



专栏：低空通信

用户为中心的多RIS辅助无人机协作传输机制

冯艳如¹, 闫俊杰², 刘汉永³

(1. 山西应用科技学院, 山西 太原 030062;

2. 广西科技大学电子工程学院, 广西 柳州 545006;

3. 国家电网重庆市电力公司大足供电分公司, 重庆 402360)

摘要: 针对多可重构智能超表面 (reconfigurable intelligent surface, RIS) 辅助的无人机 (unmanned aerial vehicle, UAV) 通信网络中用户链路需求与动态资源管理所带来的挑战, 提出了一种基于深度强化学习的用户中心化协作传输机制。构建MINLP模型, 联合优化UAV飞行轨迹、传输功率、RIS反射单元分配及相移策略, 在满足用户差异化服务质量需求的同时, 最大化系统能效。为解决传统方法复杂度高、动态环境适应性不足的问题, 设计了一种多阶段优化框架: 第一阶段采用联盟博弈算法实现RIS反射单元的动态分配, 并引入用户自主切换机制以降低同频干扰; 第二阶段利用双延迟深度确定性策略梯度 (twin delayed deep deterministic policy gradient, TD3) 算法, 联合优化资源分配与UAV轨迹。仿真结果表明, 所提方案在系统能效方面均优于对比基准方案, 且相较于SAC基线算法, 系统能效提升19%, 有效验证了该方案在动态环境下的鲁棒性与对用户需求差异化服务质量的保障能力。

关键词: RIS; UAV通信网络; 用户中心化; 联盟博弈; TD3

中图分类号: TN929.5

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2026100

User centric cooperative transmission mechanism in UAV-based multi-RIS networks

Feng Yanru¹, Yan Junjie², Liu Hanyong³

1. Shanxi College of Applied Science and Technology, Taiyuan 030062, China

2. School of Electronic Engineering, Guangxi University of Science and Technology, Liuzhou 545006, China

3. Dazu Power Supply Company of State Grid Chongqing Electric Power Company, Chongqing 402360, China

Abstract: To address the challenges of differentiated user link demands and dynamic resource management in unmanned aerial vehicle (UAV) communication networks assisted by multiple reconfigurable intelligent surfaces (RIS), a user-centric cooperative transmission mechanism based on deep reinforcement learning was proposed. A mixed integer nonlinear programming (MINLP) model was formulated to jointly optimize UAV flight trajectory, transmission

收稿日期: XXXX-XX-XX; 修回日期: XXXX-XX-XX

基金项目: 国家自然科学基金 (No. 62501178); 山西省高等学校科技创新项目 (No.2023L502)

Foundation Items: The National Natural Science Foundation of China(No. 62501178), Science and Technology Innovation Projects of Colleges and Universities in Shanxi Province (No.2023L502)



power, RIS reflection unit allocation, and phase shift control, aiming to maximize system energy efficiency while satisfying differentiated quality of service (QoS) requirements. To address the issues of high computational complexity and the limited adaptability to dynamic environments, a multi-stage optimization framework was designed. In the first stage, a coalition game-based algorithm was introduced to dynamically allocate RIS elements, and a user autonomous switching mechanism was introduced to mitigate co-channel interference. In the second stage, the twin delayed deep deterministic policy gradient (TD3) algorithm was adopted to jointly optimize resource allocation and UAV trajectory planning. Simulation results demonstrated that the proposed scheme consistently outperforms all benchmark methods in terms of system energy efficiency, achieving a 19% improvement compared to SAC algorithms, thereby effectively verifying its robustness in dynamic environments and its capability to guarantee differentiated QoS for users.

Key words: RIS, UAV communication network, user-centric, coalition game, TD3

0 引言

随着第五代移动通信 (the 5th generation mobile communication, 5G) 技术在全球范围内加速商用, 各行业对 5G 连接的需求日益呈现出多样化和复杂化的趋势。根据 Transforma Insights 的预测报告, 到 2027 年, 全球物联网设备规模预计将达到 300 亿台, 而我国物联网设备连接数也预计在 2027 年达到百亿规模。智能设备连接的激增和海量数据的实时处理, 已成为新一代蜂窝网络面临的重要挑战之一。无人机 (unmanned aerial vehicle, UAV) 因其灵活性高、成本低等优点而备受关注, 由 UAV 等飞行器构成的空中网络层, 被视为未来第六代移动通信 (the 6th generation mobile communication, 6G) 实现无缝、多维、全覆盖网络的重要补充^[1-2]。

如何在 UAV 执行飞行、监控、边缘服务等任务时, 为智能设备提供更丰富的计算、存储和通信资源, 是当前有关 UAV 研究的一个热点问题。尤其值得注意的是, 受体积、重量等因素限制, 无人机在能量、通信、缓存和计算等方面的资源往往很有限。此外, 在多无人机协同工作的场景下, 其资源表现出明显的异构性和动态性特征。为实现资源的高效利用, 无人机资源分配管理存在诸多挑战, 主要包括能效优化、频谱资源分配、轨迹优化以及计算资源分配等问题^[3-4]。除

此之外, 在建筑密集的城市环境中, UAV 与用户之间的直连链路容易受到障碍物遮挡, 从而严重影响 UAV 传输的性能。为应对 UAV 移动过程中因障碍物导致的无线信道传输性能下降问题, 可利用可重构智能超表面 (reconfigurable intelligent surface, RIS) 来改善无线信道条件^[5-6]。RIS 作为一种全新的技术, 通过在平面上集成大量低成本的无源反射元件, 能够智能重新配置无线传播环境, 从而显著提高无线通信网络的性能^[7]。

目前, 多数关于 RIS 赋能 UAV 的研究主要集中在单个 RIS 构建的小型网络优化问题^[8-9]。然而, 由于无人机具有动态移动的特性, 现有基于单 RIS 的 UAV 边缘网络方案难以持续保证用户链路的稳定需求。因此, 进一步研究多 RIS 赋能 UAV 无线网络的协同传输和计算卸载算法, 具有重要的理论意义和现实价值^[10-11]。为扩大信号覆盖范围并兼顾小区边缘用户, 许多研究将多个 RIS 与 UAV 结合作为空中中继, 通过构建特定控制信道来保障视距 (line-of-sight, LoS) 传输链路, 并建立多 RIS 下行链路通信模型, 从而为用户提供无缝、协同的多点网络覆盖^[12-13]。

然而, 现有关于多 RIS 赋能 UAV 网络的研究多从网络自身性能出发, 例如多 RIS 协助下最大化系统总传输速率^[14], 缺乏从用户视角出发的系统性考量。随着无人机通信应用场景日趋多样

化、用户需求日趋个性化，仅从网络侧进行优化已难以全面满足复杂环境下用户对通信质量、服务可靠性及体验差异化的要求。因此，有必要从用户视角出发，主动感知用户位置、业务类型及服务需求，动态调整 RIS 协作方式和 UAV 部署策略，实现面向用户体验优化的自适应网络构建。

为满足用户对链路的差异化需求，本文针对多 RIS 辅助的 UAV 通信场景，研究以用户为中心的多点协作传输策略。在该策略中，用户可根据自身连接需求自主选择合适的水RIS节点以接入服务。此外，考虑到无人机移动轨迹的动态性、资源的有限性与异构性，本文采用深度确定性策略梯度算法，实现对 UAV 轨迹、多 RIS 归属关系及相位的动态分配。本文的主要贡献如下。

(1) 针对多 RIS 辅助的 UAV 通信场景，提出一种以用户为中心的差异化链路需求保障机制。该机制允许地面用户根据实时信道状态与服务质量 (quality of service) 需求，动态选择最优的水RIS节点集合进行关联。

(2) 为实现上述目标，建立了一个联合优化模型，将功率分配、RIS 反射单元分配、相位调控以及 UAV 轨迹规划整合为一个多变量耦合的混合整数非线性规划 (mixed integer nonlinear programming, MINLP) 问题。针对 MINLP 问题的高复杂度与动态环境适应性不足的挑战，采用交替优化方法将原问题解耦为 RIS 反射单元动态分配与资源-轨迹联合优化两个子问题。

(3) 设计了一个基于联盟博弈和双延迟深度确定性策略梯度 (twin delayed deep deterministic policy gradient, TD3) 算法的多阶段优化框架。第一阶段，通过联盟博弈算法获得稳定的 RIS 单元分配方案；第二阶段，采用 TD3 算法优化剩余参数，通过交替迭代得到稳定且能效最大化的策略。

1 系统模型

多 RIS 辅助下的无人机边缘网络多点联合传输模型如图 1 所示。在基站覆盖区域构建一架 UAV u 服务 I 个地面用户的下行链路， u 在时间周期 T 内以固定高度 Z_u 飞行，在不同的时间间隔为用户提供服务。为便于分析，将 u 的 T 分为 N 个等距时间间隔，即 $\mathbf{N} = \{1, \dots, N\}$ ，步长为 λ ，则时间周期 $T = N\lambda$ 。同时，为了时刻保证用户对连接能力的差异化要求，网络中部署 K 个 RIS 来辅助用户，其中 $\mathbf{K} = \{1, \dots, K\}$ 。RIS 均配有 M 个反射单元，并且每个反射单元的相位都可以由 UAV 控制，则第 k 个 RIS 在第 $n \in \mathbf{N} = \{1, \dots, N\}$ 时隙中的相移矩阵为 $\Theta_k[n] = \text{diag}\{e^{j\theta_{k1}[n]}, e^{j\theta_{k2}[n]}, \dots, e^{j\theta_{km}[n]}\}$ ，其中， $\theta_{km}[n]$ 是在第 n 时隙中第 k 个 RIS 的第 m 个反射单元的相移，且 $\theta_{km}[n] \in [0, 2\pi)$ ， $m \in \mathbf{M} = \{1, 2, \dots, M\}$ 。

假设 u 和用户 $i \in \mathbf{I} = \{1, 2, \dots, I\}$ 均配置单个天线，不失一般性，所有通信节点都放置在三维笛卡尔坐标系中，在第 n 个时隙中 u 、 i 和第 k 个 RIS 的水平坐标分别为 $H_u[n] = [x_u^n, y_u^n]^T$ ， $H_i[n] = [x_i^n, y_i^n]^T$ ， $H_k[n] = [x_k^n, y_k^n]^T$ 。为定期向用户提供服务，UAV 在 T 内从初始位置到最终位置的约束条件为：

$$\|H_u[n+1] - H_u[n]\| \leq D_u^{\max} \quad (1)$$

$$H_u[0] = H_u^S, H_u[N+1] = H_u^F \quad (2)$$

其中， H_u^S 和 H_u^F 分别表示 u 的初始和最终水平位置，UAV 在每个时隙飞行的最大水平距离为 $D_u^{\max} = v_u^{\max} \lambda$ ， v_u^{\max} 分别表示 u 的最大飞行速度。

1.1 UAV-RIS-UE 传输模型

一般情况下，UAV 到地面用户的信道以 LoS 传播信道为主，而在特殊场景中，空地 LoS 信道容易遭受堵塞。在多 RIS 辅助空中地面通信方案中，部署多个 RIS 可以改善环境，通过相位反射将传输的信号转移到地面用户。假设 $G_{u,i}$ 表示

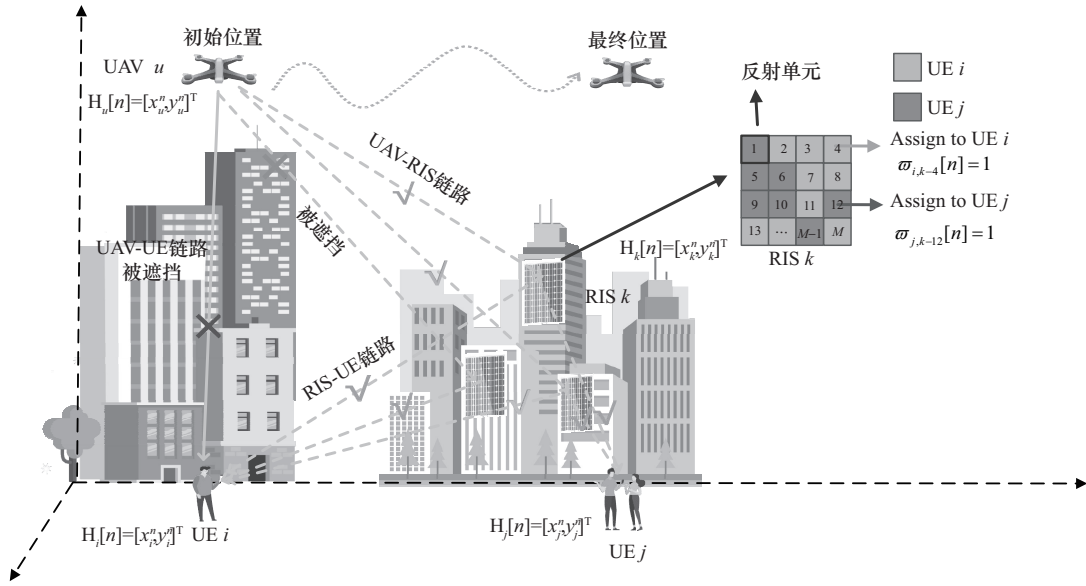


图1 多RIS辅助下的无人机边缘网络多点联合传输模型

u 到用户 i 的无线信道增益, $G_{u,k}$ 表示 u 到第 k 个RIS的无线信道增益, $G_{k,i}$ 表示第 k 个RIS到用户 i 的无线信道增益, 则第 n 个时隙 u 到用户 i 的无线信道增益 $G_{u,i}[n]$ 可以表示为:

$$G_{u,i}[n] = \sqrt{\varphi d_{u,i}^{-\varepsilon_{u,i}}} \phi_{u,i} \in \mathbb{C}^{1 \times 1} \quad (3)$$

其中, φ 表示在参考距离 $d_0 = 1$ m处的信道功率增益, $d_{u,i}$ 表示 u 到用户 i 的距离, $\phi_{u,i}$ 表示零均值和单位方差复高斯随机变量, $\varepsilon_{u,i}$ 表示无人机与用户路径损耗指数。

同理, 第 n 个时隙 u 到第 k 个RIS的无线信道增益 $G_{u,k}[n]$ 可以表示为:

$$G_{u,k}[n] = \sqrt{\varphi d_{u,k}^{-\varepsilon_{u,k}}} \phi_{u,k}[n] \in \mathbb{C}^{M \times 1} \quad (4)$$

$$\phi_{u,k}[n] = \left[1, e^{-j\frac{2\pi}{v}d_k\delta_{u,k}[n]}, \dots, e^{-j\frac{2\pi}{v}(M-1)d_k\delta_{u,k}[n]} \right]^T \quad (5)$$

其中, $\varepsilon_{u,k}$ 表示 u 与第 k 个RIS的路径损耗指数, $d_{u,k}$ 表示 u 到第 k 个RIS的距离, $\phi_{u,k}[n]$ 表示第 n 个时隙第 k 个RIS的响应阵列, $\delta_{u,k}[n] = \frac{x_u^n - x_k^n}{d_{u,k}}$ 表示第 n 个时隙无人机 u 与第 k 个RIS发射角度 (angle of departure, AoD) 的余弦, v 表示波长, d_k 表示天线的间隔。

同理, 在第 n 个时隙中, 第 k 个RIS到用户 i 的无线信道增益 $G_{k,i}[n]$ 可以表示为:

$$G_{k,i}[n] = \sqrt{\varphi d_{k,i}^{-\varepsilon_{k,i}}} \phi_{k,i}[n] \in \mathbb{C}^{M \times 1} \quad (6)$$

其中, $\varepsilon_{k,i}$ 表示第 k 个RIS与用户 i 的路径损耗指数, $d_{k,i}$ 表示第 k 个RIS到用户 i 的距离, $\phi_{k,i}[n]$ 表示第 n 个时隙第 k 个RIS到用户 i 的响应阵列。

第 n 个时隙第 i 个用户通过第 k 个RIS接收来自UAV的反射链路信号增益, 其数学表达式如下:

$$G_i^k[n] = G_{u,k}[n] \Theta_k[n] G_{k,i}[n] \quad (7)$$

为了评估系统在受遮挡下的传输性能, 构建了一种在城市环境下的空地信道模型^[15]。其中, u 与用户 i 被遮挡的概率为:

$$\Pr_{u-i} = 1 - \frac{1}{1 + \alpha \exp\left(-\beta\left(\arctan\left(\frac{Z_u}{d_{u,i}}\right) - \alpha\right)\right)} \quad (8)$$

其中, α 和 β 是取决于环境的恒定值 (文献[15]中 $\alpha = 9.61$, $\beta = 0.16$)。

因此, 第 i 个用户的平均信道增益可以表示为:

$$G_i^k = (1 - \Pr_{u-i})G_{u,i} + \Pr_{u-i}G_{u,k}\Theta_k G_{k,i} \quad (9)$$

为满足用户对链路的差异化需求, 基于无人

机功率控制，构建以用户为中心的多点协作传输策略。用户 i 可以接收多个 RIS 为其提供服务。假设 $\Gamma_{I,K-M}[n]=\{\varpi_{i,k-m}[n], i \in I, k \in K, m \in M\}$ ，其中， $\Gamma_{I,K-M}[n]$ 表示第 n 个时隙用户与 RIS 的关联矩阵， $\varpi_{i,k-m}[n]$ 表示第 n 个时隙用户 i 与第 k 个 RIS 的第 m 个反射单元之间的关联系数。如果 $\varpi_{i,k-m}[n]=1$ ，则表示第 n 个时隙第 k 个 RIS 的第 m 个反射单元分配给了用户 i ；否则， $\varpi_{i,k-m}[n]=0$ 。考虑到分配到的 RIS 与 UAV 或 UE 存在非视距 (non-line-of-sight, NLoS) 链路，本文限制了用户 i 分配 RIS 的最大数量。假设 RIS 与 UAV 或 UE 的信道分布服从 Rician 衰落^[16]， $\mu_{u,k-m}$ 和 $\mu_{k,m-i}$ 分别表示 UAV 与 RIS、RIS 与用户 i 的 Rician 系数。可以为用户 i 分配的 RIS 数量为：

$$\Phi_{i,k} = \frac{\Gamma_{I,K-M}}{M} = \frac{\sum_{k=1}^K \sum_{m=1}^M \varpi_{i,k-m}}{M} \leq K \cdot \sqrt{\frac{\mu_{u,k-m}}{\mu_{u,k-m}+1}} \sqrt{\frac{\mu_{k,m-i}}{\mu_{k,m-i}+1}} \quad (10)$$

第 n 个时隙用户 i 接收到 Φ_i 个相关联的 RIS 协助后，总的信道增益为：

$$G_i[n] = \sum_{k=1}^K \sum_{m=1}^M \varpi_{i,k-m} G_i^k[n] \quad (11)$$

第 n 个时隙用户 i 总的传输速率为：

$$R_{u,i}[n] = B \log_2(1 + [\text{SINR}]) \quad (12)$$

其中， B 为信道带宽， $[\text{SINR}]$ 表示第 n 个时隙接收相关联 RIS 协助后用户 i 的信干噪比 (signal-to-interference-plus-noise ratio, SINR)，可以表示为：

$$[\text{SINR}] = \frac{P_{u,i} |G_i[n]|^2}{\sum_{q=1, q \neq i}^m P_{u,q} |G_q[n]|^2 + \sigma^2} \quad (13)$$

其中， $P_{u,i}$ 表示 u 对用户 i 的传输功率， $\sum_{q=1, q \neq i}^m P_{u,q} |G_q[n]|^2$ 表示第 n 个时隙其他用户 q 与用户 i 复用相同 RIS 造成的干扰， σ 表示加性高斯白

噪声的功率。

1.2 UAV 能耗模型

尽管无人机通信具有视距传输信道和高移动性的优势，但其续航能力通常有限。因此，在研究无人机通信系统时，需要充分考虑其能耗问题。无人机的服务能耗主要包括传输能耗和飞行能耗。若定义第 n 个时隙无人机的总能耗为 $E_u^{\text{consume}}[n]$ ，则可将其表示为无人机传输能耗 $E_u^{\text{Trans}}[n]$ 和飞行能耗 $E_u^{\text{fly}}[n]$ 之和，即：

$$E_u^{\text{consume}}[n] = E_u^{\text{Trans}}[n] + E_u^{\text{fly}}[n] \quad (14)$$

(1) 传输能耗：无人机通过无线回程链路进行数据传输，其能耗与传输的数据总量成正比。在第 n 个时隙，无人机传输能耗 $E_u^{\text{Trans}}[n]$ 可表示为：

$$E_u^{\text{Trans}}[n] = \begin{cases} \sum_{i=1}^{I'} P_{u,i}[n] \lambda, & \mu_{u,i}[n] \geq R_{u,i}[n] \\ \sum_{i=1}^{I'} P_{u,i}[n] \frac{\mu_{u,i}[n]}{R_{u,i}[n]}, & \mu_{u,i}[n] < R_{u,i}[n] \end{cases} \quad (15)$$

其中， I' 表示无人机在第 n 个时隙服务的用户个数， $\mu_{u,i}[n]$ 表示用户 i 在当前时隙的剩余文件大小。

(2) 飞行能耗：旋翼无人机需要消耗推进能量以维持悬停和飞行。根据无人机运动模型，其在飞行速度 v 下的功率消耗^[16]可表示为：

$$P^{\text{fly}}[n] = P_0 \left(1 + \frac{3v^2}{U_{\text{tip}}^2} \right) + P_1 \left(\sqrt{1 + \frac{v^4}{4v_0^4}} - \frac{v^2}{2v_0^2} \right)^{\frac{1}{2}} + \frac{1}{2} d_0 \rho s v^3 \quad (16)$$

其中， P_0 、 U_{tip} 、 P_1 、 v_0 、 d_0 、 ρ 、 s 和 Δ 分别表示基于无人机重量、机翼面积、空气密度等因素相关的常数，具体见表 1。

对应的，无人机 u 在任意时隙 n 的飞行功耗为：

$$E_u^{\text{fly}}[n] = P^{\text{fly}}(\|v_n[n]\|) \quad (17)$$

无人机在实际飞行中时受到电池容量 E_u^{max} 的



限制。因此，在第 n 个时隙，无人机的剩余能量为最大能量减去无人机 u 的飞行能耗与传输能耗，可表示为：

$$E_u^{\text{rem}}[n] = \left[E_u^{\text{max}} - \int_0^{\lambda n} E_u^{\text{fly}}[n] dt - E_u^{\text{Trans}}[n] \right] \quad (18)$$

2 问题构建

为满足用户对链路的差异化需求，允许用户根据自身连接需求，自主选不同 RIS 为自己提供连接服务。同时，无人机也可以根据用户的差异化需求，结合传输功率、RIS 相位矩阵、无人机轨迹的调节，以最大化能效为目标，进行相关算法设计。该优化问题可以建模为：

$$\begin{aligned} & \max_{\Gamma_{L,K-M}, P, \Theta, H_u} \frac{\sum_{n=1}^N \sum_{i=1}^I R_{u,i}[n]}{\sum_{n=1}^N E_u^{\text{consume}}[n]} \quad (19) \\ \text{s.t. } & \text{C1: } 0 < x_u^n < x_{\text{max}} \\ & \text{C2: } 0 < y_u^n < y_{\text{max}} \\ & \text{C3: } \|H_u[n+1] - H_u[n]\| \leq D_u^{\text{max}} \\ & \text{C4: } \Phi_{i,k} \leq K \cdot \sqrt{\frac{\mu_{u,k-m}}{\mu_{u,k-m} + 1}} \sqrt{\frac{\mu_{k,m-i}}{\mu_{k,m-i} + 1}} \\ & \text{C5: } P_{u,i} \leq P_u^{\text{max}} \\ & \text{C6: } R_i[n] \geq R_{\text{min}}^i \\ & \text{C7: } |\theta_{km}[n]| = 1 \\ & \text{C8: } \theta_{km}[n] \in [0, 2\pi) \end{aligned}$$

其中，C1、C2 表示无人机 u 的空间约束条件；C6 表示用户 i 的速率需求约束， R_{min} 表示用户的基本需求；C7 表示反射单元的增益约束。

该问题需要先优化 RIS 与用户的关联矩阵 $\Gamma_{L,K-M}[n]$ ，再根据关联结果，联合优化传输功率 P 、RIS 相位 Θ 和无人机轨迹 H_u ，以最大化系统能效。然而，由于目标函数的非线性和约束条件的非凸性，问题（式（19））是一个多变量耦合的非凸优化问题，很难直接求得最优解。

3 基于多阶段的问题解决方案

考虑到直接求解式（19）具有较高的复杂性，并且在动态变化的通信环境中，传统优化方法难以实现自适应的资源调度。为此，本文提出一种交替优化的方式，将原问题分解为两个阶段依次求解：第一个阶段，基于联盟博弈理论动态优化 RIS 与用户的分配问题；第二阶段，在已知 RIS 与用户分配方案的基础上，采用 TD3 算法对无人机的传输功率、RIS 相位、无人机飞行轨迹进行联合智能优化，以最大化系统能效。基于联盟博弈和 TD3 算法的多阶段优化过程如图 2 所示。

3.1 基于联盟博弈的 RIS 分配问题求解

如第 1.1 小节所述，为满足用户对通信链路的差异化需求，用户 i 可同时接受多个 RIS 的服务。然而，受无人机动态飞行影响，RIS 与 UAV 或 UE 之间的链路也随之动态变化。为避免链路因遮挡等原因落入 NLoS 状态，需要动态地为用户更新与其关联的 RIS 集合。为此，本小节采用联盟博弈算法来求解 RIS 动态分配问题。联盟博弈（coalition game）是指决策者以同盟、合作的方式与其他决策者形成稳定联盟的过程，经常用来求解无线网络中的动态资源分配问题^[17]。

在本文场景中，当多个用户共享同一个 RIS，将引发严重的同频干扰，导致各用户传输速率和服务满意度的急剧下降。因此，用户倾向于选择负载较轻的 RIS，或从高干扰链路切换到低干扰链路。用户通过不断调整所关联的 RIS 组合，以追求个体效用的最大化，这一过程与联盟的形成和调整机制高度吻合。

3.1.1 用户与 RIS 联盟博弈模型构建

在本文框架中，用户可以根据自身连接需求自主加入由不同 RIS 构成的联盟，以获取通信服务。因此，RIS 分配问题可建模为联盟博弈。具体而言，联盟可以表示为 $C = \{C_1, C_2, \dots, C_I\}$ ，其

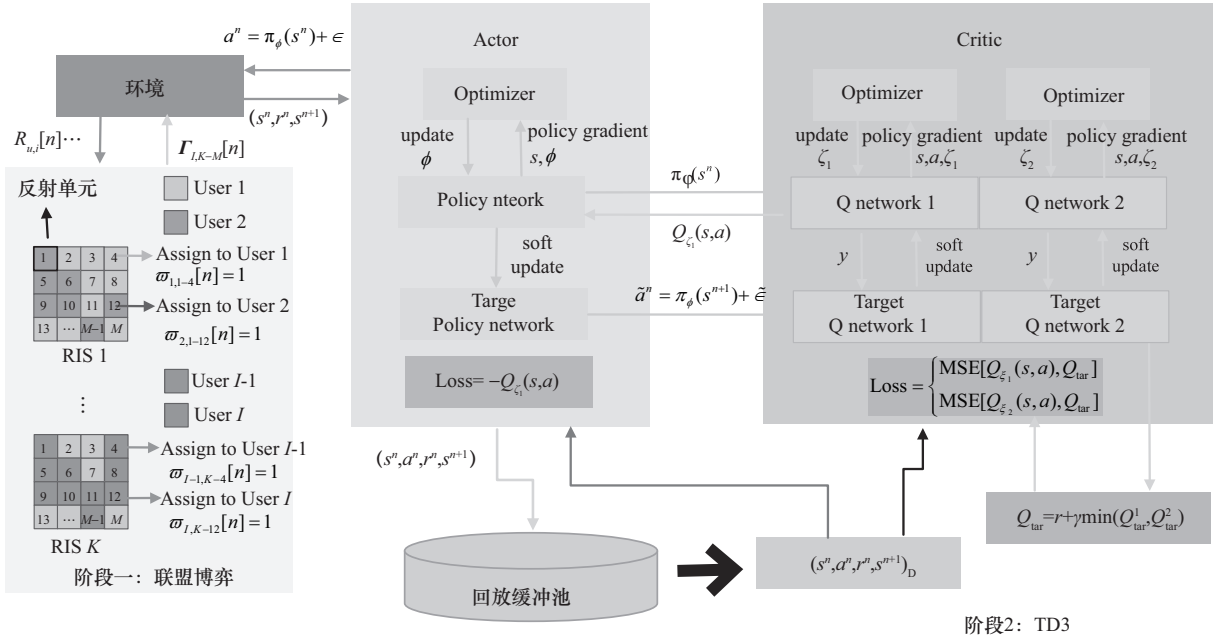


图2 基于联盟博弈和TD3算法的多阶段优化过程

中, $C_i \cap C_j = \emptyset, i \neq j \in I$ 。此外, 联盟博弈模型包括两个要素, 一是需要形成联盟的参与者集合; 二是联盟收益, 用来定量描述联盟的价值, 其定义直接影响联盟博弈的性质与求解目标。下面给出本文RIS动态分配问题中参与者集合和联盟收益的具体定义。

定义1: 本文RIS分配问题可表述为一个可转移效用的联盟博弈 (Ω, U) 。在联盟博弈中, Ω 表示参与者集合, 即RIS反射单元集合, $\Omega = \{(\Omega_{1-1}, \dots, \Omega_{1-M}), \dots, (\Omega_{K-1}, \dots, \Omega_{K-M})\}^T$; U 表示联盟收益。

考虑到多用户共享同一个RIS时, 会因同频干扰加剧而导致传输速率下降, 为此, 本文将每个联盟的传输速率作为该联盟的收益, 即

$$U_{C_i}[n] = R_{u,i}[n] \quad (21)$$

基于此, RIS动态分配问题可以转化为如下联盟结构优化问题:

$$\max_{\Gamma_{L,K-M}} \sum_{i=1}^I U_{C_i}[n] \quad (22)$$

$$\text{s.t. } \Phi_{i,k} \leq K \cdot \sqrt{\frac{\mu_{u,k-m}}{\mu_{u,k-m} + 1}} \sqrt{\frac{\mu_{k,m-i}}{\mu_{k,m-i} + 1}}$$

3.2.2 求解策略

在联盟博弈中, 可以随机分配不同的反射单元到不同的用户联盟中形成初始的联盟分区。在获得初始联盟分区后, 各反射单元可以通过“离开当前用户联盟、加入另一个用户联盟”来更新联盟结构, 形成新的联盟分区, 即联盟切换操作 (switch operation), 如定义2所示。

定义2: 联盟切换操作。切换操作 $\chi_{C_i \rightarrow C_j}^{Q_{k-m}}$ 表示第 k 个RIS的第 m 个反射单元离开当前联盟 C_i , 加入到另一个联盟 C_j 。此时, 联盟 C_i 和 C_j 都发生了变化, 即 $C'_i = C_i \setminus Q_{k-m}, C'_j = C_j \cup Q_{k-m}$ 。为衡量切换操作的优劣, 定义切换操作的切换收益为:

$$U_{C'_i, C'_j}^{Q_{k-m}} = U_{C'_i} + U_{C'_j} - U_{C_i} - U_{C_j} \quad (22)$$

在联盟博弈中通常根据切换收益定义参与者的偏好关系, 如定义3所示。

定义3: 当切换操作 $\chi_{C'_i \rightarrow C'_j}^{Q_{k-m}}$ 产生的切换收益大于另一个切换操作 χ' 的切换收益时, 就会偏向于选择切换操作 $\chi_{C'_i \rightarrow C'_j}^{Q_{k-m}}$ 。此时, 切换操作的偏好



关系 \succ 可以表示为:

$$\chi_{\mathcal{C}_i, \mathcal{C}_j}^{\Omega_{k-m}} \succ \chi' \Leftrightarrow U_{\mathcal{C}_i \rightarrow \mathcal{C}_j}^{\Omega_{k-m}} > U' \quad (23)$$

基于上述偏好关系, 基于联盟博弈的反射单元动态分配算法通过多轮迭代执行切换操作, 如算法1所示。在每一轮迭代中, 每个RIS反射单元遍历所有可能的联盟, 若发现满足偏好关系的切换操作则立即执行并更新联盟结构。由于系统总速率存在物理上限, 且每次切换操作均带来正向收益, 该博弈过程最终将收敛至一个纳什稳定状态, 即没有任何反射单元可以通过单方面切换联盟来增加系统总收益。

算法1 基于联盟博弈的反射单元动态分配算法

输入: 用户联盟 $\mathbf{C} = \{C_1, C_2, \dots, C_I\}$, RIS反射单元集合 Ω , 最大迭代次数

输出: 稳定联盟分区 \mathbf{C}^*

初始化: 随机分配不同的反射单元到各用户联盟, 形成初始联盟分区, 且满足 $\Phi_{i,k} \leq K$.

$$\sqrt{\frac{\mu_{u,k-m}}{\mu_{u,k-m}+1}} \sqrt{\frac{\mu_{k,m-i}}{\mu_{k,m-i}+1}}$$

当iter从1遍历至MaxIter时:

- (1) 根据式(21)计算每个联盟的收益
- (2) For Ω_{k-m} in Ω
- (3) 查找当前所属联盟 \mathcal{C}_i
- (4) 遍历所有可能的目标联盟 $\mathcal{C}_j (\mathcal{C}_j \neq \mathcal{C}_i)$
- (5) 执行切换操作 $\chi_{\mathcal{C}_i, \mathcal{C}_j}^{\Omega_{k-m}}$ 后根据式(22)计算切换收益
- (6) If $\chi_{\mathcal{C}_i, \mathcal{C}_j}^{\Omega_{k-m}} > 0$ then
- (7) 将反射单元 Ω_{k-m} 归属到新的联盟 \mathcal{C}_j , 并终止遍历
- (8) End if
- (9) End for
- (10) 计算当前总效用, 总效用不再上升时终止循环
- (11) End for

3.2 基于TD3的优化问题求解

通过上述联盟博弈过程, 可以实现RIS反射单元在用户联盟中的动态分配。由于无人机位置实时移动, 还需进一步动态调整RIS相位、无人机轨迹, 以最大化目标函数。然而, 目标函数的优化问题具有多约束、非凸的特性, 传统优化方法难以直接求解。为此, 本文采用TD3算法进行求解。在该框架中, 系统被视为一个马尔可夫决策过程 (Markov decision process, MDP), 智能体通过与未知环境的交互学习最优策略, 以最大限度地提高回报。本文将目标函数最大化问题作为奖励函数的设计目标, 即以最大化UAV的能效作为强化学习的优化导向。具体而言, 将多RIS辅助的无人机边缘网络建模为环境, UAV本身被视为智能体, 通过与环境交互获取状态、执行动作并获得奖励。TD3算法在传统DDPG (deep deterministic policy gradient) 算法基础上引入了双重Q网络、延迟策略更新和目标策略平滑等机制, 能有效缓解Q值过估计问题, 从而提升学习的稳定性和最终性能。接下来, 将详细阐述如何将原优化问题转化为MDP形式, 并对TD3算法在该场景下的具体设计与实现进行说明。

在动态网络环境中, 本文采用无模型强化学习算法, 通过对环境交互数据的学习来估计状态-动作值函数 (即Q值函数)。首先, 构建环境模型。在该模型中, UAV的位置和RIS的相位值都具有马尔可夫性质, 可将其映射到MDP的交互环境, 网络结构如上述图2所示。在数学上, MDP被定义为四元组 $\langle \mathbf{S}, \mathbf{A}, p, r \rangle$, 即状态空间、动作空间、状态转移概率和奖励函数。在每个时隙 n 中, 智能体从环境中获取当前状态 s^n , 执行动作 a^n , 与环境交互后转移到下一个状态 s^{n+1} , 并得到奖励值 r^n 。UAV将这种迭代转换作为经验存储在回放缓冲池中, 并从回放缓冲池中随机采样用于训练神经网络, 即Actor神经网络和Critic神经网络, 以分别逼近动作函数和

动作值函数。UAV 的状态空间、动作空间和奖励设计如下。

(1) 状态空间

\mathcal{S} 为智能体状态的集合, s^n 表示智能体在时隙 n 的状态, $s^n \in \mathcal{S}$, 由当前 UAV 的坐标、用户剩余数据量和 UAV 剩余能量组成, 即 $s^n = \{H_u[n], \mu_k[n], E_u^{\text{rem}}[n]\}$ 。

(2) 动作空间

\mathcal{A} 为智能体动作的集合, a^n 表示智能体在时隙 n 的动作, $a^n \in \mathcal{A}$, 智能体的动作由四部分组成: UAV 速度 $v_u[n]$, UAV 角度 $\theta^n \in (0, 2\pi)$, RIS 相位 $\Theta_k[n]$, 用户传输功率 $P_{u,i}[n]$ 。

(3) 奖励设计

TD3 目标是最大化奖励的累积期望, 本文以 UAV 的能效作为总奖励。同时, 还考虑了 UAV 速度和移动范围、功率和最小速率等约束的影响。因此, 奖励函数可以表示为:

$$r = \frac{\sum_{n=1}^N \sum_{i=1}^I R_{u,i}[n]}{\sum_{n=1}^N E_u^{\text{consume}}[n]} - C_0 \quad (24)$$

其中, C_0 表示违反约束时作为惩罚的常数。

TD3 算法作为 DDPG 算法的改进版本, 通过三重协同优化机制显著提升了智能体在连续动作空间中的学习效率与稳定性。TD3 算法的主要突破在于对传统 Actor-Critic 架构的系统性改进, 即针对 DDPG 中普遍存在的价值高估问题, 创新性地引入双重 Critic 评估、延迟策略优化和目标策略平滑机制, 形成多维度的动态平衡机制^[18]。

在算法初始化阶段, 系统构建了六组深度神经网络构成的并行架构, 包含主 Actor 网络 ϕ 、两组独立 Critic 网络 Q_{ζ_1} 和 Q_{ζ_2} 及其对应目标网络 ($\phi', \zeta_1', \zeta_2'$)。所有目标网络参数通过镜像映射与主网络保持一致:

$$\zeta_1' \leftarrow \zeta_1, \zeta_2' \leftarrow \zeta_2, \phi' \leftarrow \phi \quad (25)$$

这一设定确保了目标网络与主网络在训练初期具有一致的行为模式, 而参数空间的解耦可以有效规避单一评估网络的认知偏差。

探索阶段采用动态噪声注入策略, 将高斯扰动 $\epsilon \sim N(0, \sigma)$ 融入确定性策略输出:

$$a^n = \phi(s^n) + \epsilon \quad (26)$$

智能体执行动作后获得经验 (s^n, a^n, r^n, s^{n+1}) 并存储于回放缓冲池 \mathcal{B} , 构建算法进化的记忆库。在 Critic 网络更新时, 采用双重评估的最小化选择策略来计算目标 Q 值, 并引入截断高斯噪声 $\tilde{\delta} \sim \text{clip}(N(0, \tilde{\sigma}), -c, c)$ 实施策略平滑:

$$y = r + \gamma \min_{i=1,2} Q_{\zeta_i'}(s^{n+1}, \tilde{a}^n), \tilde{a}^n = \phi(s^{n+1}) + \tilde{\delta} \quad (27)$$

其中, $\gamma \in [0, 1)$ 表示折扣因子。

Critic 网络参数通过最小化双重均方误差实现参数更新:

$$L(\zeta_i) = \mathbb{E}_{(s^n, a, r, s^{n+1}) \sim \mathcal{D}} [(y - Q_{\zeta_i}(s, a))^2], \quad i=1, 2 \quad (28)$$

这种双重评估并取最小值的机制, 在数学上构建了 Q 值的下界约束, 显著抑制了单网络过拟合导致的高估风险。

Actor 网络采用延迟更新策略, 以 Critic 网络更新周期 σ 倍 (如 $\sigma=2$) 的节奏进行梯度迭代。这一延迟更新机制确保了 Critic 网络能够优先收敛至稳定状态, 避免早期策略波动对价值评估的干扰。策略梯度由第一个 Critic 网络 Q_{ζ_1} 提供:

$$\nabla_{\phi} J(\phi) \approx E_{s \sim \mathcal{D}} [\nabla_a Q_{\zeta_1}(s, a)|_{a=\phi_{\theta}(s)} \nabla_{\phi} \phi(s)] \quad (29)$$

时序异步设计确保 Critic 网络优先收敛至稳定状态, 避免策略突变对价值函数的干扰。目标网络则通过软更新系数 $\tau \in (0, 1)$ 实现参数渐进同步:

$$\phi' \leftarrow \tau \phi + (1 - \tau) \phi' \quad (30)$$

$$\zeta_i' \leftarrow \tau \zeta_i + (1 - \tau) \zeta_i', \quad i=1, 2 \quad (31)$$

整体的优化算法如算法 2 所示。

算法 2 基于 TD3 的 UAV 轨迹优化算法

输入: RIS 单元的分配策略 $\Gamma_{l,K-M}$, TD3 算法所需的超参数



输出：功率分配、RIS相位和UAV轨迹

- (1) 初始化主网络 ϕ 、 Q_{ζ_1} 、 Q_{ζ_2} 和目标网络 ϕ' 、 $Q_{\zeta_1'}$ 、 $Q_{\zeta_2'}$ ，其中目标网络参数遵循式 (24)；
- (2) 初始化经验回放缓冲池 \mathcal{B}
- (3) for 训练回合数 in Epoch_{max}
- (4) 重置环境得到状态 s_1
- (5) for step in Step_{max}
- (6) 根据式 (25) 选择动作 a^n
- (7) 执行动作 a^n ，观测奖励 r^n 和下一时刻状态 s^{n+1}
- (8) 存储 (s^n, a^n, r^n, s^{n+1}) 到经验回放缓冲池中
- (9) 更新状态 $s^n \leftarrow s^{n+1}$
- (10) if $|\mathcal{B}| \geq \mathcal{D}$ then
- (11) 从 \mathcal{B} 采样批量 D 的数据
- (12) 通过式 (26) 和式 (27) 计算损失，并更新 Critic 网络
- (13) if 训练回合数 mod $d = 0$ then
- (14) 通过式 (28) 更新 Actor 网络
- (15) 通过式 (29) 和式 (30) 软更新目标神经网络
- (16) end if
- (17) end if
- (18) end for
- (19) end for

3.3 算法复杂度分析

联盟博弈阶段，复杂度主要取决于迭代次数 I_{\max} 和切换操作的遍历。单次迭代复杂度为 $O(M \cdot K)$ ，总复杂度为 $O(I_{\max} \cdot M \cdot K)$ ，属于多项式复杂度，在实际规模的网络中可以高效求解。

TD3 算法阶段，训练过程的复杂度主要取决于神经网络规模。设 Actor 和 Critic 网络均为全连接层，每层输入与输出维度分别为 D_{in} 和 D_{out} ，层数为 H ，则单步前向传播复杂度约为 $O(\sum_{l=1}^H n_l \cdot$

$n_{l-1})$ 。考虑到 TD3 采用双重 Critic 网络及目标网络，且每步训练需进行多次前向与反向传播，实际每步训练复杂度约为该值的常数倍，仍保持在可接受的范围内。

4 性能评估

4.1 参数设置

为验证本文所提算法的性能，建模了一个 RIS 辅助的 UAV 通信系统，其中 UAV 从 $H_u^s = [75, 0]$ 飞往 $H_u^f = [75, 150]$ ，高度固定为 $Z_u = 40 \text{ m}$ ，最大飞行速度 $v_u^{\max} = 20 \text{ m/s}$ ，RIS 的坐标为 $[25, 50]$ 与 $[25, 100]$ ，高度统一固定为 $Z_k = 25 \text{ m}$ ，用户的数量 $I = 6$ ，在规定的范围内随机分布。其他参数设置见表 1。

表 1 仿真参数设置

参数	含义	设定值
λ	时隙长度	1 s
ν	载波波长	0.125 m
$\epsilon_{u,i}, \epsilon_{u,k}, \epsilon_{k,i}$	路径损耗指数	2.5、2.0、2.0
φ	参考距离增益	-30 dBm
P_{\max}	最大发射功率	1 W
$\mu_{u,k-m}, \mu_{k,m-i}$	Rician 系数	10 dB
P_0	悬停功耗参数	79.86
U_{tip}	旋翼尖端速度	120 m/s
d_0	机身阻力比	0.6
s	旋翼实度	0.05
\mathcal{B}	经验回放缓冲区	100 000
τ	软更新频率	0.001
M	RIS 单元数	40
R_{\min}	最小信道容量	3 Mb/s
d_k	单元间隔	$\nu/2$
σ^2	信道噪声功率	-60 dBm
B	带宽	10 MHz
P_1	飞行功耗参数	88.63
ν_0	平均旋翼速度	4.03 m/s
ρ	空气密度	1.225 kg/m ³
Δ	旋翼盘面积	0.503 m ²
\mathcal{D}	批量大小	128
$\zeta_{\text{Actor}}, \zeta_{\text{Critic}}$	学习率	5e-4、1e-4

4.2 性能对比

不同深度强化学习算法（DDPG、TD3、SAC）的性能对比如图3所示。其中，TD3算法在稳定性和收敛效率方面展现出显著优势。作为对比的基准算法DDPG和SAC（soft Actor-Critic）虽然最终实现了收敛，但其训练过程存在明显缺陷：DDPG算法的奖励曲线呈现剧烈震荡，最大波动幅度超过40%；SAC算法虽相对平稳，但在训练至1500轮后仍出现较大波动。相比之下，TD3算法在训练初期（100~300轮）即表现出优异的收敛特性，这种稳定性优势在收敛后期（1500轮后）更为显著，奖励值波动范围被有效控制在了±2.5%以内。实验数据表明，TD3算法通过双重Critic网络架构和延迟更新策略，显著缓解了价值高估与策略震荡问题，从而实现了更优的收敛性能。当算法均收敛至稳定状态后，TD3算法所能达到的平均系统能效约为 6.2×10^4 bits/J，而SAC算法的平均能效约为 5.2×10^4 bits/J。相较于SAC基线算法，本文方案的系统能效提升了约19%。

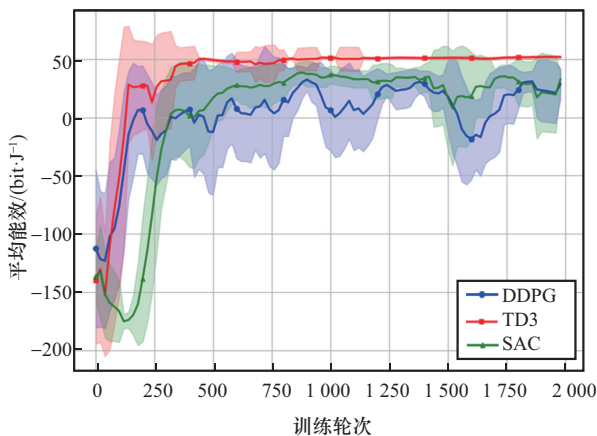
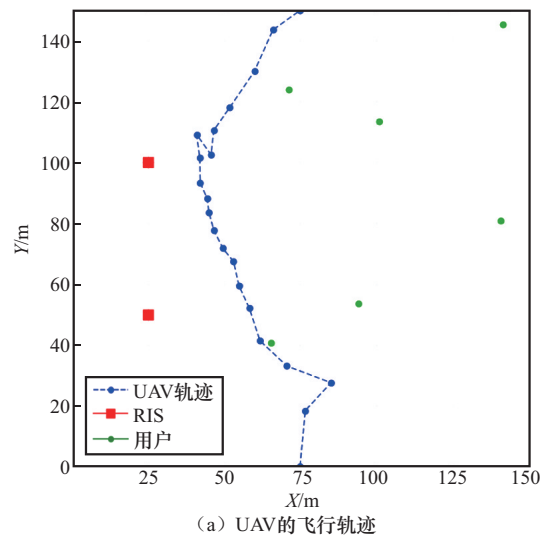


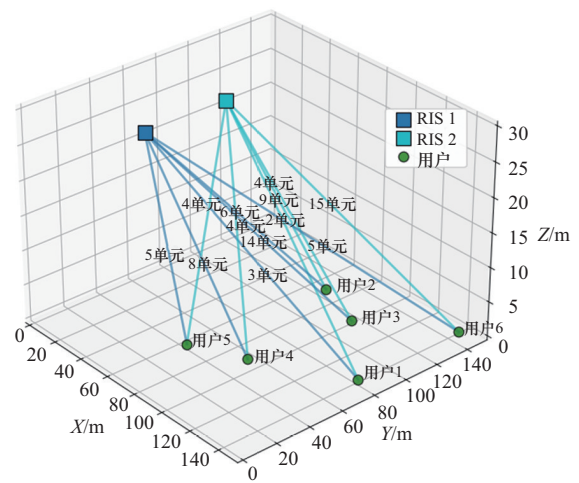
图3 不同深度强化学习算法的性能对比

UAV的飞行轨迹和RIS的配对结果如图4所示。从图4(a)可以看出，UAV的飞行轨迹呈现出向RIS靠近的趋势，这是因为当部署RIS辅助通信时，UAV需要在直接链接和RIS反射链路间做出

妥协。利用本文所提算法调整RIS的相移，可以实现反射信号的相干叠加，从而增强用户的接收信号强度，并扩展UAV的服务覆盖范围。从图4(b)可以看出，两个RIS分别与3个用户关联。考虑到RIS可关联的最大用户数量，即多用户复用RIS可能带来的干扰，本文所提算法能够有效提高RIS资源分配的均衡性。



(a) UAV的飞行轨迹



(b) RIS的配对结果

图4 UAV的飞行轨迹和RIS的配对结果

为验证所提算法在“以用户为中心”方面的表现，统计了在一个完整飞行任务周期内，网络中所有6个用户的平均传输速率分布，如图5所示。结果显示，尽管用户随机分布且信道条件各



异，基于TD3算法与联盟博弈的联合优化通过动态调整RIS相位与关联策略，确保了所有用户的平均速率均显著高于预设的QoS阈值，有效证明了算法的有效性。

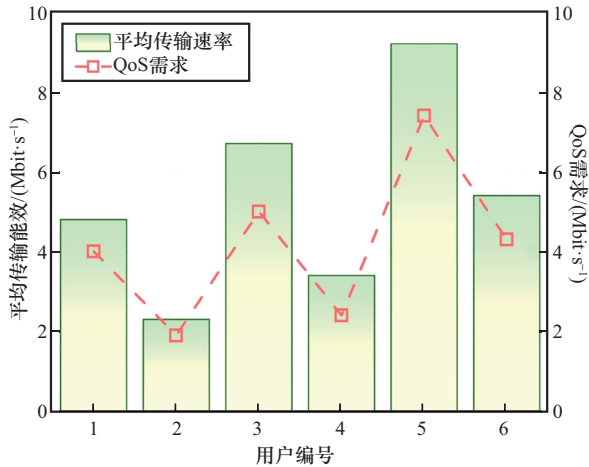


图5 各用户平均传输速率分布

在后续对比评估中，本文所提方案将与以下5种基准方案进行比较：(1) 无RIS卸载方案，即环境中不部署RIS，仅依靠UAV直接传输；(2) 随机方案，即随机生成RIS的单元分配和相位，UAV的轨迹和功率分配依旧使用TD3算法；(3) DDPG方案，即使用DDPG算法代替TD3算法来优化RIS相位、UAV轨迹和传输功率；(4) SAC方案，即使用SAC算法代替TD3算法；(5) GA方案，即用遗传算法来代替TD3算法。

总传输数据与能效之间的关系如图6所示。实验数据表明，在部署了RIS后，通过联合优化传输功率、RIS单元分配、相移和无人机轨迹，本文提出的联合优化方案可以实现比其他基准方案更高的能效，相比于无RIS的基准方案平均提升了72%。值得注意的是，采用随机相位的RIS方案虽然同样部署了RIS，但其能效较基准方案仅提升了1.7%，所带来的增益微乎其微，这一结果表明了相移优化在RIS辅助无人机网络中的重要性。

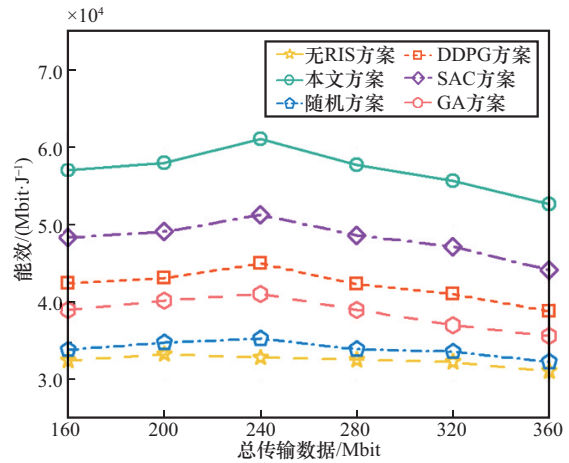


图6 总传输数据与能效之间的关系

最大传输功率与能效之间的关系如图7所示。实验数据表明，当最大传输功率从10 dBm提升至40 dBm时，本文方案的能效从 5.2×10^4 bits/J增至 6.4×10^4 bits/J，且表现出先急后缓的趋势。这种现象源于香农容量公式的边际递减效应，和复用RIS带来的干扰因素。研究表明，在RIS辅助系统中，单纯提高发射功率无法突破容量干扰的物理极限，必须通过跨层联合优化实现功率资源的最优配置。

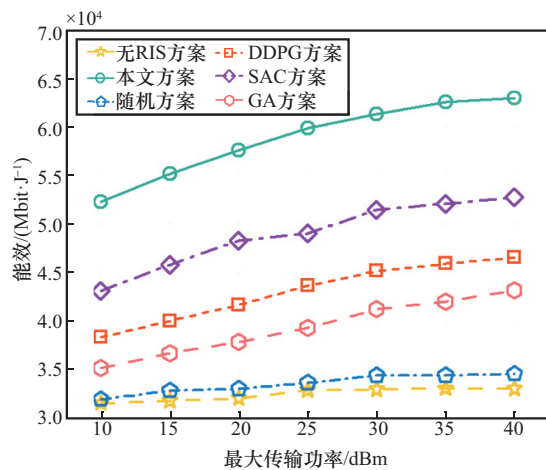


图7 最大传输功率与能效之间的关系

RIS单元数量与系统能效之间的关系如图8所示。实验数据表明，当RIS单元数从20增至70时，除了无RIS基准方案外，其余方案的能效均随RIS

单元数量的增加而提升。这表明,增加RIS反射单元能为信道提供更高的增益,特别是在结合高效相移策略的情况下,且受益于RIS的无源特性,其额外单元的引入不会产生更多的能耗。值得注意的是,RIS单元的增加会扩大动作空间的维度,容易引发“维数灾难”,从而显著加剧智能体的探索难度。因此,在实际应用中需综合考虑增益与复杂度,将RIS单元数量控制在一个适合的区间。

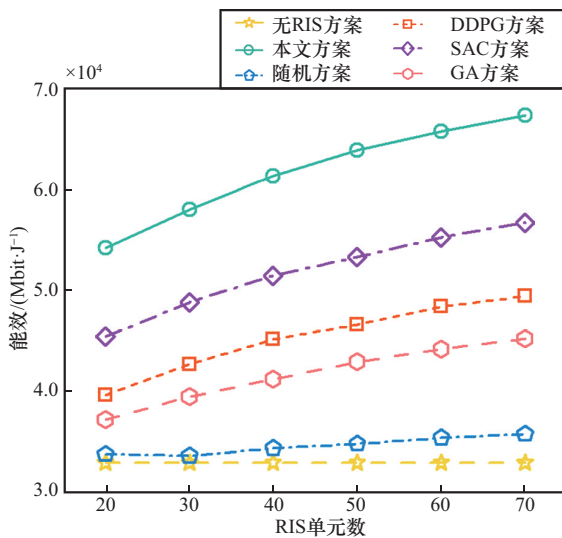


图8 RIS单元数量与系统能效之间的关系

5 总结

针对多RIS辅助的UAV通信网络,本文提出了一种以用户为中心的协作传输机制及多阶段联合优化框架。该机制突破传统网络中心化设计的局限,允许用户根据实时信道条件与业务需求动态关联最优RIS节点,显著提升了差异化QoS保障能力。为协调信道容量和能耗之间的关系,本文构建了以两者比值为基础的系统能效目标函数,并设计了一种联合联盟博弈与TD3算法的多阶段优化框架,对传输功率分配、RIS单元分配、相位调控以及UAV轨迹规划进行联合优化,以实现系统能效的最大化。实验结果表明,所提方案在能效优化方面均优于现有基准方案,尤其

是在高密度RIS部署场景下,为未来空天地一体化网络部署提供了重要参考。未来工作将围绕多无人机协同调度、非理想信道状态信息下的鲁棒性优化、通信与边缘计算联合设计,以及三维轨迹优化与多RIS波束赋形联合设计等方向展开,以应对更加复杂的城市通信场景,进一步推动智能反射面技术在6G网络中的实际部署与应用。

参考文献:

- [1] Feng W, Wang Y M, Chen Y F, et al. Structured satellite-UAV-terrestrial networks for 6G Internet of Things[J]. IEEE Network, 2024, 38(4): 48-54.
- [2] Xia X Y, Fattah S M M, Babar M A. A survey on UAV-enabled edge computing: resource management perspective[J]. ACM Computing Surveys, 2024, 56(3): 78.
- [3] Dai Z J, Zhang Y, Zhang W C, et al. A multi-agent collaborative environment learning method for UAV deployment and resource allocation[J]. IEEE Transactions on Signal and Information Processing over Networks, 2022, 8: 120-130.
- [4] Zhang C M, Xu C Y, He B. A distributed task allocation approach for multi-UAV persistent monitoring in dynamic environments[J]. Scientific Reports, 2025, 15(1): 6437.
- [5] Guo H Y, Liang Y C, Chen J, et al. Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks[J]. IEEE Transactions on Wireless Communications, 2020, 19(5): 3064-3076.
- [6] Wu Q Q, Zhang R. Towards smart and reconfigurable environment: intelligent reflecting surface aided wireless network[J]. IEEE Communications Magazine, 2020, 58(1): 106-112.
- [7] Pan C H, Ren H, Wang K Z, et al. Reconfigurable intelligent surfaces for 6G systems: principles, applications, and research directions[J]. IEEE Communications Magazine, 2021, 59(6): 14-20.
- [8] Peng Z J, Liu R J, Pan C H, et al. Energy minimization for active RIS-aided UAV-enabled SWIPT systems[J]. IEEE Communications Letters, 2024, 28(6): 1372-1376.
- [9] Peng Z, et al. Energy minimization for active RIS-Aided UAV-enabled sWIPT systems[J]. IEEE communications letters, 2024(6):28.
- [10] Jiao L, Gao L, Zheng J, et al. Resource allocation in RISs-assisted UAV-enabled MEC network with computation capacity improvement[J]. Computer Communications, 2024, 228: 107953.



- [11] Elsherbini M M, Omer O A, Salah M. Low-complexity cooperative active and passive beamforming multi-RIS-assisted communication networks[J]. EURASIP Journal on Wireless Communications and Networking, 2024, 2024(1): 65.
- [12] Adesokan A, Halabi A E, Houjaj F. RIS-assisted coverage maximization using multi-UAVs in LTE networks[J]. International Journal of Wireless and Microwave Technologies, 2023, 13(4): 28-34.
- [13] Yao Y, et al. UAV trajectory and energy efficiency optimization in RIS-Assisted Multi-User Air-to-Ground communications networks[J]. Drones, 2023, 7(4): 272.
- [14] Gao Z, Sun S, Su X, et al. Multi-RIS aided multicell wireless networks: joint beamforming design combined with selection strategy[J]. IEEE Internet of Things Journal, 2025, 12(14): 27943-27956.
- [15] Mei H B, Yang k, Liu Q, et al. 3D-trajectory and phase-shift design for RIS-assisted UAV systems using deep reinforcement learning[J]. IEEE Transactions on Vehicular Technology, 2022, 71(3): 3020-3029.
- [16] Li Y B, Zhang H J, Long K P, et al. Exploring sum rate maximization in UAV-based multi-IRS networks: IRS association, UAV altitude, and phase shift design[J]. IEEE Transactions on Communications, 2022, 70(11): 7764-7774.
- [17] Ruan L, Li G X, Cheng J, et al. Multistage clustering-based localization for remote UAV swarm: a coalitional game framework[J]. IEEE Communications Letters, 2022, 26(9): 2047-2051.

- [18] Zheng C X, Pan K, Dong J D, et al. Multi-agent collaborative optimization of UAV trajectory and latency-aware DAG task offloading in UAV-assisted MEC[J]. IEEE Access, 2024, 12: 42521-42534.

[作者简介]



冯艳如（1990-），女，山西应用科技学院讲师，主要研究方向为无线通信 UAV 网络、信息传输及信息安全。



闫俊杰（1990-），男，博士，广西科技大学电子工程学院副教授，主要研究方向为 MEC、UAV 通信、D2D 通信。



刘汉永（1992-），女，硕士，国家电网重庆市电力公司大足供电分公司工程师，主要研究方向为区块链、网络安全、密码学、自动化、电网营销。