值以及此动作的及时奖励 r_{t+1} ,并从目标网络输出当前状态动作对的目标值。



图中①~⑨表示算法基本训练过程步骤。

图 2 基于确定性策略梯度的深度强化学习算法框架

Fig. 2 The framework of deep reinforcement learning algorithm based on deterministic policy gradient

DDPG 算法的训练过程为^[22]: 在每个时隙 t , 由虚拟管理系统观测并收集到当前环境的状态 s_t , 之后演员网络在一次训练之后由状态 s_t 会输出动作 a_t ,

$= \mu(s_t; \theta^{\mu})$ 。执行过此动作后会产生新状态 s_{t+1} 及奖励 r_{t+1} , 并将四元组放在经验回放池中。当经验回放池中的样本数量足够多时, 从中取出 K 个样本数据, 输入到演员网络和评论家网络中进行卷积神经网络参数的训练。在更新参数时, 评论家网络会通过 Adam 优化器计算在线网络的梯度, 通过评论家在线网络获得目标值 Q 。接着根据此梯度更新评论家在线网络的策略参数 θ^{μ} , 随后计算演员网络中的在线网络梯度, 并根据此梯度更新演员在线网络策略参数 θ^{ν} 以及演员目标网络参数 $\theta^{\nu'}$ 。最后更新演员目标网络中的参数 $\theta^{\mu'}$ 。至此, 本回合内的一次训练过程结束, 转移至下一个状态继续更新参数, 直到终止状态。评论家在线网络与评论家目标网络之间的损失函数 $L(\theta^{\nu})$ 的定义如下

$$L(\theta^{\nu}) = \frac{1}{2K} \sum_{k=1}^K [Q - Q(s_k, a_k; \theta^{\nu})]^2 = \frac{1}{2K} \cdot \sum_{k=1}^K [(r_t + \rho Q'(s_{k+1}, \mu'(s_{k+1}; \theta^{\mu'}); \theta^{\nu})) - Q(s_k, a_k; \theta^{\nu})]^2 \quad (26)$$

式中: $\theta^{\mu'}$ 是演员目标网络的策略参数; Q' 是评论家目标网络; $\theta^{\nu'}$ 是评论家目标网络的策略参数; ρ 是衰减因子。在线策略网络更新方式为链式法则, 此算法找到最优解的标准就是判断损失函数的导数是否趋于平稳。结合式(22), $L(\theta^{\nu})$ 的导数为

$$\nabla_{\theta^{\nu}} L(\mu) = \frac{1}{2K} \sum_{k=1}^K (\nabla_a Q(s_k, \mu(a_k); \theta^{\nu}) * \nabla_{\theta^{\mu}} \mu(s_k; \theta^{\mu})) \quad (27)$$

另外, 无论是在演员网络还是在评论家网络中, 在线网络的参数通过式(28)更新, 而目标网络的参数更新方式均采用软更新的方式

$$\begin{cases} \theta^{\nu'} \leftarrow \varphi \theta^{\nu} + (1 - \varphi) \theta^{\nu'} \\ \theta^{\mu'} \leftarrow \varphi \theta^{\mu} + (1 - \varphi) \theta^{\mu'} \end{cases} \quad (28)$$

式中: φ 是更新速率, $\varphi \in [0, 1]$; $\leftarrow$ 表示网络参数迭代更新过程, 指示当前迭代中的参数更新由前一轮迭代中的参数计算得来。基于确定性策略梯度的深度强化学习算法具体流程见算法 2。

算法 2 基于确定性策略梯度的深度强化学习算法
输入: 更新速率 τ , 衰减速率 ρ , 评论家在线网络参数 θ^{ν} , 演员在线网络参数 θ^{μ} , 训练回合数 N_i , 训练步长 N_j , 评论家网络学习率 α_{critic} , 演员网络学习率 α_{actor} , 经验回放池大小 B_m , 样本数据大小 B_n
输出: 无人机最优动态部署策略 Δ

1 for $i=0$ to N_i do

```
2  初始化无人机飞行环境；
3  获得初始化状态  $s$ ；
4  for  $j=0$  to  $N_j$  do
5      状态归一化得到  $s_t$ ；
6      根据状态  $s_t$  得到动作  $a_t$ ；
7      获得及时奖励  $r_{t+1}$ ，并转移到  $s_{t+1}$ ；
8      if 经验回放池未满足 then
9          将  $(s_t, a_t, r_{t+1}, s_{t+1})$  存储到经验池；
10     else
11         随机采取  $2K$  个样本；
12         计算评论家目标网络中的  $Q$  值；
13         根据式(26)计算评论家网络中的目标网络与在线网络的损失函数；
14         根据式(27)计算评论家网络中的损失函数梯度；
15         更新演员在线网络的参数  $\theta^{\mu}$ ；
16         根据式(35)更新参数  $\theta^v, \theta^{\prime}$ ；
17     end if
18 end for
19 end for
```

3 仿真与结果分析

3.1 实验参数设置

本文针对部署无人机为海量地面设备服务的典型场景进行仿真实验,地面用户设备随机分布在边长为1的正方形目标区域中。考虑设备数量可变性,设定每隔一定时间在目标区域有新的设备请求边缘计算服务。本文所用深度强化学习神经网络基于TensorFlow 1.13.0框架搭建,采用2张3层的卷积神经网络拟合基于确定性策略梯度算法中演员网络和评论家网络,每层网络的神经元为100。仿真参数设定如表1所示^[23-24]。

表 1 仿真参数设定

Table 1 Simulation parameter setting

参数	数值	参数	数值
N	500	N_i	4 000
χ	3	N_j	100
l/m	400	ϵ	0.000 5
$V/(\text{m} \cdot \text{s}^{-1})$	25	经验回放池大小	10 000
H/m	75	K	64
d_{max}/m	400	ρ	0.9
P_t/W	80	φ	0.001
P_r/W	0.1	C	1 000
σ^2/dB	-100	f_c/GHz	2
f_d/GHz	2	B_m/MHz	1
γ_c/F	10^{-27}	τ/s	20

此外,为评估和验证提出的基于 DDPG 算法的无人机部署方案的性能,与以下几种对比方案进行对比:单无人机巡航动态部署方案,即部署单无人机按规划好的路线服务地面设备;多人机 DQN 动态部署方案,即采用深度强化学习算法优化多无人机部署;多无人机 AC 动态部署方案,即采用演员-评论家(AC)算法动态部署无人机;多无人机 DDPG 动态部署方案,即采用基于确定性策略梯度的深度强化学习算法自主动态部署无人机服务地面设备。

3.2 部署结果和算法收敛性分析

图 3 为激活 20%设备的情况下,DDPG 方案与单无人机方案的无人机轨迹对比。可以看出,由于用户设备数量较少,采用一架无人机即可满足设备的服务质量,而基于 DDPG 部署方案的无人机在寻找悬停位置时,会考虑用户设备的位置,悬停在用户设备较为密集的区域;当设备激活比例达到 60%时,单无人机无法满足设备的服务质量要求,新增无人机 2 进入目标区域,如图 4 所示。由图 4 可见:无人机 2 从(400,400)位置出发与无人机 1 同时为用户设备提供服务;当无人机 1 从位置 2 飞向位置 3 时,无人机 2 会向远离无人机 1 的位置飞行;当无人机 1 从位置 4 飞向位置 5 时,无人机 2 同样也会远离无人机 1。这是因为无人机间存在着干扰,从而影响设备与无人机之间的传输功率,在经过多轮训练后,生成优化的多无人机飞行轨迹,在保证用户体验质量的同时,达到系统能耗最小化的目标。

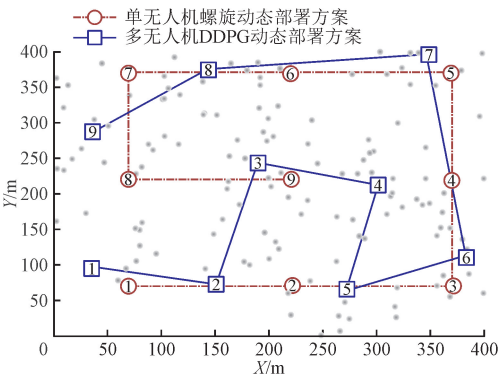


图 3 设备激活 20%场景下无人机动态部署轨迹
Fig. 3 UAV deployment trajectories for 20% device activation

图 5 展示了设备激活比例从 20%到 100%过程中的无人机动态部署轨迹,图中数字表示时隙。当地面激活比例由 20%增加到 8%时,部署无人机 1,轨迹线见绿色线条。当设备激活比例增加为 80%时,部署 2 架无人机为地面设备提供服务,轨迹线为

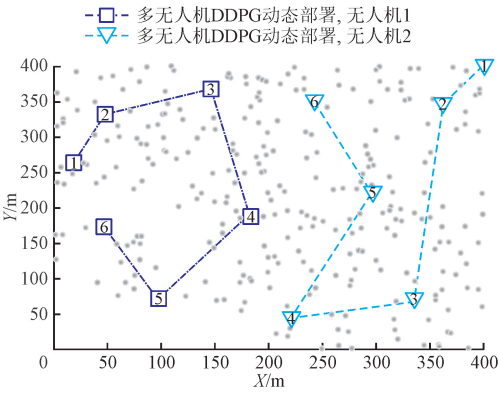


图 4 设备激活 60%场景下无人机动态部署轨迹

Fig. 4 UAV deployment trajectories for 60% device activation

蓝色线条。在第 19 个时隙,无人机 2 从目标区域右上角加入系统,此时无人机 1 和无人机 2 同时进行位置部署优化为用户设备提供服务。当设备 100% 激活时,在第 29 个时隙,两架无人机无法为地面 500 个用户设备提供服务,因此新增部署第 3 架无人机 3(轨迹线为红色线条)由区域右下角起飞进入目标区域提供服务。

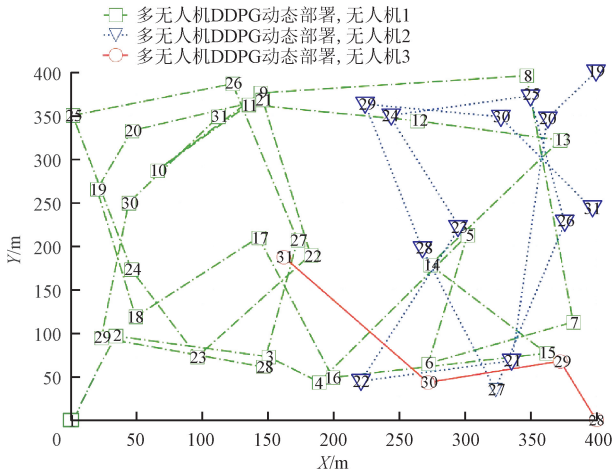


图 5 设备激活 20%到 100%场景下无人机动态部署轨迹

Fig. 5 UAV deployment trajectories for 20%~100% device activation

图 6 显示在相同网络参数下 DDPG 算法、DQN 算法和 AC 算法的训练性能对比。可以看到,AC 算法无法达到收敛状态,因为在该算法中演员网络和评论家网络存在同时更新的问题,演员网络的动作选择取决于评论家网络中的价值函数,但评论家网络较难收敛;DDPG 算法和 DQN 算法的累积奖励都随着回合数的增加而增加,这是由于其均由目

标网络和在线网络组成,可减少不同迭代生成数据之间的关联找到最优策略。此外,DQN 算法相比于 DDPG 算法收敛得更快,但 DDPG 算法收敛时获得的累积奖励为 20,而 DQN 算法收敛时获得的累积奖励为一 22。这是由于 DQN 算法是基于值函数的强化学习算法,不仅动作空间有限,而且容易陷入局部最优。

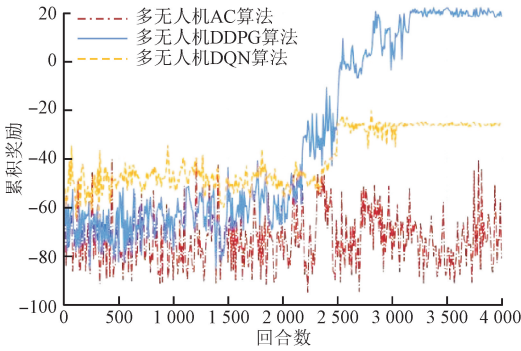


图 6 算法收敛性能比较

Fig. 6 Algorithm convergence performance comparison

3.3 系统能耗分析

图 7 为不同部署方案的系统能耗随用户设备增加的变化趋势。可以看出,当用户设备激活 20%、40%时,基于 DDPG 的单无人机和多无人机方案总能耗相差不大,这是因为在用户设备数量较少时,两种方案均部署单无人机服务地面用户。当用户设备增加到 60%时,单无人机 DDPG 动态部署方案的系统总能耗陡增,这是因为单个无人机在服务目标区域时,会在多个位置进行悬停,移动的次数较多,会消耗大量的无人机能量。当用户设备激活达到 100%时,采用多无人机 DDPG 动态部署方案相比单无人机方案能耗降低约 30%,由此体现出多无人机 DDPG 动态部署方案的优越性。

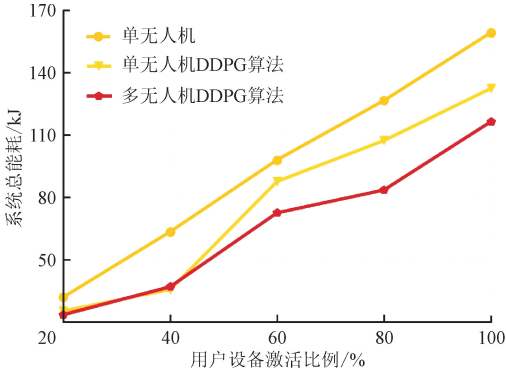


图 7 不同部署方案的系统总能耗

Fig. 7 Total system energy consumption for different deployment schemes

图 8 对比了在不同用户设备数量下无人机采用正交多址接入 (OMA) 和 NOMA 技术的系统总能耗。可以看到,系统总能耗随着用户设备激活比例增加而显著提高,采用 NOMA 技术的系统总能耗低于采用 OMA 技术的能耗。当用户设备激活 20%、40% 时系统能耗差距较小,这是因为一方面,在设备数量较少时 NOMA 技术优势体现不明显;另一方面,通过 DDPG 算法优化了无人机的部署方案,使得系统总能耗最小。随着用户设备的大量增加(激活比例高于 60%),采用基于 NOMA 技术和 DDPG 算法的动态部署方案在能耗方面的优势较为显著。当激活比例达到 100% 时,采用 NOMA 技术的系统总能耗相较于 OMA 系统降低约 40%,进一步说明本文提出的优化部署算法同时适用于 OMA 和 NOMA 系统。

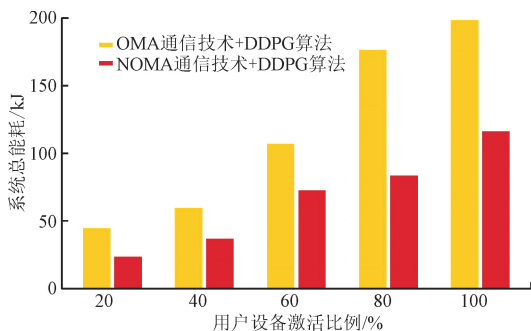


图 8 采用 OMA 和 NOMA 技术的系统能耗比较

Fig. 8 Comparison of energy consumption for OMA and NOMA systems

3.4 用户体验质量满意度分析

图 9 是基于传统地面基站网络和 DDPG 算法动态部署多无人机方案在用户服务质量满意度方面的对比结果。部署基站方案中地面基站与用户设备通信模型和参数见文献[25],部署的基站数量与本文方案所需的无人机数量保持一致。在相同条件(发射功率、频谱资源)下,基于地面基站方案在覆盖范围上低于无人机部署方案,这是由于基于 DDPG 算法的无人机能够灵活移动到最优部署位置以确保服务到大部分用户。当用户设备激活 20% 时,部署一架无人机可以达到用户设备 100% 的满意度,而采用传统的基站服务设备用户满意度仅 30%,大部分用户未得到满意服务。当用户数量继续增多时,本文方案较传统基站方案在用户服务质量满意度上最大提升 60%。图 10 为不同算法用户体验满意度的比较结果。可以看出,在设备激活比例为 20% 时,采用单无人机方案、多无人机 DQN 动态部署方

案、多无人机 DDPG 动态部署方案均可获得较好用户体验质量,而 AC 算法无法收敛,难以满足用户体验质量要求;随着用户数量的增加(激活比例超过 40%),采用优化单无人机飞行轨迹的方案难以满足用户需求;当地面设备激活比例增加到 60% 后,采用多无人机 DQN 动态部署方案难以获得较好的用户体验质量;当设备激活比例增加到 100% 时,所提方案用户服务质量满意度较单无人机场景提高约 50%。

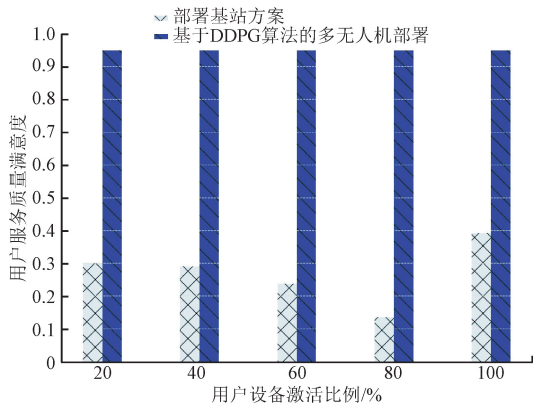


图 9 本文方案与传统地面基站方案与的用户体验满意度比较

Fig. 9 Comparison of user experience satisfaction between the proposed scheme and the traditional ground station deployment approach

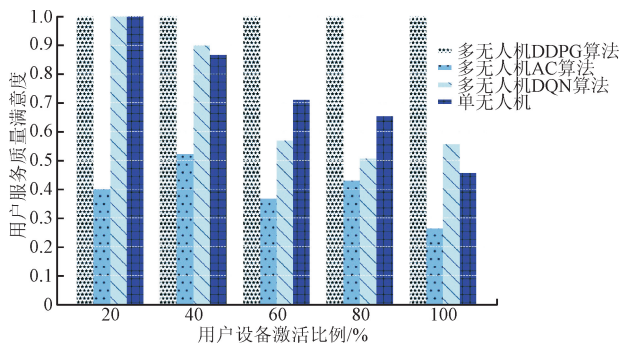


图 10 不同部署方案的用户体验满意度比较

Fig. 10 Comparison of user experience satisfaction for different schemes

4 结 论

(1) 在训练阶段,采用 DDPG 算法和 DQN 算法较 AC 算法有更好的收敛性,其中 DDPG 算法可以在复杂的交互环境中取得更高的累积奖励。

(2) 在系统能耗方面,当用户设备激活低于 40% 时,单无人机、多无人机 DDPG 部署方案系统的总能耗

耗相当;当设备激活超过 60% 时,采用多无人机 DDPG 方案较传统单无人机方案能耗降低约 30%,若进一步采用 NOMA 技术,能耗可降低约 40%。

(3) 在用户体验质量方面,当设备激活低于 20% 时,采用单无人机方案、多无人机 DQN 和 DDPG 部署方案均可满足用户需求;当激活高于 60% 时,多无人机 DDPG 方案能够取得最好效果,较单无人机和传统地面基站方案分别提高约 50% 和 60%。

参考文献:

- [1] 陈新颖, 盛敏, 李博, 等. 面向 6G 的无人机通信综述 [J]. 电子与信息学报, 2022, 44(3): 781-789.
CHEN Xinying, SHENG Min, LI Bo, et al. Survey on unmanned aerial vehicle communications for 6G [J]. Journal of Electronics & Information Technology, 2022, 44(3): 781-789.
- [2] 薛德鑫, 单涛, 徐宁骏, 等. 军用作战无人机未来发展研究 [J]. 指挥控制与仿真, 2022, 44(6): 1-6.
XUE Dexin, SHAN Tao, XU Ningjun, et al. Research on the future development of military combat UAV [J]. Command Control & Simulation, 2022, 44(6): 1-6.
- [3] ZHANG Shangwei, LIU Jiajia, ZHU Yajie, et al. Joint computation offloading and trajectory design for aerial computing [J]. IEEE Wireless Communications, 2021, 28(5): 88-94.
- [4] WANG Liang, WANG Kezhi, PAN Cunhua, 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.
- [5] SHEN Lingfeng, WANG Ning, ZHANG Di, et al. Energy-aware dynamic trajectory planning for UAV-enabled data collection in mMTC networks [J]. IEEE Transactions on Green Communications and Networking, 2022, 6(4): 1957-1971.
- [6] ZHU Botao, BEDEER E, NGUYEN H H, et al. UAV trajectory planning in wireless sensor networks for energy consumption minimization by deep reinforcement learning [J]. IEEE Transactions on Vehicular Technology, 2021, 70(9): 9540-9554.
- [7] TAN Zhenjie, QU Hua, ZHAO Jihong, et al. UAV-aided edge/fog computing in smart IoT community for social augmented reality [J]. IEEE Internet of Things Journal, 2020, 7(6): 4872-4884.
- [8] ZHAN Cheng, HU Han, SUI Xiufeng, et al. Completion time and energy optimization in the UAV-enabled mobile-edge computing system [J]. IEEE Internet of Things Journal, 2020, 7(8): 7808-7822.
- [9] HUANG Zhengrui, CHEN Chongcheng, PAN Miaoxin. Multiobjective UAV path planning for emergency information collection and transmission [J]. IEEE Internet of Things Journal, 2020, 7(8): 6993-7009.
- [10] ZHU Lipeng, ZHANG Jun, XIAO Zhenyu, et al. 3-D beamforming for flexible coverage in millimeter-wave UAV communications [J]. IEEE Wireless Communications Letters, 2019, 8(3): 837-840.
- [11] CHENG Nan, WU Shen, WANG Xiucheng, et al. AI for UAV-assisted IoT applications: a comprehensive review [J]. IEEE Internet of Things Journal, 2023, 10(16): 14438-14461.
- [12] LIU Yi, XIE Shengli, ZHANG Yan. Cooperative offloading and resource management for UAV-enabled mobile edge computing in power IoT system [J]. IEEE Transactions on Vehicular Technology, 2020, 69(10): 12229-12239.
- [13] 夏景明, 刘玉凤, 谈玲. 基于蜂窝网络的多无人机能量消耗最优化解算法研究 [J]. 通信学报, 2023, 44(2): 185-197.
XIA Jingming, LIU Yufeng, TAN Ling. Research on multi-UAV energy consumption optimization algorithm for cellular-connected network [J]. Journal on Communications, 2023, 44(2): 185-197.
- [14] 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). Piscataway, NJ, USA: IEEE, 2016: 1-5.
- [15] BITHAS P S, NIKOLAIDIS V, KANATAS A G, et al. UAV-to-ground communications: channel modeling and UAV selection [J]. IEEE Transactions on Communications, 2020, 68(8): 5135-5144.
- [16] CELIK A, RADAYDEH R M, AL-QAHTANI F S, et al. Resource allocation and cluster formation for imperfect NOMA in DL/UL decoupled HetNets [C]//2017 IEEE Globecom Workshops (GC Wkshps). Piscataway, NJ, USA: IEEE, 2017: 1-6.
- [17] NA Zhenyu, LIU Yue, SHI Jingcheng, et al. UAV-supported clustered NOMA for 6G-enabled internet of things: trajectory planning and resource allocation [J]. IEEE Internet of Things Journal, 2021, 8(20): 15041-15048.
- [18] SHINKUMA R, KISHI N, OTA K, et al. Smarter base station sleeping for greener cellular networks [J].

- IEEE Network, 2021, 35(6): 98-103.
- [19] ANDREA E J H. STELLA: a scheme for a learning machine [J]. IFAC Proceedings Volumes, 1963, 1(2): 497-502.
- [20] WANG Yunpeng, FANG Weiwei, DING Yi, et al. Computation offloading optimization for UAV-assisted mobile edge computing: a deep deterministic policy gradient approach [J]. Wireless Networks, 2021, 27(4): 2991-3006.
- [21] FAN Jianqing, WANG Zhaoran, XIE Yuchen, et al. A theoretical analysis of deep Q-learning [C]//Proceedings of the 2nd Conference on Learning for Dynamics and Control. Chia Laguna Resort, Sardinia, Italy: PMLR, 2020: 486-489.
- [22] TAN Fuxiao, YAN Pengfei, GUAN Xinping. Deep reinforcement learning: from Q-learning to deep Q-learning [C]//Proceedings of the 24th International Conference on Neural Information Processing. Cham: Springer International Publishing, 2017: 475-483.
- [23] LIU Qian, SHI Long, SUN Linlin, 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.
- [24] WANG Liang, WANG Kezhi, PAN Cunhua, et al. Deep reinforcement learning based dynamic trajectory control for UAV-assisted mobile edge computing [J]. IEEE Transactions on Mobile Computing, 2022, 21(10): 3536-3550.
- [25] ZHANG Shangwei, WANG Xiao, SHI Zhenjiang, et al. Reinforcement learning based RSS-threshold optimization for D2D-Aided HTC/MTC in dense NOMA systems [J]. IEEE Transactions on Wireless Communications, 2023, 22(10): 6489-6503.

(编辑 亢列梅)