

面向低空通信的无蜂窝大规模 MIMO 智能关联及功控方法

张琦¹, 贾旭日¹, 蔡曙¹, 张军¹, 吴怡²

(1. 南京邮电大学通信与信息工程学院, 江苏 南京 210003; 2. 福建师范大学光电与信息工程学院, 福建 福州 350117)

摘要: 随着低空经济快速发展, 无蜂窝大规模多输入多输出 (Multiple-Input Multiple-Output, MIMO) 的通信支撑作用日益重要。针对传统基于瞬时位置的聚类方法更新滞后、固定功率控制难以适应位置动态变化的问题, 本文构建了无人机-接入点关联与上行功率控制的联合优化问题, 并将其建模为马尔可夫决策过程, 提出基于深度确定性策略梯度的两阶段交替优化算法, 将无人机整个飞行周期的大尺度衰落序列作为全局先验输入, 实现整个飞行周期的聚类与功控序列的联合规划学习。仿真结果表明, 所提算法相比传统方案能够获得更高且稳定的系统性能, 并在多种无人机空间分布拓扑下表现出良好的适应性。

关键词: 无蜂窝大规模 MIMO; 无人机; 强化学习; 关联; 功率控制

中图分类号: TN929.5

文献标志码: A

doi: TXXB260251

An Intelligent Association and Power Control Method for Cell-free Massive MIMO in the Low-Altitude Communications

Zhang Qi¹, Jia Xuri¹, Cai Shu¹, Zhang Jun¹, Wu Yi²

1. School College of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

2. College Optoelectronics and Information Engineering, Fujian Normal University, Fuzhou 350117, China

Abstract: With the rapid development of the low-altitude economy, the communication support role of cell-free massive multiple-input multiple-output (MIMO) systems has become increasingly important. To address the limitations of conventional clustering methods based on instantaneous locations, which were subject to delayed updates, and fixed power control schemes, which were unable to adapt to dynamic position variations, the joint optimization problem of unmanned aerial vehicle (UAV) –access point association and uplink power control was formulated as a Markov decision process. A two-stage alternating optimization algorithm based on deep deterministic policy gradient was then proposed, in which the large-scale fading sequence over the entire UAV flight cycle was taken as global prior input to enable joint planning-oriented learning of clustering and power control sequences throughout the full flight period. Simulation results showed that the proposed algorithm achieved higher and more stable system performance than conventional schemes and exhibited good adaptability under various UAV spatial topology distributions. These results demonstrate that the proposed method provides an effective solution for joint UAV–access point association and uplink power control in dynamic low-altitude communication scenarios.

Key words: Cell-free massive MIMO, UAV, Reinforcement Learning, Association, Power Control

收稿日期: 2026-05-06; 修回日期: 2026-07-07

通信作者: 张琦, zhangqi1212@njupt.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62571267); 江苏省重点研发计划子课题 (No.BE2023022-3)

Foundation Items: National Natural Science Foundation of China (No. 62571267), Sub-project of the Jiangsu Provincial Key Research and Development Program (No. BE2023022-3)

0 引言

随着 2024 年低空经济首次被写入《政府工作报告》，我国无人机产业随即开始了迅猛发展。在空天地一体化通信的发展趋势下，无人机 (unmanned aerial vehicle, UAV) 作为重要的信息采集与通信节点，正逐步从辅助性设备转变为网络体系的关键组成部分^[1-3]，其在物流配送、农业生产、应急救援、城市基础建设和维护等领域发挥着越来越重要的作用^[4-8]。无蜂窝大规模多输入多输出 (Multi-Input Multi-Output, MIMO) 是第六代移动通信系统 (6th generation wireless systems, 6G) 的主要发展方向之一。该系统内广泛分布的接入点 (access points, AP) 通过回传链路与中央处理单元 (central processing unit, CPU) 相连，共同为所有用户提供服务^[9]。与传统蜂窝网络不同，无蜂窝大规模 MIMO 打破了固定小区边界的限制，不会因频繁的小区切换而产生中断或信号衰减，从而在移动过程中能够持续保持稳定的通信连接，满足 UAV 在大范围、动态、高速移动场景下的高可靠通信需求，实现了真正意义上的连续覆盖与无缝服务，为低空经济下的 UAV 群通信提供重要支撑^[10]。然而在实际部署中，仍然有着许多挑战。一方面，在通信过程中高速移动的 UAV 导致大尺度衰落落在时间与空间上快速变化，所提供服务的 AP 集合需要频繁重构；另一方面，UAV 电池容量有限使得功率控制变得更为关键：功率过低会削弱有效信号强度，影响通信质量，功率过高则会加速能耗导致飞行时间减短并加大用户之间的同频干扰。

为了解决上述问题，近年来研究者提出多种用户-接入点关联方法。文献[11]最早提出了以用户为中心的思想，即每个用户由与其链路最强的若干 AP 组成的“小型虚拟小区”来服务，从而降低回程开销并提高大多数用户速率。文献[12]在此基础上面向毫米波无蜂窝大规模 MIMO 提出波束对准方法，利用 6GHz 以下的频段建立控制信道，通过估计用户与 AP 间的最强传播路径来优化关联。文献[13]引入干扰感知奖励衡量用户与 AP 链路质量，并采用迭代优化算法提升系统公平性和频谱效率。文献[14]提出自适应模糊逻辑的 AP 选择方案，在以用户为中心的框架下综合考虑大尺度衰落、小尺度衰落、AP 分配功率及用户间干扰等因素，实现兼顾可扩展性与精细化的关联策略。文献[15]采用

分数规划、拉格朗日对偶和惩罚函数等工具，对用户-AP 关联与功率因子进行联合优化，在约束前传负载的条件下提升了系统总吞吐量。在上述以凸优化理论为主的工作基础上，近年来也有学者尝试将强化学习 (reinforcement learning, RL) 引入无蜂窝大规模 MIMO 系统中。文献[16]提出了一种多智能体强化学习 (deep reinforcement learning, DRL) 框架，通过共享策略网络实现上行功控，允许用户动态加入或离开，从而降低训练与部署开销。文献[17]面向地面低速移动用户，构建了联邦多智能体 RL 框架，将 AP 建模为智能体，通过本地训练和参数聚合实现 AP 聚类策略。仿真结果表明，该方法在不同网络拓扑下均优于传统凸优化算法，且学习到的网络参数能够较好地迁移到新环境中，展现出良好的泛化能力。文献[18]则以最大化用户速率的比例公平性为目标，引入深度确定性策略梯度 (deep deterministic policy gradient, DDPG) 算法，对无蜂窝大规模 MIMO 网络中的协作聚类与下行功率控制进行联合优化，在保证公平性的同时提升系统吞吐量。现有的 RL 方案大多面对的是地面静止用户，仅依赖当前时隙的瞬时大尺度衰落系数进行局部的静态决策，难以在用户完整运动周期尺度上形成具有前瞻性的全局策略。在高机动 UAV 场景下，信道状态与可服务 AP 集合会随飞行快速变化，此类策略需频繁重构决策模型，不仅增加计算负担，而且易导致策略短视与系统不稳定。

基于上述原因，本文研究了无蜂窝大规模 MIMO 系统中 UAV 用户与 AP 的关联以及上行功率控制的问题。将 UAV 的高度与视距 (line-of-sight, LoS) 和非视距 (non-line-of-sight, NLoS) 传播相关联，通过最小均方误差 (minimum mean square error, MMSE) 进行信道估计，推导出了所有用户可达的上行总速率闭式表达式。在此基础上，本文主要贡献如下：(1) 将 UAV-AP 关联和上行功率控制这一联合优化问题建模为 (markov decision process, MDP)，并采用 DDPG 算法求解，将 UAV 的整个飞行周期的大尺度衰落序列作为全局先验输入，该先验蕴含未来时隙的信道演化信息，从而实现对整个飞行周期的聚类与功控序列规划式学习。(2) 仿真结果表明，所提出的算法在此联合问题上取得的奖励值均高于其他 RL 算法，系统性能远高于全 AP 连接满功率发射，且在多重的空间分布轨

迹拓扑下均具有良好的适用性与稳定性。

1 系统模型

考虑一个由 K 个 UAV 用户和 M 个 AP 组成的无蜂窝大规模 MIMO 系统。如图 1 所示, AP 均匀分布在面积为 S 的区域内, UAV 用户在空中既定轨道匀速飞行, 飞行高度为 H , 均假设配备单根天线。假设第 k 个 UAV 的水平坐标为 (x_k, y_k) , 其中 $k = 1, 2, \dots, K$, 第 m 个 AP 的水平坐标为 (X_m, Y_m) , 其中 $m = 1, 2, \dots, M$ 。假设一个 UAV 可以由多个 AP 服务, 一个 AP 也可以服务多个 UAV。因从, 设 $\mathcal{M}_k$ 为服务于第 k 个 UAV 的 AP 集合, $\mathcal{K}_m$ 为由第 m 个 AP 所服务的 UAV 集合。所有 AP 在相同的时频资源上以时分双工模式为 UAV 服务。在上行训练阶段, UAV 将导频序列发送给 AP, AP 估计其覆盖范围内 UAV 的信道。在上行传输中, AP 无需与其他 AP 交换信道状态信息, 只需使用本地获得的信道估计对接收信号进行共轭加权, 从而实现最大比合并 (maximum ratio combining, MRC) [19]。

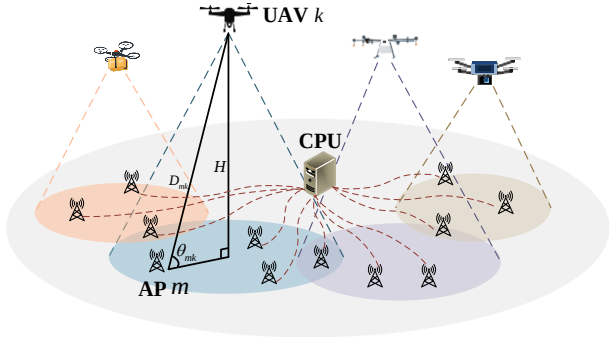


图1 以UAV为用户的无蜂窝大规模MIMO系统

1.1 信道模型

由于信道模型受到大尺度衰落和小尺度衰落的影响, 定义 $g_{mk} \in \mathbb{C}$ 表示第 m 个 AP 与第 k 个 UAV 之间的信道增益, 可表示为: $g_{mk} = \sqrt{\beta_{mk}} h_{mk}$, 其中 h_{mk} 表示第 m 个 AP 与第 k 个 UAV 之间的小尺度衰落系数。假设 h_{mk} 是独立同分布的随机变量, 服从 $\text{CN}(0, \mathbf{I}_N)$ 分布, 大尺度衰落 β_{mk} 与链路状态有关: 当第 m 个 AP 与第 k 个 UAV 之间为 LoS 链路时, $\beta_{mk} = D_{mk}^{-\alpha_1}$, 其发生概率为 p_{mk}^L ; 当该链路为 NLoS 链路时, $\beta_{mk} = D_{mk}^{-\alpha_2}$, 其发生概率为 p_{mk}^{NL} 。其中 α_1 和 α_2 分别代表 LoS 链路和 NLoS 链路的路径损耗指数, D_{mk} 表示第 k 个 UAV 与第 m 个 AP 之间的欧几

里得距离, 可表示为 $D_{mk} = \sqrt{(X_m - x_k)^2 + (Y_m - y_k)^2 + H^2}$, 在设计该无线通信系统时, 应该考虑到 LoS 链路和 NLoS 链路的随机性。因此, LoS 链路传输的概率如下[20]:

$$p_{mk}^L = \frac{1}{1 + \zeta_1 \exp\{-\zeta_2(\theta_{mk} - \zeta_1)\}} \quad (1)$$

其中, ζ_1 和 ζ_2 为取决于环境类型的常数参数, θ_{mk} 第 m 个 AP 与第 k 个 UAV 之间的仰角, 在传输过程中, 每条链路都有其特定的发生概率, 这主要取决于 UAV 与 AP 之间的相对位置与仰角, 以及环境因素。从图一可以计算出仰角 θ_{mk} 的公式为:

$$\theta_{mk} = \frac{180}{\pi} \tan^{-1} \frac{H}{\sqrt{(X_m - x_k)^2 + (Y_m - y_k)^2}} \quad (2)$$

其中 $m \in \mathcal{M}_k, k \in \mathcal{K}_m$, 根据(2)可以得到 NLoS 概率为 $p_{mk}^{NL} = 1 - p_{mk}^L$, 并且可以发现 LoS 概率会随着 UAV 高度和仰角的增加而变大。

从式(2)可以得到, 第 m 个 AP 与第 k 个 UAV 之间的平均大尺度衰落可以表示为:

$$\bar{\beta}_{mk} = p_{mk}^L D_{mk}^{-\alpha_1} + (1 - p_{mk}^L) D_{mk}^{-\alpha_2} \quad (3)$$

小尺度衰落在每个相干时间间隔内独立变化, 而大尺度衰落变化较慢, 可能在多个相干时间间隔内保持不变。

1.2 信道估计

设每个相干区间的长度为 τ_c , 其中上行训练阶段的长度为 τ_p , 且 $\tau_p < \tau_c$, 则上行载荷数据传输持续时间为 $\tau_u = \tau_c - \tau_p$ 。在上行训练阶段, 每个 UAV 发送一个 $\tau_p \times 1$ 的导频序列向量 $\sqrt{\tau_p} \boldsymbol{\varphi}_k$, 其中 $\|\boldsymbol{\varphi}_k\|^2 = 1$ 。因此, 上行训练阶段中第 m 个 AP 接收的导频信号为:

$$\mathbf{y}_{p,m} = \sqrt{\tau_p \rho_p} \sum_{k=1}^K g_{mk} \boldsymbol{\varphi}_k + \mathbf{w}_{p,m} \quad (4)$$

其中, ρ_p 为各导频的平均发射功率, $\mathbf{w}_{p,m}$ 为第 m 个 AP 接收的 $\tau_p \times 1$ 加性高斯白噪声向量, 其各元素为独立同分布的复高斯随机变量, 服从 $\text{CN}(0, 1)$ 。

基于接收到的导频信号, 以及对导频序列的已知信息, 第 m 个 AP 对其服务 $\mathcal{K}_m$ 集合中的 UAV 信道进行估计。将式(5)与导频向量 $\boldsymbol{\varphi}_k$ 同乘以后, 可得:

$$\begin{aligned}\hat{\mathbf{y}}_{p,mk} &= \boldsymbol{\varphi}_k^H \mathbf{y}_{p,m} \\ &= \sqrt{\tau_p \rho_p} \mathbf{g}_{mk} + \sqrt{\tau_p \rho_p} \sum_{k' \neq k}^K \mathbf{g}_{mk'} \boldsymbol{\varphi}_k^H \boldsymbol{\varphi}_{k'} + \boldsymbol{\varphi}_k^H \mathbf{w}_{p,m}\end{aligned}\quad (5)$$

根据(6), $\mathbf{g}_{mk}$ 的信道估计由 MMSE 估计得到, 表示如下:

$$\hat{\mathbf{g}}_{mk} = \frac{E\{\hat{\mathbf{y}}_{p,mk}^* \mathbf{g}_{mk}\}}{E\{|\hat{\mathbf{y}}_{p,mk}|^2\}} \hat{\mathbf{y}}_{p,mk} = C_{mk} \hat{\mathbf{y}}_{p,mk} \quad (6)$$

进一步地, 可以推导出 $\hat{\mathbf{g}}_{mk}$ 的表达式为:

$$\hat{\mathbf{g}}_{mk} = \frac{\sqrt{\tau_p \rho_p} \beta_{mk}}{\sqrt{\tau_p \rho_p} \sum_{k'=1}^K \beta_{mk'} |\boldsymbol{\varphi}_k^H \boldsymbol{\varphi}_{k'}|^2 + 1} \hat{\mathbf{y}}_{p,mk} \quad (7)$$

因此可以得到估计信道 $\hat{\mathbf{g}}_{mk}$ 的均方值 γ_{mk} , 具体为:

$$\gamma_{mk} \triangleq E\{|\hat{\mathbf{g}}_{mk}|^2\} = \frac{\tau_p \rho_p \beta_{mk}^2}{\tau_p \rho_p \sum_{k'=1}^K \beta_{mk'} |\boldsymbol{\varphi}_k^H \boldsymbol{\varphi}_{k'}|^2 + 1} \quad (8)$$

根据以上推导, 进一步得到估计误差 $\tilde{\mathbf{e}}_{mk} = \mathbf{g}_{mk} - \hat{\mathbf{g}}_{mk}$, 其是服从均值为 0、方差为 $\beta_{mk} - \gamma_{mk}$ 的复高斯随机变量。

特别地, 在本文所考虑的无蜂窝系统中, 每个 AP 仅对其关联服务集合 $\mathcal{M}_k$ 内的 UAV 进行信道估计, 因此关联集合 $\mathcal{M}_k$ 会影响信道估计结果并且进一步影响后续的上行速率表达式。由此可见本节信道估计过程是后续速率推导和联合优化建模的必要基础。

1.3 上行链路数据传输

在上行链路传输的过程中, UAV 将估计信道作为真实信道, 每个 AP 使用本地的信道估计对接收信号进行复共轭加权, 从而实施 MRC。由于 MRC 合并仅依赖于 AP 自身的局部信道状态信息, 因此无需在 AP 与 CPU 之间交换瞬时信道状态信息 (channel state information, CSI), 极大降低了系统的前传负载, 显著提升上行链路的整体接收性能。上行链路中, 第 m 个 AP 接收到的信号为:

$$\mathbf{y}_{u,m} = \sqrt{\rho_u} \sum_{k=1}^K \sqrt{\eta_k} \mathbf{g}_{mk} q_k + \mathbf{w}_{u,m} \quad (9)$$

其中 ρ_u 表示 UAV 的发送功率, η_k 是第 k 个 UAV 的功率控制系数, 满足 $0 \leq \eta_k \leq 1$, q_k 是第 k 个 UAV 发出的符号, 满足 $E\{|q_k|^2\} = 1$, 假设

$\mathbf{w}_{u,m} \sim \text{CN}(0,1)$, 表示第 m 个 AP 处接收到干扰的噪声。

为了检测来自第 k 个 UAV 的符号 q_k , 第 m 个 AP 使用其本地获得的信道估计 $\hat{\mathbf{g}}_{mk}$ 的复共轭对接收信号 $\mathbf{y}_{u,m}$ 进行加权。随后, 所得的量 $\hat{\mathbf{g}}_{mk}^* \mathbf{y}_{u,m}$ 通过前传链路发送至 CPU。CPU 接收到的信号为:

$$\begin{aligned}R_{u,k} &= \sum_{m \in \mathcal{M}_k} \hat{\mathbf{g}}_{mk}^* \mathbf{y}_{u,m} \\ &= \sum_{k'=1}^K \sum_{m \in \mathcal{M}_k} \hat{\mathbf{g}}_{mk}^* \sqrt{\rho_u \eta_k} \mathbf{g}_{mk'} q_{k'} + \sum_{m \in \mathcal{M}_k} \hat{\mathbf{g}}_{mk}^* \mathbf{w}_{u,m} \\ &= \sqrt{\rho_u \eta_k} \sum_{m \in \mathcal{M}_k} \hat{\mathbf{g}}_{mk}^* \mathbf{g}_{mk} q_k + \sum_{k' \neq k}^K \sum_{m \in \mathcal{M}_k} \hat{\mathbf{g}}_{mk}^* \sqrt{\rho_u \eta_k} \mathbf{g}_{mk'} q_{k'} \\ &\quad + \sum_{m \in \mathcal{M}_k} \hat{\mathbf{g}}_{mk}^* \mathbf{w}_{u,m}\end{aligned}\quad (10)$$

从而可以从 $R_{u,k}$ 检测出信号 q_k 。

1.4 可实现上行速率

每个 UAV 不知道瞬时 CSI, 但知道信道统计信息。因此, 将式(11)中的接收信号 $R_{u,k}$ 改写为:

$$R_{u,k} = \text{DS}_k q_k + \text{BU}_k q_k + \sum_{k' \neq k}^K \text{UI}_{kk'} q_{k'} + \sum_{m \in \mathcal{M}_k} \hat{\mathbf{g}}_{mk}^* \mathbf{w}_{u,m} \quad (11)$$

其中 DS_k 表示期望信号强度, BU_k 表示合并增益不确定性, $\text{UI}_{kk'}$ 为用户干扰。它们对应的详细公式如下:

$$\text{DS}_k = \sqrt{\rho_u} E \left\{ \sum_{m \in \mathcal{M}_k} \eta_k^{1/2} \hat{\mathbf{g}}_{mk}^* \mathbf{g}_{mk} \right\} \quad (12)$$

$$\begin{aligned}\text{BU}_k &= \sqrt{\rho_u} \left(\sum_{m \in \mathcal{M}(k)} \eta_k^{1/2} \hat{\mathbf{g}}_{mk}^* \mathbf{g}_{mk} - \right. \\ &\quad \left. E \left\{ \sum_{m \in \mathcal{M}(k)} \eta_k^{1/2} \hat{\mathbf{g}}_{mk}^* \mathbf{g}_{mk} \right\} \right)\end{aligned}\quad (13)$$

$$\text{UI}_{kk'} = \sqrt{\rho_u} \sum_{m \in \mathcal{M}(k)} \eta_k^{1/2} \hat{\mathbf{g}}_{mk}^* \mathbf{g}_{mk'} \quad (14)$$

由于 (12) 中的后三项与第一项不相关, 将后三项视为加性高斯噪声, 可以得到第 k 个 UAV 的可实现上行速率, 其表达式如下:

$$R_{u,k} = \log_2 \left(1 + \frac{|DS_k|^2}{E\left\{\left|BU_k\right|^2\right\} + \sum_{k' \neq k}^K E\left\{\left|UI_{kk'}\right|^2\right\} + E\left\{\left|\sum_{m \in \mathcal{M}_k} \hat{g}_{mk}^* \mathbf{w}_{u,m}\right|^2\right\}} \right) \quad (15)$$

经计算可以求得:

$$DS_k = \sqrt{p_u \eta_k} \sum_{m \in \mathcal{M}(k)} \gamma_{mk} \quad (16)$$

$$E\left\{\left|BU_k\right|^2\right\} = p_u \eta_k \sum_{m \in \mathcal{M}(k)} \gamma_{mk} \beta_{mk} \quad (17)$$

$$E\left\{\left|UI_{kk'}\right|^2\right\} = p_u \eta_{k'} \sum_{m \in \mathcal{M}(k)} \gamma_{mk} \beta_{mk'} + p_u \eta_{k'} \left(\sum_{m \in \mathcal{M}(k)} \gamma_{mk} \frac{\beta_{mk'}}{\beta_{mk}} \right)^2 |\psi_k^H \psi_{k'}|^2 \quad (18)$$

最终可以求得第 k 个 UAV 的上行总速率闭式表达式表示为:

$$R_{u,k} = \log_2 \left(1 + \underbrace{p_u \eta_k \left(\sum_{m \in \mathcal{M}_k} \gamma_{mk} \right)^2}_{|DS_k|^2} \times \left(\underbrace{p_u \sum_{k' \neq k}^K \eta_{k'} \left(\sum_{m \in \mathcal{M}_k} \gamma_{mk} \frac{\beta_{mk'}}{\beta_{mk}} \right)^2}_{CI_k} |\varphi_k^H \varphi_{k'}|^2 + \underbrace{p_u \sum_{k'=1}^K \eta_{k'} \sum_{m \in \mathcal{M}_k} \gamma_{mk} \beta_{mk}}_{UI_k} + \underbrace{\sum_{m \in \mathcal{M}_k} \gamma_{mk}}_{N_k} \right)^{-1} \right) \quad (19)$$

其中 $|DS_k|^2$ 表示第 k 个 UAV 经其服务 AP 集合 $\mathcal{M}_k$ 合并后得到的期望信号功率, 其大小由上行发射功率、功率控制系数以及服务 AP 上的信道估计均方值共同决定。 CI_k 表示与第 k 个 UAV 复用相同导频的其他用户所造成的相干干扰; UI_k 表示所有 UAV 上行传输引入的非相干干扰及信道估计误差影响;

N_k 表示接收合并后的噪声功率。由此可见, UAV-AP 关联集合 $\mathcal{M}_k$ 决定参与信号接收合并的 AP 范围, 功率控制系数 η_k 决定各 UAV 的发射功率, 二者均直接影响可达速率表达式, 是后续联合优化问题中的核心变量。

2 联合优化问题描述及求解

本节中, 研究目标是给出联合优化问题的数学表达式, 并将其建模为 MDP, 基于 DDPG 设计一种算法来找到这一问题的最优解。

2.1 优化问题描述

基于所构建的 UAV 无蜂窝大规模 MIMO 系统, 研究目标是优化 UAV-AP 的关联关系以及上行功率控制, 以最大化所有用户整个飞行周期内的平均上行速率和。前文提到的集合 $\mathcal{M}_k$, 表示第 k 个 UAV 与 AP 之间的关联关系, 即若第 m 个 AP 属于 $\mathcal{M}_k$, 说明第 m 个 AP 与该 UAV 建立服务关联并参与其上行信号接收; 反之, 则表示二者不建立服务关联, 因此对 $\mathcal{M}_k$ 的优化即体现了 UAV-AP 关联关系的优化。由此, 该优化问题可表示为如下形式:

$$\begin{aligned} \max_{\{\mathcal{M}_k, \eta_k\}} \quad & \sum_{k=1}^K R_{u,k} \\ \text{s.t.} \quad & 1 \leq |\mathcal{M}_k| \leq M, \quad \forall k \in \mathcal{K} \\ & 0 < \eta_k \leq 1, \quad k = 1, 2, \dots, K \end{aligned} \quad (20)$$

其中 $|\mathcal{M}_k|$ 表示为第 k 个 UAV 服务的接入点集合中的 AP 数量, η_k 为第 k 个 UAV 的上行功率控制系数。

聚类策略决定了 MRC 合并所利用的空间维度与分集增益, 功率控制则直接影响 UAV 的有效信干噪比及其对其他 UAV 的同频干扰。因此, 聚类策略与功率控制并不是两个独立的子问题, 而是紧密耦合、互相影响的联合优化任务。然而, 该联合优化问题(21)具有高度非线性、非凸性以及高维决策空间等特点, 传统的凸优化方法难以获得可行解。为此, 本研究引入 DRL 思想, 将 UAV-AP 关联和上行功率控制这一联合优化问题建模为 MDP。

2.2 马尔可夫决策过程构建

在 DRL 中, MDP 为随机环境的序列决策提供统一架构, 由状态空间、动作空间、奖励函数及状态转移概率构成, 用以刻画环境的不确定性与动态性。智能体在时刻 t 观测状态 s_t 后按策略 $\pi(a|s)$ 选择动作 a_t , 环境转移至 s_{t+1} 并返回即时奖励 r_t ; 学习目标是在交互中更新 π , 使折扣期望回报最大,

从而实现长期收益最优化。在这一框架下进一步提出 DDPG 算法, DDPG 算法在处理具有连续状态空间和动作空间的复杂决策问题上具有显著优势^[21]。

对于所提出的 UAV-AP 关联以及上行功率控制这一联合优化问题, 本文在 DDPG 基础上研究设计一种全时域规划引导的双智能体两阶段交替 DDPG (full-horizon alternating DDPG, FA-DDPG), 在阶段一进行聚类训练时, 功率智能体保持固定的基线动作, 在阶段二训练上行功率控制时, 使用阶段一训练好的聚类策略。在所建构的 MDP 框架下, 状态空间、动作空间和奖励函数的具体定义如下所示。

(1) 状态空间 $\mathcal{S}$ 。在时隙 t , 状态 $s_t \in \mathcal{S}$ 为一组可用于描述参数, 基于本文提出的系统模型, 状态空间表示为 $s_t = \{\text{ctx}_t, a_{t-1}\}$ 。其中首先将表示 UAV 与 AP 之间的信道状态的大尺度衰落矩阵 $\beta_{mk}(t) \in \mathbb{R}^{K \times M}$ 表向量化为 $\mathbf{b}_t \in \mathbb{R}^{KM}$, 并构造全时域序列 $\mathbf{B} = \{\mathbf{b}_0, \mathbf{b}_1, \dots, \mathbf{b}_{T-1}\}$, 再采用逐时隙共享编码器对每个 $\mathbf{b}_t$ 进行一致的非线性压缩映射, 得到低维表示 $\mathbf{e}_t = \psi(\mathbf{b}_t) \in \mathbb{R}^d$, 随后为显式刻画时隙顺序, 在每个时隙引入位置编码 $\mathbf{p}_t \in \mathbb{R}^d$, 形成 $\mathbf{x}_t = \mathbf{e}_t + \mathbf{p}_t$, 在此基础上通过 Transformer 时序汇聚对 $\{\mathbf{x}_t\}$ 进行全局信息交互得到上下文序列 $\{\text{ctx}_0, \text{ctx}_1, \dots, \text{ctx}_{T-1}\}$, 其中 $\text{ctx}_t \in \mathbb{R}^d$ 表征在第 t 个时隙、融合全时域信息后的全局上下文, 最终, 将每个时隙的 RL 状态定义为 $s_t = \{\text{ctx}_t, a_{t-1}\}$ 。

(2) 动作空间 $\mathcal{A}$ 。为了实现上行速率的最大化, 在每个决策时隙 t , 聚类智能体依据当前的状态信息 s_t 输出聚类动作 a_t^C , 并将其映射为一个二进制连接矩阵 $\mathbf{C}_t \in \mathbb{R}^{K \times M}$, 其中的元素 $c_{k,m} = 0$ 为表示不建立通信链路, $c_{k,m} = 1$ 则表示第 k 个 UAV 选择了第 m 个 AP。功控智能体输出的功控动作为 $a_t^P = [a_1^p, a_2^p, \dots, a_K^p]$, 其中 $a_k^p \in [0, 1]$ 表示第 k 个 UAV 的发射功率控制系数, 用于调节其上行信号功率与干扰平衡。

(3) 奖励 r_t 。在每个时隙执行完动作后, 环境根据当前聚类关系与功率系数计算系统的上行速率, 并将其作为即时奖励。奖励函数定义为:

$$r_t = \sum_{k=1}^K R_{u,k} \quad (21)$$

2.3 基于 actor-critic 框架的 FA-DDPG 算法

由于直接使用 DRL 算法来解决连续空间、有限时间范围的 MDP, 通常是一项具有挑战性的任务, 这主要是由于状态和动作空间的连续性增加了问题的复杂性。为了应对这一挑战, 本文提出了一种适用于高维连续动作空间的两阶段 FA-DDPG 算法来优化 UAV-AP 关联及上行功率控制。

如图 2 所示, FA-DDPG 包含聚类和功控两个智能体两个部分, 每个智能体在实现上采用 DDPG 的 actor-critic 结构, 并配备训练 actor、训练 critic 及其对应的目标网络。其中对于聚类智能体而言, 参数为 θ_a^C 的训练 actor 网络, 记为 $\mu^C(s_t; \theta_a^C)$, 用于给出用户的近似策略并产生动作。参数为 θ_c^C 的训练 critic 网络, 记为 $Q^C(s_t, a_t^C; \theta_c^C)$, 用于估计在状态 s_t 下执行动作 a_t^C 的动作值函数。参数为 θ_a^C 的目标 actor 网络和参数为 θ_c^C 的目标 critic 网络分别用 $\mu^C(s_t; \theta_a^C)$ 和 $Q^C(s_t, a_t^C; \theta_c^C)$ 表示, 为训练 actor 网络和训练 critic 网络生成目标 Q 值, 功控智能体同理。

FA-DDPG 算法使用确定性策略梯度(deterministic policy gradient, DPG)定理来更新 θ_a^C 、 θ_c^C 、 θ_a^P 、 θ_c^P 和 θ_a^P 、 θ_c^P 、 θ_a^P 、 θ_c^P , 并在 actor-critic 框架中生成动作。将当前状态 s_t 输入到训练 actor 网络中。根据 DPG 定理, 此网络通过将状态确定性映射到特定动作来形成策略, 并模拟智能体的策略函数, 从中选定动作。本文在选定的动作 a_t 上叠加了噪声, 得到最终的输出动作 $a_t = \mu(s_t; \theta_a) + N_t$ 。其中 N_t 是服从正态分布的随机噪声。特别的是, 在进行聚类训练阶段, 功控智能体只会做出一个固定的基线动作 $a_{baseline}^P$, 而在进行功控训练阶段, 聚类智能体只会输出训练好的无噪声动作 $a_t^C = \mu^C(s_t; \theta_a^C)$ 。在执行动作后, 智能体会收到奖励 r_t , 并进入新的状态 s_{t+1} 。随后智能体会将这次的经验 (s_t, a_t, r_t, s_{t+1}) 保存在经验回放池 $\mathcal{N}$ 中。同时, 训练 critic 网络会对本次动作进行评估。

为了稳定训练过程, 本文采用目标网络构造贝尔曼目标。对于聚类智能体, 从经验回放池 $\mathcal{N}^C$ 中均匀采样一个批次样本 $(s_i, a_i^C, r_i, s_{i+1})$, 其目标 Q 值写为:

$$y_i^C = r_i + \lambda Q^C(s_i^C, \mu^C(s_i^C; \theta_a^C); \theta_c^C) \quad (22)$$

其中 $\lambda \in (0, 1)$ 为折扣因子, 目标 actor 和目标 critic

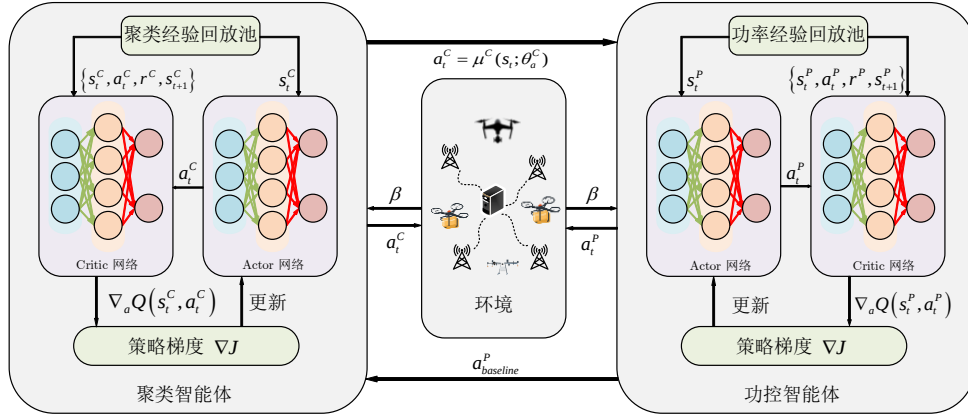


图2 FA-DDPG算法训练架构

网络的参数 θ_a^C 和 θ_c^C ，分别是原始actor网络参数 θ_a^C 和原始critic网络参数 θ_c^C 经过软更新后的版本。与之对应的critic损失函数为：

$$L^C(\theta_c^C) = \frac{1}{N_{batch}} \sum_{i=1}^{N_{batch}} (Q^C(s_i, a_i^C; \theta_c^C) - y_i^C)^2 \quad (23)$$

通过最小化上式来更新聚类critic网络参数 θ_c^C ,

$$\begin{aligned} & \nabla_{\theta_a^C} J^C(\theta_a^C) \\ & \approx \frac{1}{N_{batch}} \sum_{i=1}^{N_{batch}} [\nabla_a Q^C(s, a; \theta_c^C)|_{s=s_i^C, a=\mu^C(s_i^C)} \nabla_{\theta_a^C} \mu^C(s_i^C; \theta_a^C)] \end{aligned} \quad (25)$$

并沿着 $\nabla_{\theta_a^C} J^C$ 方向进行梯度上升。随着训练的进行，各个网络参数进行刷新。为进一步提升训练稳定性，对目标网络采用软更新策略：

$$\theta_a^{C'} \leftarrow \tau \theta_a^C + (1 - \tau) \theta_a^{C'}, \theta_c^{C'} \leftarrow \tau \theta_c^C + (1 - \tau) \theta_c^{C'} \quad (26)$$

其中 $\tau \ll 1$ 为软更新系数。在功控阶段进行相同的训练。通过上述的双智能体交替FA-DDPG训练框架，聚类智能体和功控智能体可以分别在高维连续动作空间中学习到近似最优的UAV-AP关联策略与上行功率控制策略。整个FA-DDPG网络的整体训练流程如算法1所示。

算法1 基于DDPG的全时域规划算法

输入 本地Actor和Critic神经网络参数、目标Actor和Critic网络参数、最大迭代回合

输出 训练后Actor和Critic神经网络参数、聚类策略 $\mu^C(s_i; \theta_a^C)$ 、功控策略 $\mu^P(s_i^P; \theta_a^P)$

- 1) 初始化网络参数、经验回放池
- 2) 第一阶段：聚类策略训练
- 3) while 聚类阶段迭代回合 $\leq$ 最大聚类迭代回合 do
 - 4) 初始化环境，令 $t \leftarrow 0$

并采用梯度下降进行更新 θ_c^C 。在critic网络固定时，actor网络通过最大化对应 Q 函数来更新，其目标函数可写为：

$$J^C(\theta_a^C) = \mathbb{E}_{s^C \sim \mathcal{N}^C} [Q^C(s^C, \mu^C(s^C; \theta_a^C); \theta_c^C)] \quad (24)$$

参数 θ_a^C 的梯度近似为：

- 5) while $t \leq T$ do
 - 6) 根据状态 s_t 和随机噪声 N_t^C ，通过聚类actor网络计算聚类动作 $a_t^C = \mu^C(s_t; \theta_a^C) + N_t^C$
 - 7) 执行聚类动作 a_t^C ，更新聚类矩阵 C_t ，得到中间状态 s_t^C
 - 8) 根据状态 s_t^C ，执行功率基线动作 $a_{baseline}^P$ ，获得和奖励 r_t
 - 9) 将经验 (s_t, a_t^C, r_t, s_t^C) 存储至聚类经验回放池 $\mathcal{N}^C$ 中
 - 10) $t \leftarrow t + 1$
 - 11) end while
 - 12) 调用聚类critic损失函数 $L^C(\theta_c^C)$ ，根据其梯度更新聚类网络参数 $(\theta_a^C, \theta_c^C, \theta_a^C, \theta_c^C)$
 - 13) end while
 - 14) 第二阶段：功率控制策略训练
 - 15) while 聚类阶段迭代回合 $\leq$ 最大聚类迭代回合 do
 - 16) 初始化环境，令 $t \leftarrow 0$
 - 17) while $t \leq T$ do
 - 18) 根据状态 s_t ，通过一阶段训练得到的确定

性聚类动作 $a_t^C = \mu^C(s_t; \theta_a^C)$

19) 执行聚类动作 a_t^C , 得到中间状态 s_t^C

20) 根据中间状态 s_t^C 和随机噪声 N_t^P , 通过功率 actor 网络计算功率动作 $a_t^P = \mu^P(s_t^C; \theta_a^P) + N_t^P$

21) 执行动作 a_t^P , 获得和奖励 r_t 和下一状态 s_{t+1}

22) 将经验 $(s_t^C, a_t^C, r_t, s_{t+1})$ 存储至聚类经验回放池 $\mathcal{N}^C$ 中

23) $t \leftarrow t + 1$

24) end while

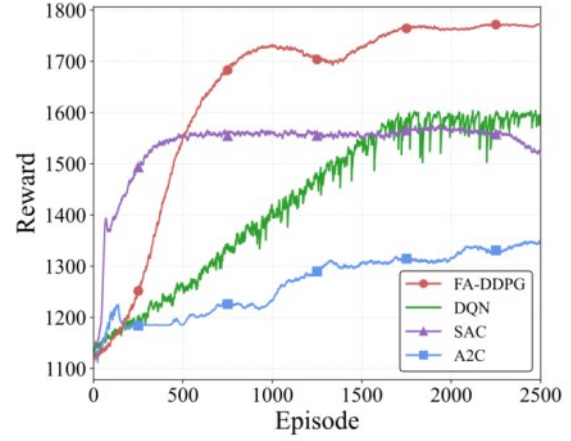
25) 调用功控 critic 损失函数 $L^P(\theta_c^P)$, 根据其梯度更新聚类网络参数 $(\theta_a^P, \theta_c^P, \theta_a^P, \theta_c^P)$

26) and while

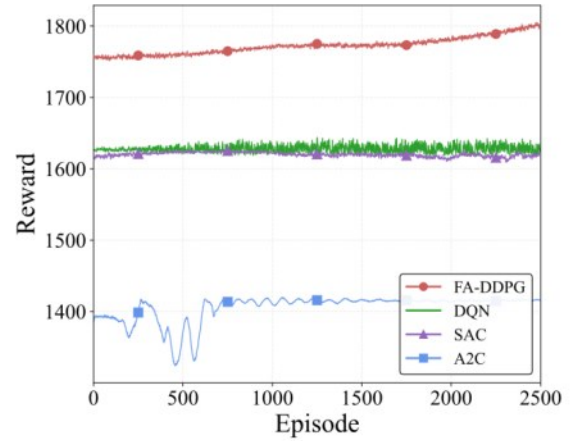
3 仿真结果与分析

在仿真时设定每个导频信号的平均发射功率是 $\rho_p = 100 \text{ mW}$, 每个 UAV 在上行链路的最大平均发射功率是 $\rho_u = 200 \text{ mW}$ 。在公式(2)中的环境参数 ξ_1 和 ξ_2 分别设为 11.95 和 0.136^[19]。LoS 和 NLoS 的衰减因子 α_1 和 α_2 分别设为 1.2 和 2.1。训练阶段块的长度为 $\tau_p = 10$ 个符号。RL 相关参数设定如下, 折扣因子 $\lambda = 0.99$, 软更新系数 $\tau = 5 \times 10^{-4}$, actor 网络学习率设为 5×10^{-5} , critic 网络学习率设为 3×10^{-4} , 经验回放缓冲区容量设为 10^5 , 采样的批量大小 N_{batch} 设为 256, 聚类和功控阶段的最大迭代回合设定均为 2500, 且只有当聚类阶段达到最大迭代回合后, 才会切换至功率控制阶段。

图 3 给出了在 UAV-AP 关联与上行功率控制联合优化场景下, 四种深度 RL 算法的训练回报曲线。从图 3(a)对应的聚类阶段可以看出在仅优化 UAV-AP 关联、功率采用固定功率的条件下, 软演员-评论家 (soft actor-critic, SAC) 在前期具有较快上升速度, 但在很快进入平台; 深度 Q 网络 (deep Q-network, DQN) 能够提升系统回报但波动较大; 优势演员-评论家 (advantage actor-critic, A2C) 整体提升有限, 长期性能明显落后于其他三种方法。本文提出的 FA-DDPG 收敛很快且最终获得最高的稳态回报并保持较好的平滑性。图 3(b)对应功率控制阶段, 此时聚类策略固定为第一阶段学得的聚类策略。从图 3(b)可以观察到, 各算法的初始回报水平整体高于聚类阶段, DQN 与 SAC 的最终性能较为接近, 但低于 FA-DDPG; A2C 依旧明显落后; 其



(a) 聚类阶段



(b) 功控阶段

图 3 不同 RL 算法的奖励值

中 FA-DDPG 仍然取得最高的稳态回报, 并在整个训练过程中保持单调、稳定的提升。两幅图表明, 在本联合优化任务中, 本文提出的 FA-DDPG 在聚类与功控两个子问题上均能获得更高的长期回报和更平滑的收敛轨迹。

图 4 比较了不同算法下的上行和速率随 UAV 数量变化趋势。为保证对比公平性, 本文将文献[18]中的 A-DDPG 算法奖励函数统一设置为最大化系统和速率, 并去除了其原有的公平性惩罚项, 使其优化目标与本文方法一致。此外, 图中还给出了基于文献[22]的“UC Top-15 Max-min”和“全 AP Max-min”两种方案。其中, UC Top-15 表示为每个 UAV 选择信道增益最强的前 15 个 AP 建立连接, 而全 AP 表示所有 AP 均参与服务; 在此基础上, 采用文献[22]中的 Max-min 方法, 并通过二分法优化功率系数。“UC Top-15 满功率”和“全 AP 连接满功

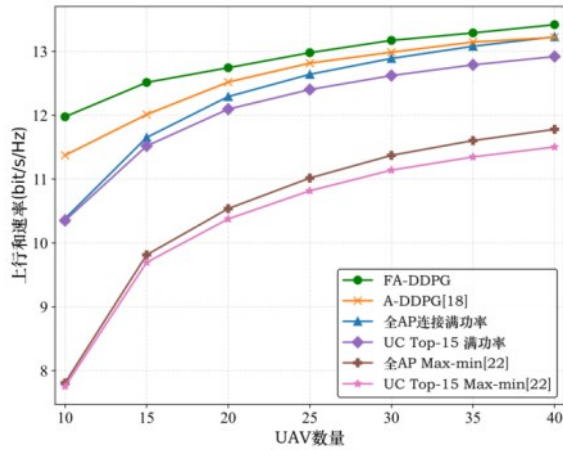


图4 不同方案下上行链路和速率随UAV数量的关系

率”则表示在选择不同数量的AP基础上以最大功率进行传输。从图中可以看出,随着UAV数量 K 增加,各方案的总速率均有所提升,但FA-DDPG在全范围内始终保持最高性能。这是因为FA-DDPG利用整个飞行周期的大尺度衰落序列进行决策,能够更合理地协调跨时隙的AP选择与功率分配。相比之下,A-DDPG和UC Top-15 Max-min、UC Top-15 满功率连接方法主要依赖当前时刻信息,难以充分刻画高机动场景下的动态变化,而全AP连接则受协作冗余和干扰增强限制。上述结果进一步说明了FA-DDPG在高机动UAV场景中的有效性与可扩展性。

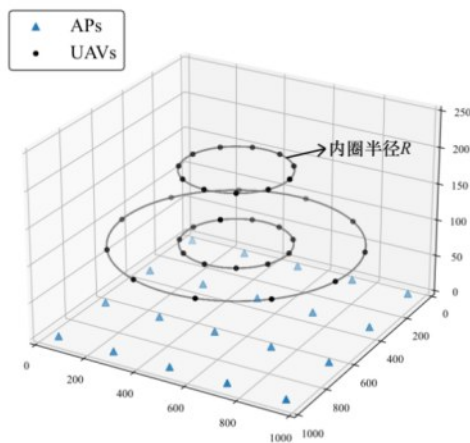


图5 UAV飞行轨迹图

图5展示了UAV的多重轨迹飞行图,将AP数量 M 设定为20,均匀分布在边长为1000米的方形区域内。其中UAV飞行线速度设为20 m/s。假设

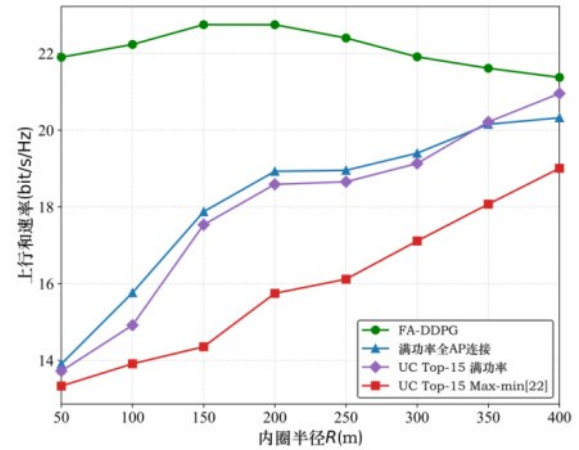


图6 不同方案下上行链路和速率随不同轨道的关系

UAV沿三条同心圆形轨道飞行。其中,外环轨道参数固定,半径为450m、高度为100m;另外两条内

环轨道,与外环同心,半径相同,其半径记为内环半径 R ,且在50-400m范围内逐步增大。两条内环轨道的飞行高度分别设置为100m和200m。该多环轨道使UAV-AP之间的LoS/NLoS关系及大尺度衰落在时间与空间上均呈现强烈非平稳性,从而显著增加关联与功率控制问题的复杂度。

图6比较了不同算法下的上行和速率随UAV飞行内圈半径变化性能表现。其中三个比较算法已在图4中做出说明。令 $K=33$,将内圈半径由50m逐步增加至400m,从而构造出由高度重叠、轨迹紧凑到空间分布较为稀疏的多环场景。可以看到,基线方案的和速率随 R 增大单调上升,主要得益于内圈扩展后,不同用户间几何间隔增大,因而其在上行传输时对AP侧造成的互用户干扰显著减弱。从图中可以看出,在所有半径下,FA-DDPG曲线始终显著高于其他优化方案。当 $R=50$ m时,三条轨道高度和半径差异较小,互用户干扰最为严重,但是FA-DDPG仍能通过自适应聚类 and 功率控制获得超过21bit/s/Hz的性能增益;当 R 增大至200到400m时,虽然总体干扰水平降低,基线方案性能接近饱和,但FA-DDPG依然保持约1.5bit/s/Hz的优势。上述结果表明,所提出的FA-DDPG策略能够根据不同轨道设定实时调整UAV-AP关联关系及上行发送功率,在多重、不均匀的空间分布轨迹拓扑下均具有良好的适用性与稳定性。

图7进一步比较了不同AP数量下三种方案的上行和速率表现。为保证对比公平性,FA-DDPG

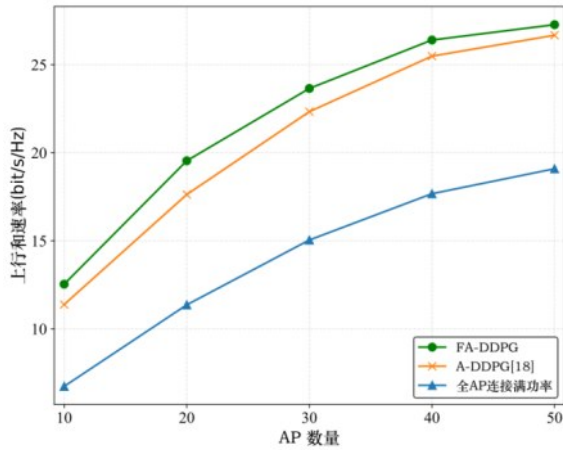


图7 不同方案下上行链路和速率随AP数量的关系

与A-DDPG的奖励函数的设定与图4中保持一致。“全AP连接满功率”算法已在图4中做出解释。可以看到，随着AP数量由10增加到50，三种方案的上行和速率均明显提升，这是因为更密集的AP部署为UAV提供了更丰富的空间分集与协作增益；但当AP数量进一步增加时，曲线斜率逐渐减小，说明在AP部署较为充分后，系统由新增AP所获得的有效协作增益呈边际递减趋势，因此总速率的提升幅度较前期明显放缓。在整个范围内，FA-DDPG始终取得最高速率，且相较A-DDPG和全AP连接满功率基线方案均保持稳定优势。其原因在于，FA-DDPG利用整个飞行周期的大尺度衰落序列构建全局上下文，能够在AP数目变化时更加前瞻地规划服务AP选择与功率分配，充分挖掘新增AP带来的潜在协作收益，同时避免冗余关联和干扰放大。相比之下，A-DDPG仅基于当前时隙信息决策，更易陷入局部最优；而全AP连接满功率虽然可利用全部AP资源，但由于同频干扰增强，性能始终受限。该结果说明，所提出的FA-DDPG不仅在不同UAV规模下有效，而且对不同AP部署密度同样具有良好的可扩展性与鲁棒性。

4 结束语

在低空经济快速发展背景下，本文研究无蜂窝大规模MIMO系统上行传输中UAV-AP关联与功率控制的联合优化问题。针对UAV多轨迹运行场景，采用LoS/NLoS复合大尺度衰落模型，推导上行传输速率的闭式表达式，并将最大化上行和速率的问题刻画为马尔可夫决策过程。在此基础上，为突破传统方法仅依赖当前时隙大尺度衰落信息所带来的

短视决策局限，提出全时隙联合学习的FA-DDPG方法，将一个飞行周期内的全时隙大尺度衰落序列经编码压缩与时序汇聚后作为强化学习状态输入，并通过两阶段交替优化实现整个飞行周期的聚类与功耗序列规划学习：第一阶段由聚类智能体在连续动作空间中自适应学习UAV-AP关联策略，第二阶段在冻结聚类策略的前提下由功率智能体优化上行功率控制。仿真结果表明，相比满功率发射全AP连接基线及DQN、SAC、A2C等DRL方法，所提FA-DDPG能够获得更高且更稳定的上行和速率增益，表明该DRL框架对高机动UAV场景具有良好的适用性和推广潜力。后续工作将进一步把能量效率、时延约束和用户公平性纳入奖励设计，构建多目标约束RL模型。

参考文献：

- [1] 于紫月.《2024 低空经济场景白皮书》出炉[J]. 中国科技财富, 2024, (10): 29.
YU Ziyue.《The 2024 Low-Altitude Economy Scenarios White Paper》Released[J]. China Science and Technology Fortune Magazine, 2024, (10): 29.
- [2] Y. Zeng, R. Zhang and T. J. Lim, Wireless communications with unmanned aerial vehicles: Opportunities and challenges[J]. IEEE Communications magazine, 2016, 54(5): 36-42
- [3] M. Mozaffari, W. Saad, M. Bennis, et al. A Tutorial on UAVs for Wireless Networks: Applications, Challenges, and Open Problems[J]. IEEE Communications Surveys & Tutorials, 2019, 21(3): 2334-2360.
- [4] 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.
- [5] A. Otto, N. Agatz, J. Campbell, et al. Optimization approaches for civil applications of unmanned aerial vehicles (UAVs) or aerial drones: A survey[J]. Networks, 2018, 72(4): 411-458.
- [6] D. C. Tsouros, S. Bibi, and P. G. Sarigiannidis, A review on UAV based applications for precision agriculture[J]. Information, 2019, 10(11): 349.
- [7] 白露, 孙铭然, 黄子蔚, 等. 面向应急通信的多无人机协同通信建模研究[J]. 通信学报, 2023, 44(7): 38-50.
Lu BAI, Mingran SUN, Ziwei HUANG, et al. Multi-UAV cooperative channel model for emergency communication[J]. Journal on Communications, 2023, 44(7): 38-50.
- [8] Ellenberg A, Branco L, Krick A., et al. Use of unmanned aerial vehicle for quantitative infrastructure evaluation[J]. Journal of Infrastructure Systems, 2015, 21(3): 04014054.
- [9] NGO H Q, ASHIKHMIN A, YANG Hong, et al. Cell-free massive MIMO versus small cells[J]. IEEE Transactions on Wireless Communications, 2017, 16(3): 1834-1850.
- [10] Interdonato G, Björnson E, Quoc Ngo H, et al. Ubiquitous cell-free

massive MIMO communications[J]. EURASIP Journal on Wireless Communications and Networking, 2019, 2019(1): 1-3.

- [11] S. Buzzi and C. D' Andrea, Cell-Free Massive MIMO: User-Centric Approach[J]. IEEE Wireless Communications Letters, 2017, 6(6): 706-709.
- [12] S. Buzzi, C. D' Andrea, M. Fresia, et al. Multi-UE Multi-AP Beam Alignment in User-Centric Cell-Free Massive MIMO Systems Operating at mmWave[J]. IEEE Transactions on Wireless Communications, 2022, 21(11): 8919-8934.
- [13] S. Chen, J. Zhang, E. Björnson, et al. Improving Fairness for Cell-Free Massive MIMO Through Interference-Aware Massive Access[J]. IEEE Transactions on Vehicular Technology, 2023, 72(4): 5468-5472.
- [14] H. Cao, H. Zeng, X. Zhu, et al. Adaptive Fuzzy Logic Based AP Selection for Scalable User-Centric Cell-Free Massive MIMO[J]. IEEE Transactions on Vehicular Technology, 2025, 74(2): 3429-3433.
- [15] M. S. A. Khan, S. Agnihotri and R. M. Karthik, Joint AP-UE Association and Power Factor Optimization for Distributed Massive MIMO [C]. 2024 IEEE 35th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Valencia, Spain, 2024: pp. 1-6.
- [16] X. Zhang, V. A. Le, M. Kaneko, et al. Multi-Agent Deep Reinforcement Learning-Based Uplink Power Control in Cell-Free Massive MIMO With Mobile Users[J]. IEEE Transactions on Vehicular Technology, 2025, 74(11):17796-17811.
- [17] B. Banerjee, R. C. Elliott, W. A. Krzymieñ, et al. Access Point Clustering in Cell-Free Massive MIMO Using Conventional and Federated Multi-Agent Reinforcement Learning[J]. IEEE Transactions on Machine Learning in Communications and Networking, 2023, 1:107-123.
- [18] D. Mingjun, S. Xinghua, Z. Yue, et al. Deep reinforcement learning based joint cooperation clustering and downlink power control for cell-free massive MIMO[J]. China Communications, 2024, 21(11):1-14.
- [19] S. Elhoushy, M. Ibrahim and W. Hamouda, Cell-Free Massive MIMO: A Survey[J]. IEEE Communications Surveys & Tutorials, 2022, 24(1): 492-523.
- [20] A. Al-Hilo, M. Samir, M. Elhattab, et al. RIS-Assisted UAV for Timely Data Collection in IoT Networks[J]. IEEE Systems Journal, 2023, 17(1): 431-442.
- [21] L. Luo, J. Zhang, S. Chen, et al. Downlink Power Control for Cell-Free Massive MIMO With Deep Reinforcement Learning[J]. IEEE Transactions on Vehicular Technology, 2022, 71(6): 6772-6777.
- [22] H. Q. Ngo, A. Ashikhmin, H. Yang, et al. Cell-Free Massive MIMO: Uniformly great service for everyone[C].in Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC), Stockholm, Sweden, 2015:pp. 201 - 205.



张琦 (1987-), 女, 河南郑州人, 博士, 南京邮电大学教授、硕士生导师, 主要研究方向为 6G 无线通信基础理论与关键技术研究, 包括大规模 MIMO、物联网、人工智能、无人机通信等。



贾旭日 (2001-), 男, 内蒙古包头人, 南京邮电大学硕士生, 主要研究方向为无蜂窝大规模 MIMO、强化学习、无人机通信等。



张军 (1980-), 男, 安徽安庆人, 博士, 南京邮电大学教授、博士研究生导师, 主要研究方向为未来移动通信理论与关键技术, 包括大规模 MIMO、智能超表面辅助通信、通感一体化、无人机通信、人工智能通信、毫米波通信、以及物联网中的通信理论与技术等。



蔡曙 (1984-), 男, 江苏泰州人, 博士, 南京邮电大学、研究生导师, 主要研究方向包含阵列和通信信号处理、统计信号处理、检测与估计理论。目前, 主要专注于 5G 无线通信中的新技术, 包括无人机通信、大规模天线阵列、人工智能等。



吴怡 (1980-), 女, 辽宁葫芦岛人, 博士, 福建师范大学教授、博士研究生导师, 主要研究方向为光电传感、定位、自组织网络通信等。