面向低空通信的无蜂窝大规模MIMO智能关联及功率控制方法

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通信的支撑作用日益重要。针对传统基于瞬时位置的聚类方法存在更新滞后、固定功率控制策略难以适应位置动态变化的问题，构建无人机－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。在此基础上，提出基于深度确定性策略梯度的两阶段交替优化算法，以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，从而实现对整个飞行周期聚类与功率控制序列的联合规划学习。仿真结果表明，与传统方案相比，所提方法能够获得更高且更稳定的系统性能，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：无蜂窝大规模多输入多输出；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：10.11959/j.issn.1000–436x.TXXB260251

Intelligent association and power control method for cell-free massive MIMO in low-altitude communications

Zhang Qi[1], Jia Xuri[1], Cai Shu[1], Zhang Jun[1], Wu Yi[2]
1. School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
2. College of Photonic and Electronic 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 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-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 proposed, in which the large-scale fading sequence over the entire unmanned aerial vehicle 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 show that the proposed method achieves higher and more stable system performance than conventional schemes and exhibits good adaptability under various unmanned aerial vehicle spatial topology distributions.

Key words: cell-free massive multiple-input multiple-output, 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, No.U25A20398）；江苏省重点研发计划基金资助项目（No.BE2023022-3）
Foundation Items：The National Natural Science Foundation of China (No.62571267, No.U25A20398), The Key Research and Development Program of Jiangsu Province (No.BE2023022-3)

0　引言

随着2024年低空经济被写入《政府工作报告》，我国无人机（unmanned aerial vehicle, UAV）产业迎来迅猛发展。在空天地一体化通信的发展趋势下，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]，在物流配送、农业生产、应急救援、城市基础设施建设和维护等领域发挥着日益重要的作用[4-8]。无蜂窝大规模多输入多输出（multiple-input multiple-output, MIMO）是第六代移动通信系统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打破固定小区边界的限制，不会由于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，在实际部署中仍面临许多挑战。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，使功率控制变得更关键，功率过低会削弱有效信号强度，影响通信质量；功率过高则会增加能耗，导致飞行时间缩短并加大用户间的同频干扰。

为解决上述问题，近年来研究者提出多种用户–接入点关联方法。文献[11]提出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“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6 GHz以下频段建立控制信道，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]提出基于自适应模糊逻辑的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，在约束前传负载条件下提升系统总吞吐量。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。文献[16]提出一种深度强化学习（deep reinforcement learning, DRL）框架，通过共享策略网络实现上行功率控制，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。实验结果表明，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较好地迁移到新环境中，展现出良好的泛化能力。文献[18]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RL方案大多面向地面静止用户，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，还易导致策略短视与系统不稳定。

基于上述原因，本文研究无蜂窝大规模MIMO系统中UAV用户与AP的关联以及上行功率控制问题。将UAV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，并推导出所有用户可达的上行总速率闭式表达式。本文的主要贡献如下。

（1）将UAV-AP关联和上行功率控制的联合优化问题建模为马尔可夫决策过程（Markov decision process, MDP），并采用DDPG算法求解。

（2）将UAV整个飞行周期内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聚类与功率控制序列的联合规划学习。仿真结果表明，本文方法在此联合优化问题中取得的奖励值均高于其他RL算法，系统性能显著优于全AP连接满功率发射方案，且在多重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，且均配备单根天线。假设第k个UAV的水平坐标为(x_k,y_k)，其中$k=1,2,\cdots,K$，第m个AP的水平坐标为(X_m,Y_m)，其中$m=1,2,\cdots,M$。假设一个UAV可以由多个AP服务，一个AP也可以服务多个UAV。因此，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无须与其他AP交换信道状态信息（channel state information, CSI），仅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]。

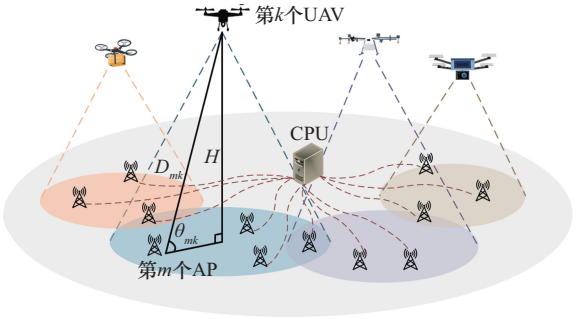


图1　以UAV为用户的无蜂窝大规模MIMO系统

1.1　信道模型

由于信道模型同时受大尺度衰落和小尺度衰落的影响，定义$g_{mk}\in\mathbb{C}$（$\mathbb{C}$表示复数集合）为第m个AP与第k个UAV之间的信道增益，表示为$g_{mk}=\sqrt{\beta_{mk}}h_{mk}$，其中$h_{mk}$为第$m$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$\mathrm{CN}(0,\boldsymbol{I}_N)$分布，大尺度衰落$\beta_{mk}$与链路状态相关。当第$m$个AP与第$k$个UAV之间为LoS链路时，则$\beta_{mk}=D_{mk}^{-\alpha_1}$，其发生概率为$p_{mk}^{\mathrm{L}}$；当为NLoS链路时，则$\beta_{mk}=D_{mk}^{-\alpha_2}$，其发生概率为$p_{mk}^{\mathrm{NL}}$。其中，$\alpha_1$和$\alpha_2$分别为LoS链路和NLoS链路的路径损耗指数，$D_{mk}$为第$k$个UAV与第$m$个AP之间的欧几里得距离，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，应考虑LoS链路和NLoS链路的随机性。因此，LoS链路传输的概率为[20]：

$$p_{mk}^{L}=\frac{1}{1+\xi_1\exp\{-\xi_2(\theta_{mk}-\xi_1)\}} \tag{1}$$

其中，ξ_1和ξ_2为取决于环境类型的常数参数，θ_{mk}为第m个AP与第k个UAV之间的仰角。在传输过程中，每条链路都有特定的发生概率，这主要取决于UAV与AP之间的相对位置、仰角以及环境因素。通过图1可以计算出仰角θ_{mk}为：

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

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

由式(2)可以得到，第m个AP与第k个UAV之间的平均大尺度衰落表示为：

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

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$\tau_p\times1$的导频序列向量$\sqrt{\tau_p}\boldsymbol{\varphi}_k$，其中$\|\boldsymbol{\varphi}_k\|^2=1$。因此，上行训练阶段第$m$个AP接收的导频信号为：

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

其中，ρ_p为各导频的平均发射功率，$\boldsymbol{w}_{p,m}$为第m个AP接收的$\tau_p\times1$加性白高斯噪声向量，其元素为独立同分布的复高斯随机变量，服从CN (0,1)分布。

基于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$\boldsymbol{y}_{p,m}$与导频向量$\boldsymbol{\varphi}_k^{\mathrm{H}}$相乘可得：

$$\hat{y}_{p,mk}=\boldsymbol{\varphi}_k^{\mathrm{H}}\boldsymbol{y}_{p,m}=\sqrt{\tau_p\rho_p}\,g_{mk}+\sqrt{\tau_p\rho_p}\sum_{k'\neq k}^{K}g_{mk'}\boldsymbol{\varphi}_k^{\mathrm{H}}\boldsymbol{\varphi}_{k'}+\boldsymbol{\varphi}_k^{\mathrm{H}}\boldsymbol{w}_{p,m} \tag{5}$$

g_{mk}的信道估计由MMSE估计得到，表示为：

$$\hat{g}_{mk}=\frac{\mathrm{E}\left\{\hat{\boldsymbol{y}}_{p,mk}^{*}g_{mk}\right\}}{\mathrm{E}\left\{\left|\hat{\boldsymbol{y}}_{p,mk}\right|^{2}\right\}}\hat{\boldsymbol{y}}_{p,mk}=C_{mk}\hat{\boldsymbol{y}}_{p,mk}\tag{6}$$

进一步地，可以推导出$\hat{g}_{mk}$的表达式为：

$$\hat{g}_{mk}=\frac{\sqrt{\tau_p\rho_p}\,\beta_{mk}}{\sqrt{\tau_p\rho_p}\sum\limits_{k'=1}^{K}\beta_{mk'}\left|\boldsymbol{\varphi}_k^{\mathrm{H}}\boldsymbol{\varphi}_{k'}\right|^{2}+1}\hat{\boldsymbol{y}}_{p,mk}\tag{7}$$

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

$$\gamma_{mk}\triangleq\mathrm{E}\left\{\left|\hat{g}_{mk}\right|^{2}\right\}=\frac{\tau_p\rho_p\beta_{mk}^{2}}{\tau_p\rho_p\sum\limits_{k'=1}^{K}\beta_{mk'}\left|\boldsymbol{\varphi}_{\mathrm{k}}^{\mathrm{H}}\boldsymbol{\varphi}_{k'}\right|^{2}+1}\tag{8}$$

根据以上推导，进一步得到估计误差$\tilde{e}_{mk}=g_{mk}-\hat{g}_{mk}$，其为服从均值为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仅对其关联集合$\mathcal{M}_k$内的UAV进行信道估计，因此关联集合$\mathcal{M}_k$不仅影响信道估计结果，还将影响后续的上行速率表达式。由此可见，本节的信道估计过程是后续速率推导和联合优化建模的重要基础。

1.3　上行链路数据传输

在上行链路传输过程中，各UAV以估计信道为真实信道，每个AP利用本地信道估计对接收信号进行复共轭加权，从而实现MRC。由于MRC仅依赖AP自身的局部信道状态信息，因此无须在AP与CPU之间交换瞬时信道状态信息，明显降低了系统的前传负载，显著提升了上行链路的整体接收性能。在上行链路中，第m个AP收到的信号为：

$$\boldsymbol{y}_{u,m}=\sqrt{\rho_u}\sum_{k=1}^{K}\sqrt{\eta_k}\,g_{mk}q_k+\boldsymbol{w}_{u,m}\tag{9}$$

其中，ρ_u为UAV的发射功率；η_k为第k个UAV的功率控制系数，满足$0\leqslant\eta_k\leqslant1$；$q_k$为第$k$个UAV发出的符号，满足$\mathrm{E}\left\{\left|q_k\right|^{2}\right\}=1$；假设$w_{u,m}\sim\mathrm{CN}(0,1)$为第$m$个AP收到的干扰噪声。

为检测第k个UAV的符号q_k，第m个AP利用其本地信道估计$\hat{g}_{mk}$的复共轭对接收信号$\boldsymbol{y}_{u,m}$进行加权处理。随后，所得量$\hat{g}_{mk}^{*}\boldsymbol{y}_{u,m}$通过前传链路发送至CPU，CPU收到的信号为：

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

基于上述计算，可以从$R_{u,k}$中检测出信号q_k。

1.4　可实现上行速率

每个UAV无法获取瞬时CSI，但可获取信道统计信息。因此，接收信号$R_{u,k}$可表示为：

$$R_{u,k}=\mathrm{DS}_kq_k+\mathrm{BU}_kq_k+\sum_{k'\neq k}^{K}\mathrm{UI}_{kk'}q_{k'}+\sum_{m\in\mathcal{M}_k}\hat{g}_{mk}^{*}\boldsymbol{w}_{u,m}\tag{11}$$

其中，DS_k为期望信号强度，BU_k为合并增益不确定性，$\mathrm{UI}_{kk'}$为用户干扰，对应的详细表达式分别为：

$$\mathrm{DS}_k=\sqrt{\rho_u}\,\mathrm{E}\left\{\sum_{m\in\mathcal{M}_k}\eta_k^{\frac{1}{2}}\hat{g}_{mk}^{*}g_{mk}\right\}\tag{12}$$

$$\mathrm{BU}_k=\sqrt{\rho_u}\left(\sum_{m\in\mathcal{M}_k}\eta_k^{\frac{1}{2}}\hat{g}_{mk}^{*}g_{mk}-\mathrm{E}\left\{\sum_{m\in\mathcal{M}_k}\eta_k^{\frac{1}{2}}\hat{g}_{mk}^{*}g_{mk}\right\}\right)\tag{13}$$

$$\mathrm{UI}_{kk'}=\sqrt{\rho_u}\sum_{m\in\mathcal{M}_k}\eta_k^{\frac{1}{2}}\hat{g}_{mk}^{*}g_{mk'}\tag{14}$$

由于式(11)中的后三项与第一项不相关，将后三项视为加性白高斯噪声，可以得到第k个UAV的可实现上行速率，其表达式为：

$$R_{u,k}=\mathrm{lb}\left(1+\frac{\left|\mathrm{DS}_k\right|^{2}}{\mathrm{E}\left\{\left|\mathrm{BU}_k\right|^{2}\right\}+\sum\limits_{k'\neq k}^{K}\mathrm{E}\left\{\left|\mathrm{UI}_{kk'}\right|^{2}\right\}+\mathrm{E}\left\{\left|\sum\limits_{m\in\mathcal{M}_k}\hat{g}_{mk}^{*}\boldsymbol{w}_{u,m}\right|^{2}\right\}}\right)\tag{15}$$

经计算可以求得：

$$\mathrm{DS}_k = \sqrt{p_u \eta_k} \sum_{m \in \mathcal{M}_k} \gamma_{mk} \tag{16}$$

$$\mathrm{E}\left\{\left|\mathrm{BU}_k\right|^2\right\} = p_u \eta_k \sum_{m \in \mathcal{M}_k} \gamma_{mk} \beta_{mk} \tag{17}$$

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

最终求得第 k 个 UAV 的上行总速率闭式表达式为：

$$R_{u,k} = \mathrm{lb}\left(1 + \frac{p_u \eta_k \left(\sum_{m \in \mathcal{M}_k} \gamma_{mk} \right)^2}{\left|\mathrm{DS}_k\right|^2} \times \left(\underbrace{p_u \sum_{k' \neq k}^{K} \eta_{k'} \left(\sum_{m \in \mathcal{M}_k} \gamma_{mk} \frac{\beta_{mk'}}{\beta_{mk}} \right)^2 \left| \varphi_k^{\mathrm{H}} \varphi_{k'} \right|^2}_{\mathrm{CI}_k} + \underbrace{p_u \sum_{k'=1}^{K} \eta_{k'} \sum_{m \in \mathcal{M}_k} \gamma_{mk} \beta_{mk'}}_{\mathrm{UI}_k} + \underbrace{\sum_{m \in \mathcal{M}_k} \gamma_{mk}}_{N_k} \right)^{-1} \right) \tag{19}$$

其中，$\left|\mathrm{DS}_k\right|^2$ 为第 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，其大小取决于上行发射功率、功率控制系数以及服务 AP 上的信道估计均方值；CI_k 为与第 k 个 UAV 复用相同导频的其他用户造成的相干干扰；UI_k 为所有 UAV 上行传输引入的非相干干扰及信道估计误差的影响；N_k 为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，两者均直接影响可达速率表达式，是后续联合优化问题的核心变量。

2 联合优化问题描述及求解

本节的研究目标是建立联合优化问题的数学表达式，并将其建模为 MDP，进而基于 DDPG 设计算法以寻找该问题的最优解。

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 的关联关系与上行功率控制，使所有用户在整个飞行周期内的平均上行速率之和最大化。前文定义的集合 $\mathcal{M}_k$ 表示第 k 个 UAV 与 AP 之间的关联关系。若第 m 个 AP 属于 $\mathcal{M}_k$，表明第 m 个 AP 与该 UAV 建立服务关联并参与其上行信号接收；否则，两者不建立服务关联。因此，对 $\mathcal{M}_k$ 的优化体现了 UAV-AP 关联关系的优化。由此，该优化问题可表示为：

$$\begin{aligned} \max_{\{\mathcal{M}_k, \eta_k\}} \quad & \sum_{k=1}^{K} R_{u,k} \\ \text{s.t.} \quad & 1 \leqslant \left|\mathcal{M}_k\right| \leqslant M, \quad \forall k \in \mathcal{K} \\ & 0 < \eta_k \leqslant 1, \quad k = 1,2,\cdots,K \end{aligned} \tag{20}$$

其中，$\left|\mathcal{M}_k\right|$ 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 MRC 可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，该联合优化问题式(20)具有高度非线性、非凸性以及高维决策空间等特征，传统凸优化方法难以求得可行解。为此，本文引入 DRL 框架，将 UAV-AP 关联与上行功率控制的联合优化问题建模为 MDP。

2.2 马尔可夫决策过程构建

在 DRL 框架中，MDP 为随机环境下的序列决策提供统一的建模架构，由状态空间、动作空间、奖励函数及状态转移概率构成，用于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，环境转移至下一状态 s_{t+1} 并返回即时奖励 r_t。学习目标是在与环境交互中更新策略 π，使折扣期望回报最大化，从而实现长期收益最优。在此框架基础上，本文进一步提出 DDPG 算法，该算法在处理连续状态空间和动作空间的复杂决策问题时具有显著优势[21]。

针对 UAV-AP 关联与上行功率控制的联合优化问题，本文在 DDPG 基础上提出一种全时域规划引导的 FA-DDPG（full-horizon alternating DDPG）算法。该算法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，则使用阶段一已训练好的聚类策略。在所构建的 MDP 框架下，状态空间、动作空间和奖励函数的具体定义如下。

（1）状态空间$\mathcal{S}$。在时隙t，状态$s_t \in \mathcal{S}$为一组可用于描述的参数。根据本文的系统模型，状态空间表示为$s_t = \{\mathbf{ctx}_t, a_{t-1}\}$。其中，将表示UAV与AP之间信道状态的大尺度衰落矩阵$\boldsymbol{\beta}_{mk}(t) \in \mathbb{R}^{K \times M}$向量化为$\boldsymbol{b}_t \in \mathbb{R}^{KM \times 1}$，构造全时域序列$\boldsymbol{B} = \{\boldsymbol{b}_0, \boldsymbol{b}_1, \cdots, \boldsymbol{b}_{T-1}\}$，并采用逐时隙共享编码器对每个$\boldsymbol{b}_t$进行一致的非线性压缩映射，得到低维表示$\boldsymbol{e}_t = \psi(\boldsymbol{b}_t) \in \mathbb{R}^{d \times 1}$。为显式刻画时隙顺序，在每个时隙引入位置编码$\boldsymbol{p}_t \in \mathbb{R}^{d \times 1}$，形成$\boldsymbol{x}_t = \boldsymbol{e}_t + \boldsymbol{p}_t$。在此基础上，通过Transformer时序汇聚对$\{\boldsymbol{x}_t\}$进行全局信息交互，得到上下文序列$\{\mathbf{ctx}_0, \mathbf{ctx}_1, \cdots, \mathbf{ctx}_{T-1}\}$，其中$\mathbf{ctx}_t \in \mathbb{R}^d$表示第$t$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$s_t = \{\mathbf{ctx}_t, a_{t-1}\}$。

（2）动作空间$\mathcal{A}$。为实现上行速率最大化，在每个决策时隙t，聚类智能体依据当前状态信息s_t输出聚类动作a_t^C，并将其映射为二进制连接矩阵$\boldsymbol{C}_t \in \mathbb{R}^{K \times M}$，其中元素$c_{k,m} = 0$表示不建立通信链路，$c_{k,m} = 1$表示第$k$个UAV选择第$m$个AP。功率控制智能体输出的功率控制动作为$a_t^P = [a_1^\eta, a_2^\eta, \cdots, a_k^\eta]$，其中$a_k^\eta \in [0,1]$为第$k$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：

$$r_t = \sum_{k=1}^{K} R_{u,k} \tag{21}$$

2.3　基于actor-critic框架的FA-DDPG算法

直接使用DRL算法求解连续空间、有限时间范围的MDP通常颇具挑战性，这主要源于状态和动作空间的连续性所增加的问题复杂度。为应对该挑战，本文提出一种适用于高维连续动作空间的两阶段FA-DDPG算法，以优化UAV-AP关联与上行功率控制。

如图2所示，FA-DDPG包含聚类和功率控制两个智能体模块，每个智能体均采用DDPG的actor-critic结构，并配备训练actor、训练critic及其对应的目标网络。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生成动作。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$\theta_a^{C'}$的目标actor网络和参数为$\theta_c^{C'}$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critic网络生成目标Q值，功率控制智能体同理。

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、$\theta_a^{C'}$、$\theta_c^{C'}$和θ_a^P、θ_c^P、$\theta_c^{P'}$、$\theta_a^{P'}$，并在actor-critic框架中生成动作。将当前状态s_t输入训练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。特别地，在进行聚类训练阶段，功率控制智能体只会做出一个固定的基线动作a_{baseline}^P，而在进行功率控制训练阶段，聚类智能体只会输出训练好的无噪声动作$a_t^C =$

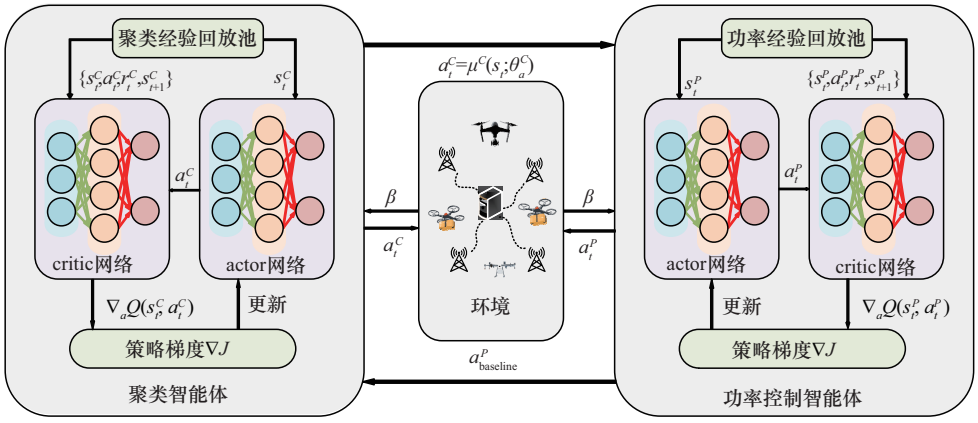


图2　FA-DDPG算法训练架构

$\mu^C(s_t;\theta_a^C)$。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_{t+1})，其目标Q值为：

$$y_i^C = r_i + \lambda Q^{C'}\left(s_i^{C'},\mu^C\left(s_i^{C'};\theta_a^{C'}\right);\theta_c^{C'}\right) \tag{22}$$

其中，$\lambda\in(0,1)$为折扣因子，目标actor网络和目标critic网络的参数$\theta_a^{C'}$和$\theta_c^{C'}$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：

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

通过最小化式(23)来更新聚类critic网络参数θ_c^C，并采用梯度下降更新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：

$$J^C(\theta_a^C)=\underset{s^C\sim\mathcal{N}^C}{}[Q^C(s^C,\mu^C(s^C;\theta_a^C);\theta_c^C)] \tag{24}$$

参数θ_a^C的梯度近似为：

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

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

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

其中，$\tau\ll 1$为软更新系数。在功率控制阶段进行相同的训练。通过上述双智能体交替FA-DDPG训练框架，聚类智能体和功率控制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。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$\mu^C(s_t;\theta_a^C)$、功率控制策略$\mu^P(s_t^C;\theta_a^P)$

1)初始化网络参数、经验回放池

2)第一阶段：聚类训练

3)while 聚类阶段迭代回合≤最大聚类迭代回合 do

4) 初始化环境，令$t\leftarrow 0$

5) while $t\leqslant T$ 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$\boldsymbol{C}_t$，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≤最大聚类迭代回合 do

16) 初始化环境，令$t\leftarrow 0$

17) while $t\leqslant T$ do

18) 根据状态s_t，通过第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) 调用功率控制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)end 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$\rho_p = 100\,\text{mW}$，每个UAV在上行链路的最大平均发射功率$\rho_u = 200\,\text{mW}$。式(1)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τ_p = 10个符号。RL相关参数设定如下：折扣因子λ = 0.99，软更新系数$\tau = 5 \times 10^{-4}$，actor网络学习率为5×10^{-5}，critic网络学习率为3×10^{-4}，经验回放缓冲区容量为10^5，采样的批量大小N_{batch}=256，聚类和功率控制阶段的最大迭代回合均为2 500，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，4种DRL算法的训练回报曲线。从图3(a)可以看出，在聚类阶段，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）在前期奖励值上升速度较快，但很快进入平台；深度Q网络（deep Q-network, DQN）能够提升系统回报但奖励值波动较大；优势演员−评论家（advantage actor-critic, A2C）奖励值整体提升有限，长期性能明显落后于其他3种算法。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)可以看出，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；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。图3表明，在本联合优化任务中，本文提出的FA-DDPG在聚类与功率控制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比较了不同算法的上行和速率与UAV数量的关系。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4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4可以看出，随着UAV数量K增加，各算法的总速率均有所提升，FA-DDPG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。

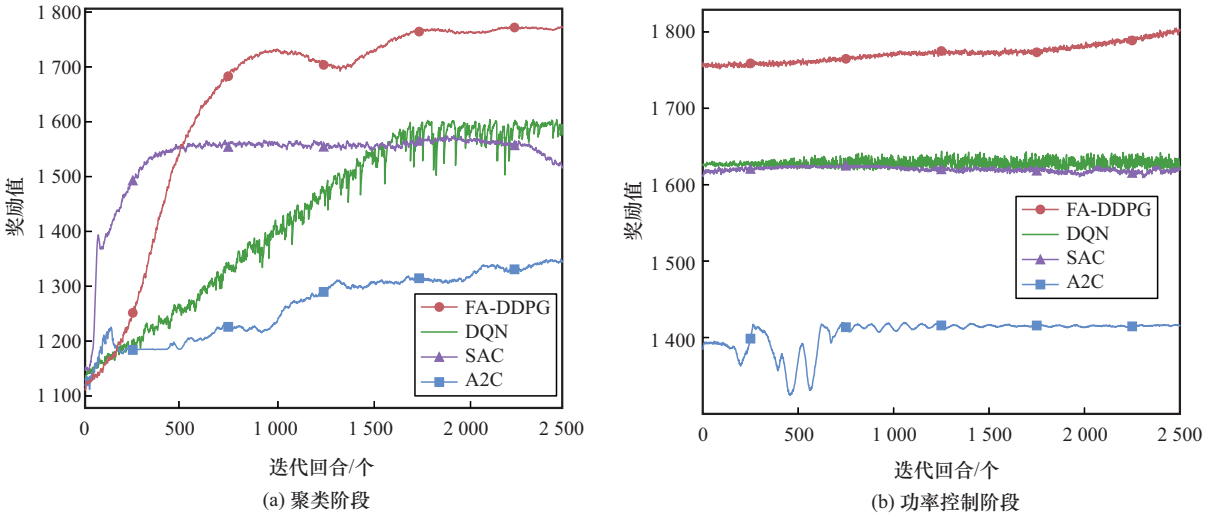


图3　不同RL算法的训练回报

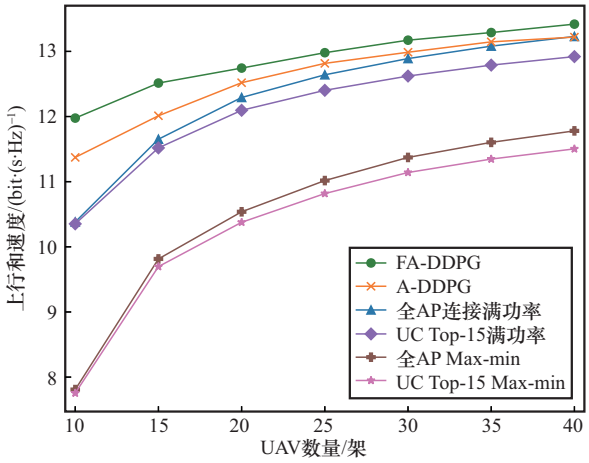


图4　不同算法的上行和速率与UAV数量的关系

图5展示了UAV的多重飞行轨迹，将AP数量M设定为20，均匀分布在边长为1 000 m的方形区域内，UAV飞行线速度为20 m/s。假设UAV沿3条同心圆形轨道飞行。其中，外环轨道参数固定，半径为450 m、高度为100 m；另外两条内环轨道与外环轨道同心，其半径R相同，R在50～400 m范围内逐步增大。2条内环轨道的飞行高度分别为100 m和200 m。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。

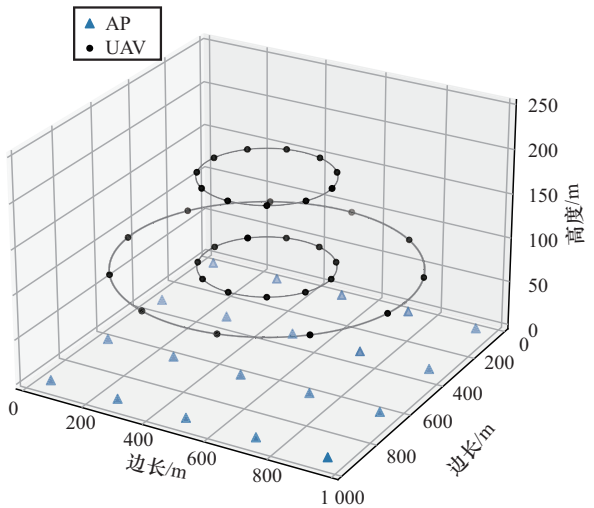


图5　UAV飞行轨迹

图6比较了不同算法下的上行和速率与UAV飞行内环轨道半径的关系。令K = 33，将内环轨道半径由50 m逐步增加至400 m，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，满功率全AP连接的基线算法的和速率随R增大单调上升，主要得益于内环轨道扩展后，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。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，3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和功率控制获得超过21 bit/(s·Hz)的性能增益；当R增大至200～400 m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1～5 bit/(s·Hz)的优势。上述结果表明，所提出的FA-DDPG能够根据不同轨道设定实时调整UAV-AP关联关系及上行发射功率，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。

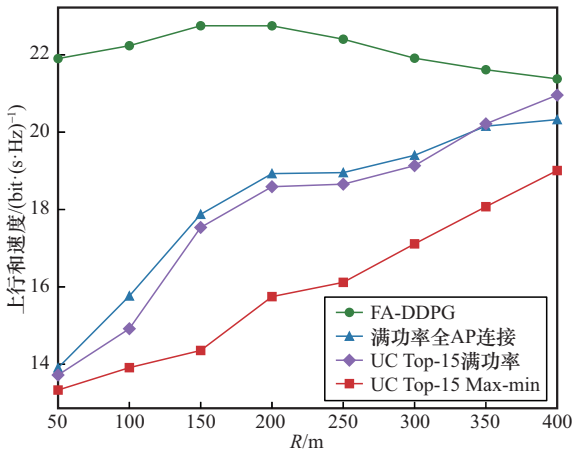


图6　不同方案下上行和速率随轨道半径的变化关系

图7进一步比较了不同AP数量下3种算法的上行和速率表现。为保证对比公平性，FA-DDPG与A-DDPG的奖励函数的设定与图4保持一致。可以看到，随着AP数量由10增加到50，3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，因此上行和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性地规划服务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部署密度具有良好的可扩展性与鲁棒性。

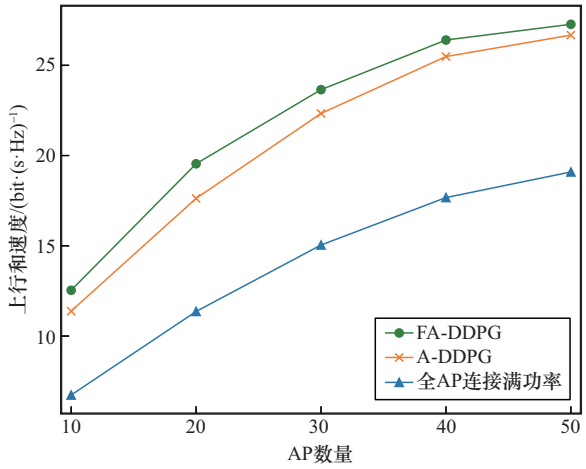


图7　不同方案下上行和速率随AP数量的关系

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率控制序列规划学习：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，FA-DDPG能够获得更高且更稳定的上行和速率增益，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 Z Y.《The 2024 low-altitude economy scenarios white paper》released[J]. China Science and Technology Fortune Magazine, 2024, (10): 29.

[2] Zeng Y, Zhang R, Lim T J. Wireless communications with unmanned aerial vehicles: opportunities and challenges[J]. IEEE Communications Magazine, 2016, 54(5): 36-42.

[3] 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.

[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 X Y, Sheng M, Li B, et al. Survey on unmanned aerial vehicle communications for 6G[J]. Journal of Electronics & Information Technology, 2022, 44(3): 781-789.

[5] Otto A, Agatz N, Campbell J, 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] Tsouros D C, Bibi S, Sarigiannidis P G. 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信道建模研究[J]. 通信学报, 2023, 44(7): 38-50.
Bai L, Sun M R, Huang Z W, et al. Multi-UAV cooperative channel model for emergency communication[J]. Journal on Communications, 2023, 44(7): 38-50.

[8] 姚昌华, 程田圆, 屈毓锛, 等. 面向异构复合任务的无人集群动态重叠联盟任务分配方法[J]. 信息网络安全, 2024, 24(2): 217-228.
Yao C C, Cheng T Y, Qu Y B, et al. A task allocation method for unmanned clusters based on dynamic overlapping coalition toward heterogeneous composite tasks[J]. Netinfo Security, 2024, 24(2): 217-228.

[9] Ngo H Q, Ashikhmin A, Yang H, 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 N H, et al. Ubiquitous cell-free massive MIMO communications[J]. EURASIP Journal on Wireless Communications and Networking, 2019, 2019: 197.

[11] Buzzi S, D'Andrea C. Cell-free massive MIMO: user-centric approach[J]. IEEE Wireless Communications Letters, 2017, 6(6): 706-709.

[12] Buzzi S, D'Andrea C, Fresia M, 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] Chen S F, Zhang J Y, Björnson E, 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] Cao H Y, Zeng H Y, Zhu X, 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] Ali K M S, Agnihotri S, Karthik R M. Joint AP-UE association and power factor optimization for distributed massive MIMO[C]//Proceedings of the 2024 IEEE 35th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). Piscataway: IEEE Press, 2024: 1-6.

[16] Zhang X Q, Le V A, Kaneko M, 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] Banerjee B, Elliott R C, Krzymieñ W A, 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] Du M J, Sun X H, Zhang Y, 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] Elhoushy S, Ibrahim M, Hamouda W. Cell-free massive MIMO: a survey[J]. IEEE Communications Surveys & Tutorials, 2022, 24(1): 492-523.

[20] Al-Hilo A, Samir M, Elhattab M, et al. RIS-assisted UAV for timely data collection in IoT networks[J]. IEEE Systems Journal, 2023, 17(1): 431-442.

[21] Luo L R, Zhang J Y, Chen S F, 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] Ngo H Q, Ashikhmin A, Yang H, et al. Cell-free massive MIMO: uniformly great service for everyone[C]//Proceedings of the 2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). Piscataway: IEEE Press, 2015: 201-205.

[作者简介]



张琦（1987– ），女，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。



贾旭日（2001– ），男，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。



蔡曙（1984– ），男，博士，南京邮电大学硕士生导师，主要研究方向为阵列和通信信号处理、统计信号处理、检测与估计理论。



张军（1980– ），男，博士，南京邮电大学教授、博士生导师，主要研究方向为未来移动通信理论与关键技术。



吴怡（1980– ），女，博士，福建师范大学教授、博士生导师，主要研究方向为光电传感、定位、自组织网络通信等。