

空地一体化网络服务连续性保障的低功耗通信与控制联合优化方法

蔡自伟^① 盛敏^{*①} 刘俊宇^① 赵晨曦^① 李建东^{①②}

^①(西安电子科技大学空天地一体化综合业务网全国重点实验室 西安 710071)

^②(鹏城实验室 深圳 518055)

摘要: 空地一体化网络(AGIN)充分利用了空中基站(ABSs)灵活部署的特点,为热点地区提供了按需覆盖与高质量服务。然而,空中基站的高动态性使得网络的服务连续性难以保障。而且,空中基站能量受限,提升服务连续性和降低功耗通常又对应不同的飞行动作,因此,低功耗的服务连续性保障尤为困难。针对上述问题,该文基于联邦深度强化学习(FDRL)提出了一种面向低功耗服务连续性保障的通信与控制联合优化方法。所提方法通过联合优化空中基站的移动控制、用户关联和功率分配来保障网络服务的连续性。针对空中基站的高动态性,通过在所提方法中设计了环境状态经验池来利用信道的时空相关性,并在奖励函数中引入速率方差来保障网络服务连续性。考虑到不同飞行动作的功耗差异,所提方法通过优化空中基站的飞行动作来降低网络功耗。仿真结果表明,该文所提算法在满足用户速率需求和速率方差需求的前提下,能够减小网络功耗,并且所提联邦深度强化学习的性能接近中心式强化学习的性能。

关键词: 空地一体化网络; 服务连续性保障; 低功耗通信与控制联合优化

中图分类号: TN929.5

文献标识码: A

文章编号: 1009-5896(2024)05-1920-11

DOI: 10.11999/JEIT231192

Low-power Communication and Control Joint Optimization for Service Continuity Assurance in Aerial-Ground Integrated Networks

CAI Ziwei^① SHENG Min^① LIU Junyu^① ZHAO Chenxi^① LI Jiandong^{①②}

^①(State Key Laboratory of ISN, Xidian University, Xi'an, 710071, China)

^②(Pengcheng Laboratory, Shenzhen 518055, China)

Abstract: The Aerial-Ground Integrated Networks (AGIN) take full advantage of the flexible deployment of Aerial Base Stations (ABSs) to provide on-demand coverage and high-quality services in hotspot areas. However, the high dynamics of ABSs pose a great challenge to service continuity assurance in AGIN. Furthermore, given the energy constraints of ABSs, ensuring service continuity with low power consumption becomes an increasingly formidable challenge. This is attributed to the inherent contradiction between enhancing service continuity and reducing power consumption, which typically necessitates distinct flight actions. Focusing on the problem mentioned above, a communication and control joint optimization approach based on Federated Deep Reinforcement Learning (FDRL) is proposed to obtain low-power service continuity assurance in AGIN. The proposed approach ensures service continuity by jointly optimizing the flight actions of ABSs, user associations, and power allocation. To cope with the high dynamics of ABSs, an environmental state experience pool is designed to capture the spatiotemporal correlation of channels, and the rate variance is introduced into the reward function to ensure service continuity. Taking into account the power consumption differences associated with various flight actions, the proposed approach optimizes the flight actions of ABSs to reduce their power consumption. Simulation results demonstrate that, under the premise of satisfying requirements for user rate and rate variance, the proposed approach can effectively reduce network power consumption. Additionally, the performance of FDRL is close to that of centralized reinforcement learning.

Key words: Aerial-Ground Integrated Networks (AGIN); Service continuity assurance; Low-power communication and control joint optimization

收稿日期: 2023-10-31; 改回日期: 2023-12-07; 网络出版: 2023-12-18

*通信作者: 盛敏 msheng@mail.xidian.edu.cn

基金项目: 国家重点研发计划(2022YFB2902300), 国家自然科学基金(62121001, 62341111, 62171344), 陕西省重点产业创新链项目(2022ZDLGY05-01, 2022ZDLGY05-06), 鹏城实验室重点项目(PCL2021A15)

Foundation Items: The National Key Research and Development Program of China (2022YFB2902300), The National Natural Science Foundation of China (62121001, 62341111, 62171344), The Key Industry Innovation China of Shaanxi (2022ZDLGY05-01, 2022ZDLGY05-06), The Major Key Project of PengCheng Laboratory(PCL) (PCL2021A15)

1 引言

在空地一体化网络(Aerial-Ground Integrated Networks, AGIN)中,无人机搭载的空中基站(Aerial Base Stations, ABSs)能够为地面用户提供灵活的无线接入服务^[1]。这种网络架构是一种能满足动态流量需求和广域覆盖的有效方案,可解决通信覆盖不足和通信热点地区网络容量限制导致网络服务劣化的问题^[2]。然而,由于空中基站的快速移动,导致网络中的信道状态信息过时。过时的信道状态信息导致资源分配决策迟滞于网络环境的变化,显著恶化网络服务速率的同时,使网络难以提供稳定的服务速率。鉴于空中基站的高动态性,实现网络的服务连续性保障非常困难,并且,为了提升服务连续性通常需要付出高的能耗为代价^[3]。因此,在能量受限的空地一体化网络中,低功耗的服务连续性保障是其实际应用过程中所面临的关键问题。

文献[4-8]研究了提升空地一体化网络服务连续性保障方法,包含提升服务速率和提高服务稳定性两个方面。文献[4]提出了一种迭代优化的方法来联合优化覆盖范围划分、频谱分配和无人机轨迹,最大限度地提高小区边缘用户的最小吞吐量。为解决地面用户速率约束下平均速率最大的问题,文献[5]提出了一种连续凸优化方法通过迭代优化用户关联、子信道分配和轨迹控制来实现平均速率的最大化。进一步考虑信道信息不完美的实际情况,文献[6]进一步提出了一种聚类辅助的强化学习方法来最大化网络的容量。文献[4-6]主要研究了服务速率提升的问题,为了进一步保障服务速率的稳定性,文献[7]提出了一种面向覆盖的计算控制体系结构,用于自适应覆盖结构生成和资源协调。该架构可以灵活调整覆盖结构和可用资源,以确保地面覆盖的空间连续性。由于空中基站的快速移动还会造成网络服务在时间上的不连续,文献[8]考虑了相邻时隙的速率方差,通过设计自适应的强化学习方法来增强空地一体化网络在时空上的覆盖连续性。上述研究已根据空地一体化网络的特点设计出提高服务连续性的方法,但提高服务连续性通常需要更高的功耗为代价。在能量受限的情况下,网络功耗高会严重影响网络服务的持续时间。

文献[9-16]对空地一体化网络的功耗问题进行了研究。文献[9]基于线性状态空间近似和序列凸优化技术提出了一种轨迹控制方案来最大化网络能效。为了解决多样化用户体验约束下的网络能效问题,文献[10]提出了一种基于凸优化的资源分配算法通过交替优化空中基站轨迹、发射功率和带宽分配来

最大化网络能效。结合非正交多址接入,文献[11]提出了一种匹配博弈和连续凸优化的算法来进一步提升网络的能效。文献[9-11]主要采用了凸优化算法,然而,凸优化算法需要实时的信道信息。在空中基站高动态环境下,这些实时信道信息难以获得^[12]。鉴于强化学习方法能够通过学习和改进自身的行为进行资源分配,从而减少对实时信道信息的依赖,强化学习方法已成为空地一体化网络中能效优化的重要手段。文献[13]提出了一种基于强化学习的轨迹控制方法在保障用户公平性的前提下最大化网络能效。考虑空中基站在3维空间的移动,文献[14]提出了一种基于多臂老虎机的多智能体深度强化学习算法,通过优化空中基站轨迹来提升能效。根据空中基站分布式部署的特点,文献[15]提出了分布式强化学习方法利用轨迹控制实现高能效网络的目标。进一步,考虑空中基站能量受限和存储受限的实际情况,文献[16]提出了一个实用的空地一体化网络节能框架,通过结合李雅普诺夫稳定性理论和强化学习算法来完成面向高能效的功率分配和轨迹控制。以上工作在提升网络能效方面取得了较好的效果,然而,高能效设计并不能完全解决能量受限网络的高能耗问题。特别值得注意的是,在空地一体化网络中服务连续性与低功耗之间存在矛盾关系。由于空中基站在飞行过程中的高动态特性,空中基站更倾向于选择悬停在用户上方,以确保网络服务的连续性。然而,悬停相比飞行需要更多的功耗^[17],这导致了服务连续性与低功耗之间的矛盾。为了实现低功耗的服务连续性保障,可以通过合理分配通信资源,结合空中基站移动控制,在保障按需服务连续性的前提下降低网络功耗。因此,如何协调通信资源和移动控制来保障低功耗服务连续性的问题有待进一步研究。

针对上述问题,本文提出一种基于联邦深度强化学习(Federated Deep Reinforcement Learning, FDRL)的低功耗通信与控制资源联合优化方法。通过联合优化通信资源和空中基站的移动控制,解决了服务连续性保障前提下网络功耗最小化问题。在联邦深度强化学习过程中,通过奖励低功耗的飞行动作来达成网络低功耗的目标。同时,设计了信道状态信息经验池,利用历史信息来获取无线信道的相关性,并在奖励函数中引入速率方差来解决空中基站高动态性导致的服务连续性问题。通过联合优化功率分配和移动控制,得到了最优的动作集合来消除低功耗飞行动作对服务连续性保障的影响。分析结果表明,所提方法的计算复杂度随空中基站数量呈线性变化。

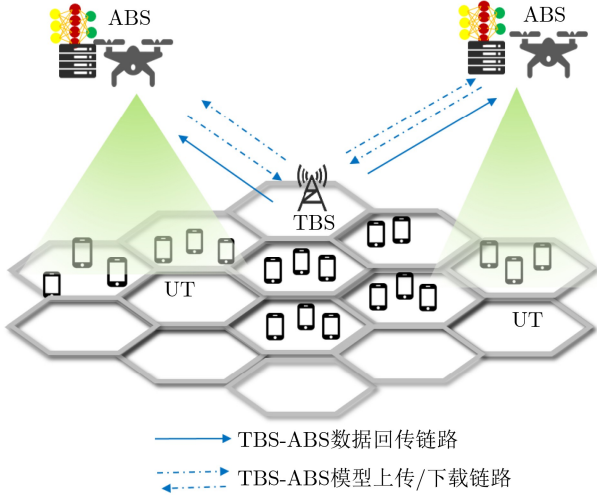


图1 空地一体化网络系统模型图

2 系统模型与问题建模

如图1所示, AGIN, 包含1个地面基站(Terrestrial Base Station, TBS)和 M 个具有一定计算资源的空中基站(ABS)和 K 个地面用户终端(User Terminal, UT)。本文考虑下行传输场景, 当用户的接入请求超过TBS的接入能力时, 将过剩的用户请求卸载到ABS上。ABS在 R 个子信道上采用非正交多址接入(Non-Orthogonal Multiple Access, NOMA)来提升网络的谱效和能效。本文采用笛卡尔坐标系, ABS_m 的位置记为 $\mathbf{ca}_m = (x_m, y_m, z_m)$, 用户 k 的位置记为 $\mathbf{cu}_k = (x_k, y_k, 0)$ 。ABS $_m$ 的飞行高度受限在特定范围内, 即 $z_m \in [z_{\min}, z_{\max}]$ 。用户按照随机游走的方式在地面自由移动。ABS可以在匀速运动中或者悬停时为用户提供无线网络服务。TBS一方面用于网络中智能体的模型聚合。另一方面, ABS与TBS的回传链路建立起空中接入网络与地面核心网络的连接。ABS的计算资源有限, 但能够用于执行深度强化学习算法的本地训练。

2.1 信道模型

本文采用了一个典型的空地信道模型^[18]来描述ABS与用户之间的信道特征。该空地信道模型由路径损耗和阴影衰落两部分组成。路径损耗包含直射路径分量和非直射路径分量, 表示为

$$PL_{m,k,r} = p_{\text{LOS}}L_{\text{LOS}} + p_{\text{NLOS}}L_{\text{NLOS}} \quad (1)$$

其中, p_{LOS} 和 p_{NLOS} 分别表示直射路径和非直射路

$$\beta_{m,\pi(k)}(t) = \frac{p_{m,\pi(k),r}(t)h_{m,\pi(k),r}(t)}{\sum_{d=1, d \neq m}^M \sum_{j=1}^K a_{d,j,r} p_{d,j,r}(t) h_{d,j,r}(t) + \sum_{j \in C_m, j=k+1}^{|C_m|} p_{m,\pi(j),r} h_{m,\pi(j),r}(t) + \delta^2} \quad (7)$$

其中, δ^2 表示噪声功率。因此, 关联到ABS $_m$ 的用户 k 的速率表示为

径的概率。 L_{LOS} 和 L_{NLOS} 分别表示直射路径信号和非直射路径信号的路径损耗。 $PL_{m,k,r}$ 由子信道 r 的频率, ABS $_m$ 的位置和用户 k 的位置决定, 具体计算方法参考文献^[18]。根据空地信号传播特征, 用户的信道增益可表示为

$$h_{m,k,r}(t) = H_{m,k,r}(t) \times 10^{-PL_{m,k,r}} \quad (2)$$

其中, $H_{m,k,r}(t)$ 表示ABS $_m$ 到用户 k 在子信道 r 上的阴影衰落, 它通常被认为服从瑞利分布^[19]。

2.2 信号传输模型

ABS $_m$ 在子信道 r 上的发射信号表示为

$$x_{m,r}(t) = \sum_{k=1}^K a_{m,k,r}(t) \sqrt{p_{m,k,r}(t)} x_{m,k,r}(t) \quad (3)$$

其中, $x_{m,k,r}(t)$ 表示从ABS $_m$ 到用户 k 的数据符号, $p_{m,k,r}(t)$ 表示ABS $_m$ 在子信道 r 上分配给用户 k 的功率, $a_{m,k,r}(t)$ 表示用户与基站子信道的关联关系。 $a_{m,k,r}(t) = 1$ 表示用户 k 关联到ABS $_m$ 的子信道 r 上。 $a_{m,k,r}(t) = 0$ 表示用户 k 没有关联到ABS $_m$ 的子信道 r 上。根据NOMA信号传播特征, 关联到ABS $_m$ 子信道 r 上的用户 k 的接收信号表示为

$$y_{m,k,r}(t) = x_{m,k,r}(t) \sqrt{p_{m,k,r}(t)h_{m,k,r}(t)} + I_{m,k,r}^{\text{inter}} + I_{m,k,r}^{\text{intra}} + \varsigma \quad (4)$$

其中, $x_{m,k,r}(t) \sqrt{p_{m,k,r}(t)h_{m,k,r}(t)}$ 表示有用信号, ς 表示高斯白噪声, $I_{m,k,r}^{\text{inter}}$ 表示其他ABS在子信道 r 上的用户对用户 k 的干扰, 表示为

$$I_{m,k,r}^{\text{inter}} = \sum_{d=1, d \neq m}^M \sum_{j=1}^K a_{d,j,r} x_{d,j,r}(t) \sqrt{p_{d,j,r}(t)h_{d,j,r}(t)} \quad (5)$$

$I_{m,k,r}^{\text{intra}}$ 表示NOMA簇中其他用户对用户 k 的干扰^[18], 表示为

$$I_{m,k,r}^{\text{intra}} = \sum_{j \in C_m, j=k+1}^{|C_m|} x_{m,j,r}(t) \sqrt{p_{m,j,r}(t)h_{m,j,r}(t)} \quad (6)$$

C_m 表示按照信道条件排序后的NOMA簇用户集合, 且 $|C_m|$ 表示集合中的用户数。令 $\pi(k)$ 表示用户 k 的译码顺序, 根据串行干扰消除原理, 用户 $\pi(k)$ 解码之前连续减去所有 $\pi(k-1)$ 用户信号, 然后解码当前信号^[20]。在确定了用户 $\pi(k)$ 接收信号中的有用信号成分和干扰信号成分以后, 用户 k 接收信号的信干噪比 $\beta_{m,\pi(k)}(t)$ 表示为

$$R_{m,k}(t) = B \log_2(1 + \beta_{m,\pi(k)}(t)) \quad (8)$$

其中, B 表示每个用户被分配的带宽。

2.3 功耗模型

ABS_{*m*}的总功耗 $P_m^{\text{tot}}(t)$ 包含飞行保障功耗 $P_m^{\text{F}}(t)$ 和基站工作所需要的功耗 $P_m^{\text{A}}(t)$ 两部分^[9], 表示为

$$P_m^{\text{tot}}(t) = P_m^{\text{F}}(t) + P_m^{\text{A}}(t) \quad (9)$$

考虑搭载ABS的无人机为旋翼无人机, 其移动过程的飞行保障功耗由3部分组成^[17], 表示为

$$P_m^{\text{F}}(t) = P_m^{\text{level}}(t) + P_m^{\text{vert}}(t) + P_m^{\text{drag}}(t) \quad (10)$$

其中, $P_m^{\text{level}}(t)$ 表示水平移动需要的牵引功耗, $P_m^{\text{vert}}(t)$ 表示垂直升降需要的牵引功耗, $P_m^{\text{drag}}(t)$ 表示克服空气阻力所需要的功耗, 具体计算方法参考文献^[17]。ABS悬停的功耗表示为 $P_m^{\text{F}}(t) = W^{3/2}/\sqrt{2\rho A}$, 其中, W 表示ABS承受的重力, ρ 是空气密度, A 是旋翼无人机叶片面积。通过计算可以发现, 不同的飞行动作对应不同的功耗, 比如, 悬停的功耗大于水平中、低速飞行的功耗。

基站的功耗包含静态功耗以及传输功耗两个部分^[21], 表示为

$$P_m^{\text{A}}(t) = P_m^{\text{S}}(t) + P_m^{\text{T}}(t) \quad (11)$$

其中, $P_m^{\text{S}}(t)$ 表示ABS上基站的静态功耗, $P_m^{\text{T}}(t)$ 表示ABS上基站的传输功耗。

2.4 服务连续性

本文参考了文献^[8]对服务连续性的定义。当满足以下条件时, 表示地面用户处于服务连续状态: (1)连续 T 个时隙的速率方差较小, (2)连续 T 个时隙的速率均值较大。

连续 T 个时隙的速率均值可以表示为

$$\bar{X}_m = \frac{1}{I_m \times T} \sum_{t=1}^T \sum_{i=1}^{I_m} R_{m,i}(t), m \in [1, M] \quad (12)$$

其中, I_m 表示关联到ABS_{*m*}的用户数。连续 T 个时隙的速率方差可以表示为

$$V_m(t) = \text{VAR} \left(\frac{1}{I_m} \sum_i^{I_m} R_{m,i}(t) \right), \quad 1 \leq t \leq T, m \in [1, M] \quad (13)$$

其中, 符号VAR表示方差。速率方差表示平均速率的抖动情况。为了完成网络低功耗的目标, 本文在设计时考虑用户需求, 提供按需服务连续性保障。因此, 要求网络满足速率需求即可, 而不是提供更高的速率。

2.5 问题建模

考虑到网络的主要功耗来源是ABS的飞行保障

功耗, 为了实现低功耗的目标, ABS会优先选择低功耗的飞行动作, 例如水平匀速飞行。然而, ABS的移动会导致网络的服务连续性难以保证。为了保障服务连续性, ABS倾向于选择悬停的动作以提供稳定的服务, 但是悬停动作会导致功耗增加。因此, 本文旨在解决满足服务连续性需求前提下的网络功耗最小化问题。鉴于通信资源优化和ABS移动控制在低功耗服务连续性保障方面的协调能力, 该问题被定义为ABS移动控制 J 、用户关联 A 和功率分配 P 的联合优化问题

$$\begin{aligned} \text{P1: } & \min_{J,A,P} \sum_{m=1}^M P_m^{\text{tot}}(t) \\ \text{s.t. } & \text{C1: } 0 < \sum_{r=1}^R \sum_{k=1}^K a_{m,k,r}(t) p_{m,k,r}(t) \leq P_{\max}, \\ & \quad \forall m \in \{1, 2, \dots, M\} \\ & \text{C2: } R_{m,k}(t) \geq R_{\min}, \forall k \in \{1, 2, \dots, K\}, \\ & \quad \forall m \in \{1, 2, \dots, M\} \\ & \text{C3: } V_m(t) \leq V_0, \forall m \in \{1, 2, \dots, M\} \\ & \text{C4: } a_{m,k,r}(t) \in \{0, 1\}, \forall k \in \{1, 2, \dots, K\}, \\ & \quad \forall m \in \{1, 2, \dots, M\}, \forall r \in \{1, 2, \dots, R\} \\ & \text{C5: } \|\mathbf{c}a_i(t) - \mathbf{c}a_j(t)\| \neq 0, \forall j \neq i, \forall i, \\ & \quad j \in \{1, 2, \dots, M\} \end{aligned} \quad (14)$$

约束C1表示每个ABS的发射功率不能超过其被允许的最大值。约束C2表示每个用户的速率不小于用户的速率需求。约束C3表示ABS的服务速率要稳定, 即速率方差要小于给定的门限 V_0 。约束C4限制了1个用户最多只能关联1个ABS。约束C5限制了多个ABS不能同时出现在同一个位置, 从而避免碰撞。由于目标中存在用户关联等离散变量, 上述问题是一个整数非凸的非多项式时间难题, 使用传统方法求解具有较高复杂度。因此, 本文拟设计联邦深度强化学习算法来求解该问题, 该算法将在下一章中进行描述。

3 基于联邦深度强化学习的通信与控制联合优化算法

本文提出的通信与控制联合优化方法包括用户关联、通信与控制的联合优化两个步骤, 它们之间的关系如图2所示。为了在无法获取信道实时信息的情况下实现高效的关联, 本文采用了均值聚类关联算法。另一方面, 基于联邦深度强化学习设计了用于功率分配和ABS移动控制的联合优化算法。ABS可以独立进行本地强化学习训练, 然后在TBS上进行联邦聚合。在TBS上最终聚合的学习模型被下载到各个ABS上, 用于功率分配和ABS移动

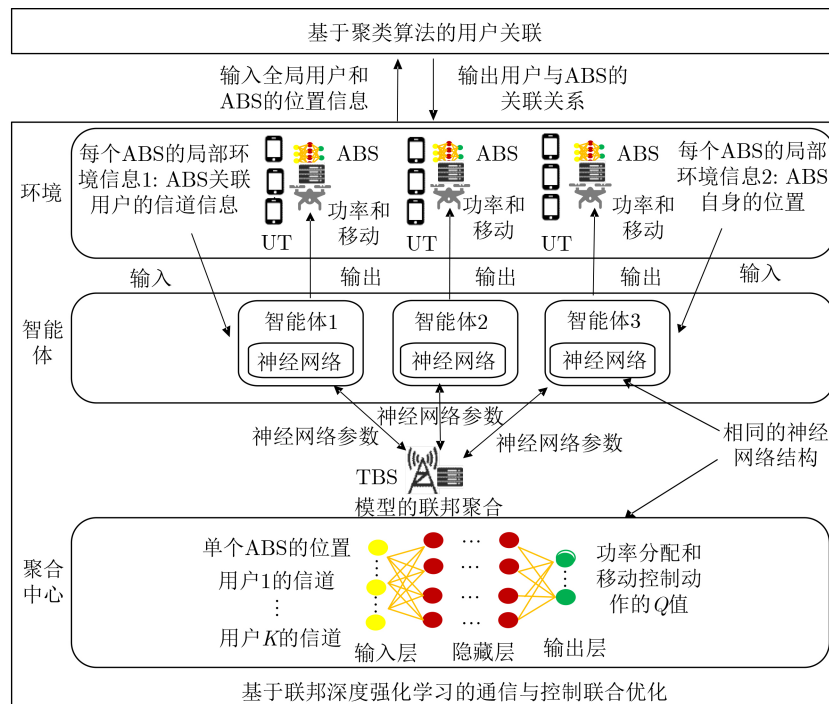


图2 空中基站的通信与控制联合优化算法框图

控制。通过不断地交互执行上述两个步骤，实现优化目标。

3.1 均值聚类关联算法

由于ABS和地面用户都在移动中，信道条件将持续变化，实时获取信道信息极为困难。因此，需要设计一种无需信道信息也能迅速建立用户与基站子信道关联关系的方法。考虑到NOMA用户配对的需求，本文提出了一种基于均值聚类算法的用户关联策略，以便快速将用户与基站子信道关联起来。TBS仅根据掌握的用户位置信息对用户进行聚类，然后采用最近关联规则来确定用户与ABS子信道的关联关系。均值聚类关联算法的具体实现请参见算法1。

3.2 基于联邦深度强化学习的通信与控制联合优化算法

为了实现低功耗的服务连续性保障，本文把ABS的移动控制和功率分配建模成一个马尔可夫决策问题，并采用基于多智能体的深度强化学习算法来解决该马尔可夫决策问题。多智能体强化学习方法的核心在于设计5元组 $\langle S, A, P, R, O \rangle$ 。其中， S 表示环境状态， A 表示动作， P 表示状态转移概率， R 表示奖励， O 表示每个智能体所能观察的局部环境状态， O 是 S 的一个子集。考虑分布式设计，每一个ABS作为一个独立的智能体，用于本地训练深度强化学习算法的深度神经网络。

状态作为决策的基础，必须提供足够的信息，以使ABS选择动作来最大化回报。本文针对该问题

算法1 均值聚类关联算法

- 输入：ABS位置 $\{(x_m, y_m, z_m), m \in [1, M]\}$ ，用户位置 $\{(x_k, y_k, 0), k \in [1, K]\}$ ，ABS数量 M ，用户数量 K ，聚类迭代次数 I
- (1) 随机选择 M 个用户作为 M 簇的簇心，它是所有用户位置的子集
 - (2) for $i = 1, 2, \dots, I$ do
 - (3) for $k = 1, 2, \dots, K$ do
 - (4) 计算用户 k 与每个簇心的距离
 - (5) 把用户划分到距离最小的簇内
 - (6) 重新计算每个簇的簇心
 - (7) end for
 - (8) 如果簇心不再变化，提前结束聚类过程
 - (9) end for
 - (10) 得到用户簇和对应的簇心
 - (11) for $m = 1, 2, \dots, M$ do:
 - (12) 计算空中基站 m 与每个簇心的距离，并将距离存储在距离数组中
 - (13) 找到距离最小的用户簇，并将 m 与其关联
 - (14) 在空中基站 m 的 R 个子信道集合中按顺序选择子信道，记录到用户关联列表中
 - (15) 标记已经选择基站的用户簇，使其不能与其他基站关联
 - (16) end for
 - (17) 输出用户关联列表

设计的状态空间包含当前ABS的位置和关联用户的信道增益。ABS的位置反映了空地一体化网络中基站的部署情况，智能体以此来判断ABS之间的干

扰。ABS能够采集到关联用户的位置信息，并用来估算关联用户的信道增益。信道增益被用作功率分配的依据。

动作空间包含每个ABS的飞行动作和功率分配动作两部分。功率采用离散的功率级别。智能体通过学习选择合适的功率和飞行动作来保障网络的服务连续性。ABS可以执行的动作包括{左、右、前、后、上、下、悬停}这7种动作。不同的飞行动作对应不同的功耗。为了实现低功耗的目标，ABS需要在每一个时隙选择一个低功耗的飞行动作来降低网络功耗，比如水平移动。然而，移动会影响网络的服务连续性。因此，这就需要通过学习选择飞行动作来平衡服务连续性和低功耗。另一方面，还需要联合功率分配来提供按需的服务连续性保障。最终，通过功率分配和飞行动作的联合优化实现低功耗的服务连续性保障。

奖励设计的目的是保障网络服务连续性前提下最小化ABS的总功耗。因此，把ABS总功耗的相反数作为强化学习的即时奖励，即 $R(t) = -P_m^{\text{tot}}(t) - o_1 - o_2$ ，其中， $P_m^{\text{tot}}(t)$ 表示每个ABS的总功耗， o_1 表示违反速率约束的惩罚， o_2 表示违反速率方差约束的惩罚。 o_1 用来保证每个用户的速率需求， o_2 用来约束速率抖动的范围，两者分别从速率需求和速率方差两个方面保障了按需的服务连续性。

智能体选择动作执行 ε 策略，该策略表示为

$$A = \begin{cases} \varphi, & \varepsilon \\ \max Q(s, a, \theta), & 1 - \varepsilon \end{cases} \quad (15)$$

其中， $\varepsilon \in (0, 1)$ 表示随机动作 φ 被选择的概率。通过随机动作的选择，使智能体能够进行探索。同时， $1 - \varepsilon$ 表示选择 Q 值最大动作的概率。令 s 和 a 是策略 π 下的状态动作对，根据Bellman等式^[13]， Q 值函数表示为

$$Q(s, a) = E_{\pi}(R(t+1) + \gamma R(t+2) + \gamma^2 R(t+3) + \dots | S_t = s, A_t = a) \quad (16)$$

其中， $\gamma \in [0, 1]$ 是折扣因子，表示对历史值的加权。 Q 值函数表示智能体在特定状态下执行特定动作后，获得累积奖励的期望值。特别的， $\gamma = 0$ 表示只使用当前的奖励，此时，智能体会显得非常短视，总是忽略长期的奖励。令 $\kappa \in [0, 1]$ 表示学习率， Q 值函数更新的过程是一个学习的过程，表示为

$$Q_{t+1}(s, a) = (1 - \kappa)Q_t(s, a) + \kappa(R(t) + \gamma \max_{a' \in A} Q_t(s', a')) \quad (17)$$

Q 表的大小由状态空间和动作空间决定。当 Q 表的空间太大时，一方面对内存的要求较高，另

外一方面最大值搜索也是非常耗时的。因此，本文设计了神经网络结构来近似 Q 表。在训练深度强化学习网络过程中，把 Q 表的状态作为输入数据， Q 值作为标签。基于输入数据和标签就可以训练一个神经网络，用来根据状态预测 Q 值。训练的目标在于最小化智能体的最小损失函数 $f_i(\theta_i)$ ，表示为

$$f_i(\theta_i) = (y_i(t) - Q_t^i(s, a, \theta_i))^2, i \in [1, M] \quad (18)$$

其中， θ_i 表示神经网络参数， $Q_t^i(s, a, \theta_i)$ 表示预测的 Q 值， $y_i(t)$ 表示真实的 Q 值，表示为

$$y_i(t) = R_i(t) + \gamma \max_{a' \in A} Q_t^i(s', a'), i \in [1, M] \quad (19)$$

针对每个智能体，本文采用全连接深度神经网络的结构，包含1个输入层， F 个隐藏层和1个输出层。本文还设计了一个状态信息池，用于存储包括信道状态信息在内的各种状态信息。在神经网络的训练过程中，每次从状态信息池中采样状态来训练深度神经网络。正是因为这样的设计，使得智能体能够利用到ABS的历史信道信息，从中学习信道的相关性。利用学习到的信道相关性，通过实时的功率分配和移动控制来减小速率方差。因此，这种设计使得本文提出的方法能够解决信道信息过时带来的服务连续性保障问题。

为了获得全局最优的通信和控制资源分配方案，通常可以选择中心式强化学习算法^[13]或分布式强化学习算法^[15]。中心式强化学习算法需要将所有智能体的状态和动作传输到一个智能体，这需要传输大量的原始数据。相比之下，分布式强化学习算法则存在收敛较慢的问题。为了综合解决上述问题，本文设计了联邦聚合机制并应用到深度强化学习算法。在该算法中，每个ABS可以通过联邦聚合后的模型感知其他ABS的位置、用户关联和功率分配信息，因此，能取得接近中心式强化学习的效果，并且收敛速度快。不同于中心式强化学习，该算法只需要让每个智能体把深度神经网络的模型参数 $\theta_i, i \in [1, M]$ 传输到用于模型聚合的中心服务器，而不需要传输原始训练数据，因此需要传输的数据量较小。通过对各个智能体深度神经网络的模型完成聚合，形成一个全局的深度神经网络模型。在TBS执行联邦模型聚合的过程表示为

$$\theta_g = \sum_{i=1}^M w_i \theta_i \quad (20)$$

其中， w_i 表示智能体 i 的模型对全局模型的贡献。联邦聚合的目标是最小化模型损失，表示为

$$P2 : \min_{\theta_i} \sum_{i=1}^M w_i f_i(\theta_i) \quad (21)$$

具体而言,每个ABS根据低功耗和服务连续性保障的需求,独立地分配功率和移动控制。每个智能体根据其观察采取行动,并从环境中接收奖励。每个智能体的观察包括ABS的位置、用户状态信息、用户数据速率和各自的功耗。在收集了训练数据后,用于独立训练每个智能体的深度神经网络。然后,每个智能体将深度神经网络的参数上传到TBS进行模型聚合。聚合完成后,智能体下载全局模型在各个智能体中进行分布式执行,并开始新的联邦学习周期。联邦深度强化学习算法的流程总结在算法2中。

算法2中算法的模型聚合采用了联邦平均方法(Federated Average, FedAvg)^[22],根据文献[22]对FedAvg方法收敛性的证明,本文所设计的联邦深度强化学习算法在迭代次数足够大的情况下能够实现收敛。

3.3 复杂度和通信开销分析

3.3.1 计算复杂度分析

本文对所提联邦深度强化学习算法的计算复杂度进行了分析。首先分析深度强化学习部分的计算复杂度。对于一个全连接神经网络而言,每个时隙的计算复杂度表示为 $\sigma_1 = o(|A| \times |S| \times M)$,其中, $|A|$ 表示动作空间的大小, $|S|$ 表示状态空间的大小。考虑到训练的幕数 E 和每一幕的时隙数 T ,所以总的计算复杂度表示为 $E \times T \times \sigma_1$ 。然后考虑

执行联邦聚合部分的计算复杂度。深度神经网络的层数表示为 N ,每一层的节点数表示为 J_n ,那么联邦聚合部分的计算复杂度为 $\sigma_2 = \sum_{n=1}^N o(J_n M)$ 。联邦深度强化学习的计算复杂度是深度强化学习部分和联邦聚合部分计算复杂度的总和,表示为 $E \times T \times \sigma_1 + (E \times T \times \sigma_2)/G$ 。由于联邦聚合部分的计算复杂度远小于深度神经网络训练部分,所以所提算法的计算复杂度主要取决于深度强化学习部分的复杂度。由于深度强化学习中每个智能体只需要处理局部信息,所以状态空间较小,从而使深度强化学习部分的计算复杂度也较小。从计算复杂度总和的表达式可以看出,本文所提算法的计算复杂度随空中基站的数量线性变化。

3.3.2 通信开销分析

在中心式强化学习算法中,ABS将本地观测的信息上传到中心服务器,并从中心服务器获取到需要执行的动作。通信开销表示为

$$D_{\text{CDRL}} = E \times T \times \sum_{m=1}^M (|s_m| + |a_m|) \quad (22)$$

其中, $|s_m|$ 表示每个智能体进行深度神经网络训练需要的原始数据大小, $|a_m|$ 表示每个智能体下载的动作空间大小。

在本文所提联邦深度强化学习算法中,ABS将本地训练的模型上传到中心服务器。并从中心服务

算法2 联邦深度强化学习算法

输入:聚合间隔 G ,折扣因子 γ ,学习率 κ ,探索因子 ε ,智能体的数量 M ,用于训练的幕数 E ,单幕的时隙数 T

- (1) 初始化环境参数,包含ABS和用户的位置,以及信道增益
- (2) 给每一个智能体和全局模型设置初始化的深度神经网络参数
- (3) for $i = 1, 2, \dots, E$ do
- (4) for $t = 1, 2, \dots, T$ do
- (5) for $m = 1, 2, \dots, M$ do
- (6) 步骤1:从聚合服务器获取全局的模型参数 θ^g ,初始化智能体 m 的神经网络
- (7) 步骤2:从网络环境中获取状态参数 $s(t)$
- (8) 步骤3:根据 ε 和式(15)选择动作 $a(t)$
- (9) 步骤4:执行动作 $a(t)$,获得下一个状态 $s(t+1)$ 和 $R(t)$
- (10) 步骤5:把 $a(t)$, $s(t)$, $s(t+1)$, $R(t)$ 记录到环境信息经验池中
- (11) 步骤6:当采样数满足训练要求时,从环境信息经验池中选择样本训练深度神经网络训练模型 θ_m ,否则返回
- (12) end for
- (13) if $(t\%G == 0)$
- (14) M 个智能体传输深度神经网络参数到TBS,执行FedAvg聚合,把聚合后的模型传输给各个智能体分布式执行
- (15) end if
- (16) end for
- (17) end for
- (18) 输出ABS移动控制和功率分配的动作

器下载神经网络模型，利用该模型在ABS上预测最优的动作。该算法主要的通信开销来自模型上传下载阶段，表示为

$$D_{\text{FDRL}} = \frac{2E \times T}{G} \times \sum_{m=1}^M (|\theta_m|) \quad (23)$$

其中， G 表示模型更新间隔， $|\theta_m|$ 代表每个智能体本地神经网络模型参数的大小。由于神经网络模型参数大小 $|\theta_m|$ 通常远小于状态空间大小 $|s_m|$ ，所以本文所提算法的通信开销要远低于中心式强化学习算法。

4 仿真结果与讨论

本节使用数值仿真验证了基于联邦深度强化学习的低功耗通信与控制联合优化方法的有效性。本文设置的覆盖范围为 $3\ 000\ \text{m} \times 3\ 000\ \text{m}$ 。在该区域内放置1个TBS和默认3个ABS。每个ABS使用100个子信道，每个子信道的带宽为180 kHz。每个子信道可以被2个NOMA用户共用，每个用户的速率需求默认设置为300 kbit/s，可根据需要调整。用户速率方差门限默认设置为900。无人机的最大飞行高度为300 m，最小飞行高度为100 m，初始位置随机分布。无人机的飞行速度设置为10 m/s，其他与飞行功耗计算有关的参数参考文献[17]。在联邦深度强化学习算法中，神经网络采用四层全连接结构。神经网络的优化器选用均方根传播(Root Mean Square propagation, RMSprop)。联邦聚合的间隔被设置为10。联邦强化学习算法中动作选择依赖参数 ϵ ，它被设置为 $[0.1, 0.9]$ 区间内线性递减的值。折扣因子 $\gamma = 0.9$ ，学习率 $\kappa = 0.1$ 。

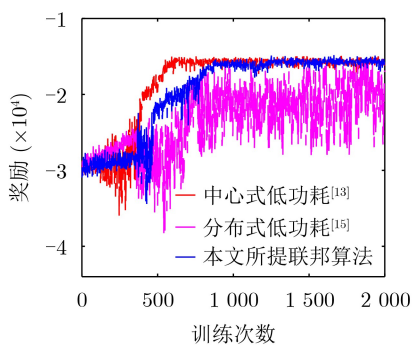
本文把所提方法与现有文献中的先进方法进行了对比研究。其中，提升速率与减小方差均衡方案的目标是最大化网络速率并最小化速率方差[8]。本文还把中心式强化学习算法[13]和分布式强化学习算法[15]使能的低功耗方案作为本文提出的联邦深度强化学习算法性能仿真的对比方案。

本文首先分析了联邦学习算法在低功耗优化方案中的收敛性能，如图3所示。

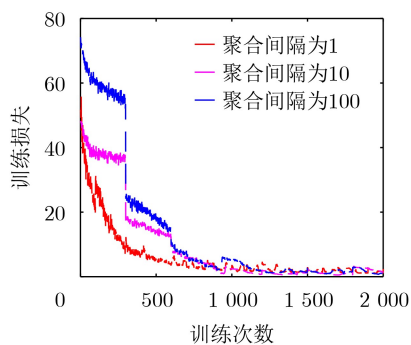
图3(a)展示了本文提出的算法与对比算法在收敛性能上的比较结果。从图3(a)可以看出，本文设计的联邦深度强化学习算法比分布式强化学习算法更易收敛，并且接近于中心式强化学习算法的收敛性能。从图3(a)中还可以看到，联邦深度强化学习算法获得了比分布式强化学习算法更高的奖励，尽管需要付出少量的通信开销为代价。图3(b)展示了联邦聚合过程中不同聚合间隔对收敛性能的影响。从图3(b)可以看出，随着模型聚合间隔的减小，模型收敛得越快。接下来，将提升速率与减小方差均衡方案与低功耗服务连续性方案进行对比。

由图4(a)可见，本文提出的低功耗连续性保障方案的吞吐量小于提升速率与减小方差均衡方案，但两者均满足网络吞吐量需求。由图4(b)可见，本文提出的低功耗连续性保障方案需要更少的功耗。因此，所提方法实现了低功耗服务连续性保障的目标。由图4(b)还可以观察到，在保持相同的空中基站数量的情况下，网络的功耗随着速率需求的提高而增加。网络整体功耗的上升来源于无人机飞行动作的改变和基站发射功率的调整。换言之，除了规划出功率最小的移动控制动作外，还需要适时增强功率输出，以确保网络服务连续性保障和用户速率需求。网络吞吐量反映了网络在特定时刻满足速率需求的能力。进一步，通过速率方差可以更好地刻画网络在多个时隙内的服务稳定性。速率方差越小表示用户平均速率的波动越小。

图5展示了用户速率方差与网络功耗随空中基站数量的变化趋势。从图5可以看出，无论ABS数量如何变化，低功耗服务连续性保障方案在速率方差和功耗两个指标上都取得了更小的值，表现出更好的性能。这说明了提出的算法具备很好的ABS规模化扩展能力。另外，从图5中还可以观察到，相较于分布式低功耗方案，本文所提联邦低功耗方案



(a) 不同强化学习算法收敛性能比较



(b) 不同聚合间隔的联邦学习收敛性能比较

图3 联邦深度强化学习的收敛性能

在功耗和速率方差方面均取得了更小的值，并接近中心式低功耗方案的性能。

图6展示了网络功耗和实际速率方差随着速率方差门限的变化趋势。从图6可以看出，实际的速率方差随速率方差门限的增加而增加，并且优化后的实际速率方差满足速率方差门限的要求。并且，随着速率方差门限的增加，网络功耗逐渐减少。速率方差的门限会影响网络功耗，是因为空中基站的飞行机动决策会受到速率方差门限的影响。门限增加意味着放宽约束，所以功耗可以降低。

5 结束语

本文针对空地一体化网络，提出一种基于联邦深度强化学习的服务连续性保障低功耗通信与控制联合优化方法。为实现低功耗的服务连续性保障，本文构建了基于通信资源和控制资源的联合优化模型，包括对ABS移动控制、用户关联和功率分配的联合优化。在联邦深度强化学习中引入了速率方差的概念，以保障无线服务在时间上的连续性，并通过空中基站的按需移动来实现低功耗的目标。另外，本文提出的联邦深度强化学习更符合ABS分布

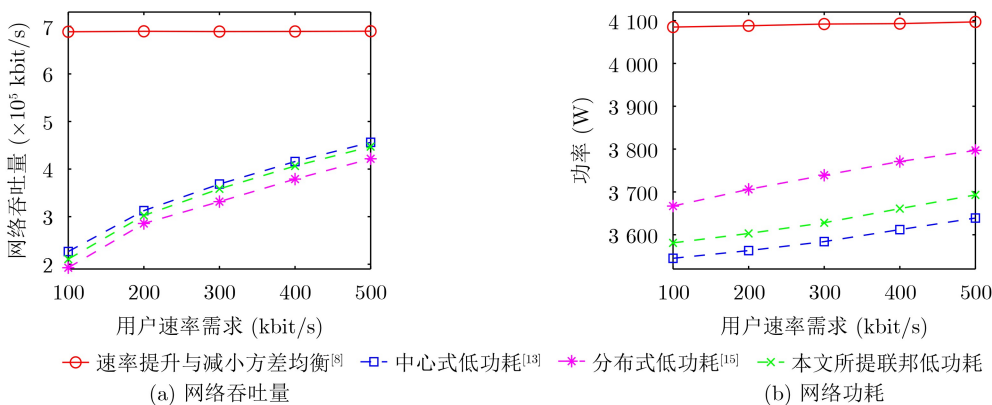


图4 网络吞吐量和网络功耗随用户速率需求的变化趋势

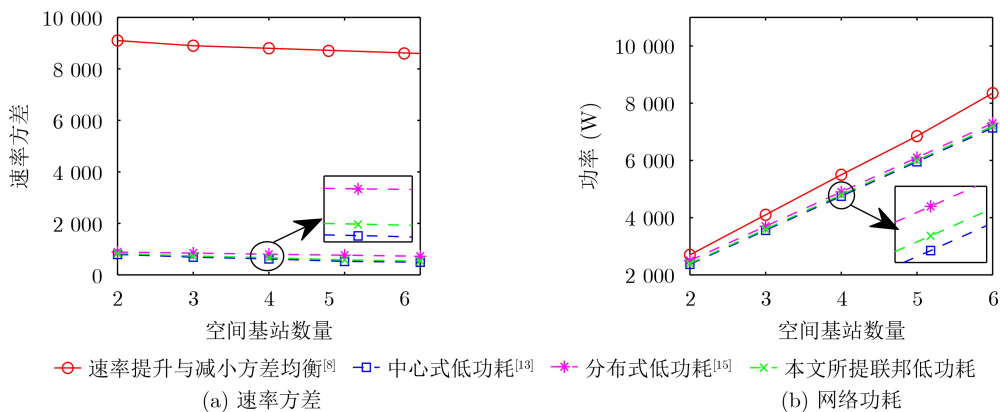


图5 用户速率方差与网络功耗随空中基站数量的变化趋势

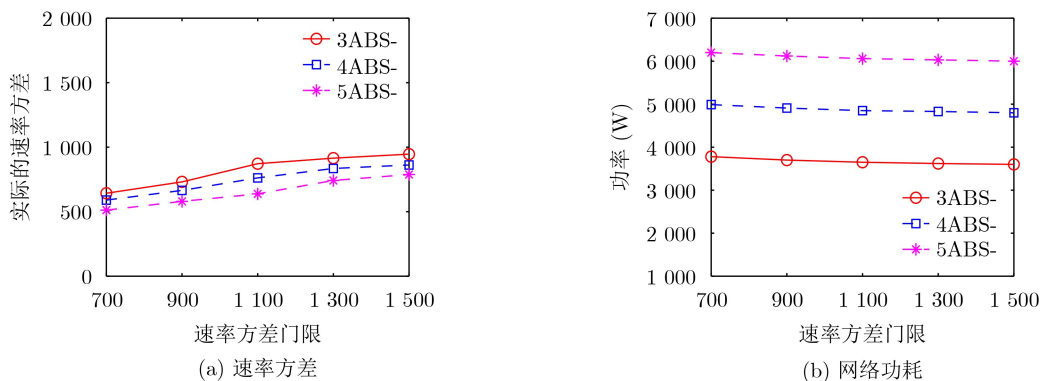


图6 实际速率方差和网络功耗随速率方差门限的变化趋势

式特性和资源受限的特点。仿真和分析结果还说明了所提的方法具备低功耗服务连续性保障的能力，具有收敛快、计算复杂度低和通信开销小的特点。本文提出的面向低功耗的服务连续性保障联合优化方法为空地一体化网络的发展提供了有力支持。在未来的研究中，可以考虑太阳能供电以进一步提高空地一体化网络的服务能力。

参 考 文 献

- [1] LYU Jiangbin, ZENG Yong, and ZHANG Rui. UAV-aided offloading for cellular hotspot[J]. *IEEE Transactions on Wireless Communications*, 2018, 17(6): 3988–4001. doi: [10.1109/TWC.2018.2818734](https://doi.org/10.1109/TWC.2018.2818734).
- [2] MOZAFFARI M, SAAD W, BENNIS M, *et al.* A tutorial on UAVs for wireless networks: Applications, challenges, and open problems[J]. *IEEE Communications Surveys & Tutorials*, 2019, 21(3): 2334–2360. doi: [10.1109/COMST.2019.2902862](https://doi.org/10.1109/COMST.2019.2902862).
- [3] JIANG Xu, SHENG Min, ZHAO Nan, *et al.* Green UAV communications for 6G: A survey[J]. *Chinese Journal of Aeronautics*, 2022, 35(9): 19–34. doi: [10.1016/j.cja.2021.04.025](https://doi.org/10.1016/j.cja.2021.04.025).
- [4] LI Yang, ZHANG Heli, JI Hong, *et al.* UAV-assisted cellular communication: Joint trajectory and coverage optimization[C]. 2021 IEEE Wireless Communications and Networking Conference (WCNC), Nanjing, China, 2021: 1–6. doi: [10.1109/WCNC49053.2021.9417264](https://doi.org/10.1109/WCNC49053.2021.9417264).
- [5] NGUYEN M D, LE L B, and GIRARD A. Integrated UAV trajectory control and resource allocation for UAV-based wireless networks with Co-channel interference management[J]. *IEEE Internet of Things Journal*, 2022, 9(14): 12754–12769. doi: [10.1109/JIOT.2021.3138374](https://doi.org/10.1109/JIOT.2021.3138374).
- [6] ZHOU Shiyang, CHENG Yufang, LEI Xia, *et al.* Resource allocation in UAV-assisted networks: A clustering-aided reinforcement learning approach[J]. *IEEE Transactions on Vehicular Technology*, 2022, 71(11): 12088–12103. doi: [10.1109/TVT.2022.3189552](https://doi.org/10.1109/TVT.2022.3189552).
- [7] XIE Ziwen, LIU Junyu, SHENG Min, *et al.* Exploiting aerial computing for air-to-ground coverage enhancement[J]. *IEEE Wireless Communications*, 2021, 28(5): 50–58. doi: [10.1109/MWC.211.2100048](https://doi.org/10.1109/MWC.211.2100048).
- [8] LI Jiandong, ZHOU Chengyi, LIU Junyu, *et al.* Reinforcement learning based resource allocation for coverage continuity in high dynamic UAV communication networks[J]. *IEEE Transactions on Wireless Communications*, 2023. doi: [10.1109/TWC.2023.3282909](https://doi.org/10.1109/TWC.2023.3282909).
- [9] ZENG Yong and ZHANG Rui. Energy-efficient UAV communication with trajectory optimization[J]. *IEEE Transactions on Wireless Communications*, 2017, 16(6): 3747–3760. doi: [10.1109/TWC.2017.2688328](https://doi.org/10.1109/TWC.2017.2688328).
- [10] ZENG Fanzi, HU Zhenzhen, XIAO Zhu, *et al.* Resource allocation and trajectory optimization for QoE provisioning in energy-efficient UAV-enabled wireless networks[J]. *IEEE Transactions on Vehicular Technology*, 2020, 69(7): 7634–7647. doi: [10.1109/TVT.2020.2986776](https://doi.org/10.1109/TVT.2020.2986776).
- [11] LI Yabo, ZHANG Haijun, LONG Keping, *et al.* Joint resource allocation and trajectory optimization with QoS in NOMA UAV networks[C]. GLOBECOM 2020 - 2020 IEEE Global Communications Conference, Taipei, China, 2020: 1–5. doi: [10.1109/GLOBECOM42002.2020.9322351](https://doi.org/10.1109/GLOBECOM42002.2020.9322351).
- [12] CUI Yuling, DENG Danhao, WANG Chaowei, *et al.* Joint trajectory and power optimization for energy efficient UAV communication using deep reinforcement learning[C]. IEEE INFOCOM 2021 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Vancouver, Canada, 2021: 1–6. doi: [10.1109/INFOCOMWKSHPS51825.2021.9484490](https://doi.org/10.1109/INFOCOMWKSHPS51825.2021.9484490).
- [13] LIU C H, CHEN Zheyu, TANG Jian, *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. doi: [10.1109/JSAC.2018.2864373](https://doi.org/10.1109/JSAC.2018.2864373).
- [14] ARANI A H, AZARI M M, HU Peng, *et al.* Reinforcement learning for energy-efficient trajectory design of UAVs[J]. *IEEE Internet of Things Journal*, 2022, 9(11): 9060–9070. doi: [10.1109/JIOT.2021.3118322](https://doi.org/10.1109/JIOT.2021.3118322).
- [15] GALKIN B, OMONIWA B, and DUSPARIC I. Multi-agent deep reinforcement learning for optimising energy efficiency of fixed-wing UAV cellular access points[C]. ICC 2022 - IEEE International Conference on Communications, Seoul, Korea, 2022: 1–6. doi: [10.1109/ICC45855.2022.9838809](https://doi.org/10.1109/ICC45855.2022.9838809).
- [16] KAKARLA N B and MAHENDRAN V. Lyapunov meets thompson: Learning-based energy-efficient UAV communication with queuing stability constraints[C]. IEEE INFOCOM 2023 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Hoboken, USA, 2023: 1–7. doi: [10.1109/INFOCOMWKSHPS57453.2023.10225765](https://doi.org/10.1109/INFOCOMWKSHPS57453.2023.10225765).
- [17] SUN Yan, XU Dongfang, NG D W K, *et al.* Optimal 3D-trajectory design and resource allocation for solar-powered UAV communication systems[J]. *IEEE Transactions on Communications*, 2019, 67(6): 4281–4298. doi: [10.1109/TCOMM.2019.2900630](https://doi.org/10.1109/TCOMM.2019.2900630).
- [18] WANG Ningyuan, LI Feng, CHEN Dong, *et al.* NOMA-based energy-efficiency optimization for UAV enabled space-air-ground integrated relay networks[J]. *IEEE Transactions on Vehicular Technology*, 2022, 71(4): 4129–4141. doi: [10.1109/TVT.2022.3151369](https://doi.org/10.1109/TVT.2022.3151369).
- [19] LUONG P, GAGNON F, TRAN L N, *et al.* Deep

- reinforcement learning-based resource allocation in cooperative UAV-assisted wireless networks[J]. *IEEE Transactions on Wireless Communications*, 2021, 20(11): 7610–7625. doi: [10.1109/TWC.2021.3086503](https://doi.org/10.1109/TWC.2021.3086503).
- [20] DE SENA A S, LIMA F R M, DA COSTA D B, *et al.* Massive MIMO-NOMA networks with imperfect SIC: Design and fairness enhancement[J]. *IEEE Transactions on Wireless Communications*, 2020, 19(9): 6100–6115. doi: [10.1109/TWC.2020.3000192](https://doi.org/10.1109/TWC.2020.3000192).
- [21] FIORANI M, TOMBAZ S, FARIAS F S, *et al.* Joint design of radio and transport for green residential access networks[J]. *IEEE Journal on Selected Areas in Communications*, 2016, 34(4): 812–822. doi: [10.1109/JSAC.2016.2544599](https://doi.org/10.1109/JSAC.2016.2544599).
- [22] WEI Xizixiang and SHEN Cong. Federated learning over noisy channels: Convergence analysis and design examples[J]. *IEEE Transactions on Cognitive Communications and Networking*, 2022, 8(2): 1253–1268. doi: [10.1109/TCCN.2022.3140788](https://doi.org/10.1109/TCCN.2022.3140788).
- 蔡自伟: 男, 博士生, 研究方向为空地一体化网络、低能耗移动通信网络.
- 盛 敏: 女, 博士, 教授, 研究方向为空天地一体化网络、智能无线网络、移动自组织网络.
- 刘俊宇: 男, 博士, 教授, 研究方向为超密无线网络、空地一体化网络、低能耗移动通信网络.
- 赵晨曦: 男, 博士, 副教授, 研究方向为6G无线网络、空地一体化网络、低能耗移动通信网络.
- 李建东: 男, 博士, 教授, 研究方向为宽带无线通信、认知无线网络、大规模自组织网络.

责任编辑: 余 蓉