

基于 DDPG 的 RIS-UAV 辅助 MEC 系统波束赋形算法

高 航¹
GAO Hang

摘 要

针对现有的无人机 (UAV) 与智能反射面 (RIS) 辅助移动边缘计算 (MEC) 系统的波束赋形方案中的能量效率 (EE) 较低的问题, 文章提出了一种基于深度确定性策略梯度 (DDPG) 的 RIS-UAV 辅助 MEC 系统波束赋形算法。该算法首先利用全连通神经网络 (FCNN) 实现的 Actor 网络和 Critic 网络; 其次使用加性噪声进行训练, 通过提供更快更平滑的收敛来提高 DDPG 的效果; 最后针对模型收敛速度慢的问题, 设计数据驱动经验回放机制 (PER), 提升 DDPG 算法在负载调度场景的收敛速度。仿真实验结果表明, 所提出的算法, 优于传统 CCCP 算法, 具备更快收敛速度和更好稳定性, 并能提高系统的 EE。

关键词

无人机; 智能反射面; 移动边缘计算; 波束赋形; 深度确定性策略梯度算法

doi: 10.3969/j.issn.1672-9528.2025.08.019

0 引言

随着移动互联网和物联网的迅速发展, 传统云计算模式因计算延迟高、网络带宽瓶颈和数据隐私问题, 无法满足低延迟、高可靠性和用户体验的需求。移动边缘计算 (MEC) 通过将计算和存储功能移至网络边缘, 显著降低计算延迟, 提高实时性和可靠性。

无人机 (UAV) 可以灵活部署 MEC 服务器, 为偏远地区提供可靠计算服务。可重构智能表面 (RIS) 技术通过调整单元相移优化信号传播特性, 进一步提高 MEC 系统性能。

无人机凭借其高移动性, 打破了传统无线网络的不灵活特性。然而, UAV-MEC 网络在高机动性情况下必须应对复杂且时变的衰落信道, 尤其在城市环境中, 这对无线传输产生不利影响。因此, 开发新技术重新配置无线信道以提升 UAV-MEC 网络的计算能力是必不可少的。

在这种背景下, 移动边缘计算 (MEC) 似乎通过增强网络边缘的计算能力来应对这些问题, 使得 UE 能够通过任务卸载来极大地缓解其计算压力, 然后以高可靠性和低延迟接收结果。同样, Sun 等人^[1]在工业物联网场景中, 将人工智能方法与无人机-MEC 网络相结合。尽管 UAV-MEC 网络在各种通信应用中得到了广泛的研究, 但它仍然面临着许多挑战或问题。因此, 开发一种新技术来重新配置无线信道是势在必行的。

可重构智能表面 (RIS), 也称为智能反射表面 (IRS), 其有望设计无线电传播环境, 受到了学术界和工业界的广泛

研究^[2]。

RIS 已与多种通信技术相结合, 如 NOMA^[3]、同时无线信息和功率传输 (SWIPT)^[4]、双向中继^[5]和物理层安全^[6]。此外, 无人机与 RIS 相遇属于一种新模式^[7], Li 等人^[8]和 Fang 等人^[9]研究了 RIS 设计改进的通信安全问题。Liu 等人^[10]的工作旨在借助基于机器学习方法的 RIS 来提高无人机的服务质量, 其中同时考虑了无人机运动和 RIS 相移。Wei 等人^[11]通过仔细设计 RIS, 使正交频分多址 (OFDMA) 系统中的和速率最大化。在最坏情况信噪比 (SNR) 最大化的目标下, 工作^[12]揭示了航空 RIS 和传统地面 RIS 之间的性能差异。接下来, Khalili 等人^[13]和 Pan 等人^[14]分别关注 RIS 辅助异构通信网络和 RIS 辅助太赫兹通信。Hua 等人^[15]提出了一种具有多个 RIS 的共生无人机-RIS 通信系统, 每个 RIS 在无人机的帮助下将自己的数据发送到基站, 同时辅助无人机的数据传输。

然而, 在实际问题中, 当地面用户卸载任务量很有可能超过 UAV 所能承担的负荷。基于这些观点, 针对无人机和 RIS 联合设计的 MEC 网络系统, 本文提出采用一种面向系统的 EE 的优化方案。本文的主要贡献总结如下:

本文提出了一种使用深度确定性策略梯度 (DDPG) 网络来解决无人机机载物联网平台中的 RIS 辅助干扰抑制问题的方法。该方法利用全连通神经网络 (FCNN) 实现的 Actor 网络和 Critic 网络, 使用加性噪声进行训练, 并通过提供更快更平滑的收敛来提高 DDPG 的效果, 尤其有利于更大的 RIS。本文还提出了两个基线来进一步验证所提出的方法的优点: 无 RIS 和 RIS 配置。

1. 沈阳航空航天大学电子信息工程学院 辽宁沈阳 110000

1 系统模型

1.1 整体架构

本文考虑在 MEC 网络中使用无人机和反射表面 (RIS) 进行中继传输的问题。具体而言, 本文考虑一个 MEC 网络, 其中无人机 (UAV) 配备 $L \geq 2$ 个天线, RIS 附着在建筑表面, 由 RIS 控制器操控。UAV 充当服务器以执行计算; 同时还充当中继, 将计算任务从一组单天线 AP (发射机 S) 传递到配备 MEC 服务器的远程地面单天线 UE (接收机 D), 在此期间, 带有 $M_y \times M_z$ 反射元素的均匀矩形阵列 (URA) RIS 协助从 UAV 到 AP 的中继传输过程, M_y 和 M_z 分别表示沿 y 轴和 z 轴的元素。具体结构如图 1 所示。

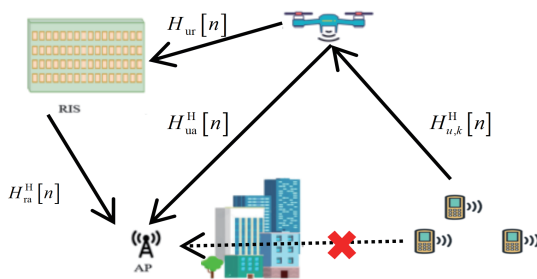


图 1 MEC 网络中使用无人机和反射表面 (RIS) 进行中继传输的示意图

本文假设 UE-RIS 通信链路和 UE-AP 通信链路由于距离长而被障碍物严重阻塞。在这种情况下, 传统的无线通信技术可能无法满足网络性能需求。因此, 本文考虑利用 RIS 中的反射来增强信号强度和质量, 从而提高网络性能。为了优化网络性能, 本文将网络周期时间 T 分成 N 个相等的时间隙, 每个时间隙的大小为 δ_t 。本文将时间隙集表示为 $\mathcal{N}$, 其中 $|\mathcal{N}| = N$, UE 集表示为 $\mathcal{K}$, 其中 $|\mathcal{K}| = K$, $|\cdot|$ 表示集的基数。 $\boldsymbol{\theta}_k[n] = \text{diag}(e^{j\theta_{k,1}[n]}, e^{j\theta_{k,2}[n]}, \dots, e^{j\theta_{k,M}[n]})$ 表示第 K 个 UE 在时间隙 n 中的反射系数矩阵, 相移满足 $0 \leq \theta_{k,m}[n] < 2\pi$ 。假设每个元素的反射幅度等于 1, 在三维笛卡尔坐标系中, 第 K 个 UE 的位置为 $(w_k, 0)$, 其中 $w_k = (x_k, y_k)$ 。因此, 本文获得 UAV 的位置 $(q[n], h_u)$, AP 的位置 $(w_a, 0)$ 和 RIS 的参考位置 (w_r, h_r) , 其中 $q[n] = (x_u[n], y_u[n])$, $w_a = (x_a, y_a)$, $w_r = (x_r, y_r)$, 表示水平位置; h_u 表示 UAV 的高度; h_r 表示 RIS 的高度。因高路径损耗, 本文忽略了 RIS 多次反应的信号功率。假设所有信道都经历了准静态平坦衰落, 且所有信道的信道状态信息 (CSI) 都是完全已知的。在这个设置中, 本文希望通过调整反射系数矩阵 $\mathbf{k}[n]$ 来最大化网络性能。

1.2 UAV 运动模型

UAV 的有限调度时间为 $N\delta_t$ s, 分为以 n 为索引的 N 个时间隙, $n = 1, 2, \dots, N$ 。一个时间隙持续 δ_t , 此时较短, 因此可以将 UAV 视为每个时间隙静止。UAV 从预定的起始位置 $q[0] = [x_u[0], y_u[0]]^T$ 飞行到预定的终止位置 $q[N] = [x_u[N], y_u[N]]^T$ 。

每个时间隙 T 的 UAV 速度和加速度分别用 $v[n] = [v_x[n], v_y[n]]^T$ 和 $A[n] = [A_x[n], A_y[n]]^T$ 表示。UAV 遵循一些机动性约束:

$$q[n+1] = q[n] + v[n]\delta_t + \frac{1}{2}A[n]\delta_t^2, \forall n \quad (1)$$

$$q[N] = q_F \quad (2)$$

$$v[n+1] = v[n] + A[n]\delta_t, \forall n \quad (3)$$

$$\|q[n] - q[n-1]\| \leq \delta_t V_{\max}, n = 2, 3, \dots, N \quad (4)$$

式中: $V_{\max}$ 是 UAV 的最大速度。

RIS 黏附在建筑物的外表面上, 该建筑物位于 xz 平面上, 并与 x 轴对齐。RIS 是一种反射能量的智能表面, 由 $M = M_y M_z$ 个反射单元组成, 这些反射单元组成了一个均匀矩形阵列 (URA), 同时还有用于控制每个单元相移的控制单元。RIS 的作用是通过改变反射单元的相位来控制入射信号的幅度和方向。每个时间隙 n , RIS 上的相移矩阵 $\boldsymbol{\theta}_k[n] = \text{diag}(e^{j\theta_{k,1}[n]}, e^{j\theta_{k,2}[n]}, \dots, e^{j\theta_{k,M}[n]})$, 相移满足 $0 \leq \theta_{k,m}[n] < 2\pi$ 。相移矩阵 $\boldsymbol{\theta}_k[n]$ 表示 UE k 在时间隙 n 中的反射系数矩阵, 用于调整反射单元的相位。其中, $\theta_{k,m}$ 表示位于 RIS 的第 M_y 行和第 M_z 列的第 m 个反射单元的相位。RIS 的右下角是 RIS 的第一个元素, 其坐标为 $q_R = [x_R, y_R, z_R]^T$ 。意味着 RIS 的第一个反射单元位于坐标 $[x_R, y_R, z_R]$ 。通过控制每个反射单元的相位, RIS 可以实现对入射信号的精确控制和调整。

1.3 信号传输与接收模型

本小节主要包含信号传输和信号接收的信道增益模型。

假设 LoS 通道主导无人机-RIS 链路。UE k 和 UAV 之间的信道矩阵为 $H_{u,k}^H[n] \in \mathbb{C}^{L \times 1}$; UAV 和 RIS 之间的信道矩阵为 $H_{ur}^H[n] \in \mathbb{C}^{M \times L}$; RIS 和 AP 之间的信道矩阵为 $H_{ru}^H[n] \in \mathbb{C}^{1 \times M}$; UAV 和 AP 之间的信道矩阵为 $H_{ua}^H[n] \in \mathbb{C}^{1 \times L}$ 。本文考虑 RIS-UAV 链路 (RU 链路) 的 LoS 信道, UE-UAV 链路的 Rician 衰落信道。

RU 链路每个时间隙中的信道为:

$$H_{ur}[n] = \frac{\beta_0}{\sqrt{\|q[n] - w_r\|^2 + (h_u - h_r)^2}} \bar{H}_{ur}^{\text{LoS}}[n] \quad (5)$$

式中: β_0 表示参考距离为 1 m 时的平均信道功率增益, 接收阵列响应 $\bar{H}_{ur}^{\text{LoS}}[n] = A_{R,r}[n] A_{T,u}^H[n]$ 。其中, $A_{T,u}^H[n]$ 示与无人机到 RIS 的偏离角 (AoD) 有关的阵列响应:

$$A_{T,u}^H[n] = \left[1, e^{-j\frac{2\pi}{\lambda} d \phi_{ur}[n]}, \dots, e^{-j\frac{2\pi}{\lambda} (L-1) d \phi_{ur}[n]} \right]^T \quad (6)$$

式中: $\phi_{ur} = \frac{[x_r - x_u[n]]}{\sqrt{\|q[n] - w_r\|^2 + (h_u - h_r)^2}}$ 表示从 UAV 到 RIS 的信号的 AoD 的余弦; d 表示天线间距; λ 表示载波波长。

$A_{R,r}[n]$ 表示 RIS 处的接收阵列响应:

$$A_{R,r}[n] = \left[1, e^{-j\frac{2\pi}{\lambda} d_r \phi_{ur}[n]}, \dots, e^{-j\frac{2\pi}{\lambda} (M_y-1) d_r \phi_{ur}[n]} \right]^T \otimes \left[1, e^{-j\frac{2\pi}{\lambda} d_z \phi_{ur}[n]}, \dots, e^{-j\frac{2\pi}{\lambda} (M_z-1) d_z \phi_{ur}[n]} \right]^T \quad (7)$$

$$\sin\theta_{ur}[n] = \frac{[h_u[n] - h_r]}{\sqrt{\|q[n] - w_r\|^2 + (h_u - h_r)^2}} \quad (8)$$

$$\sin\varphi_{ur}[n] = \frac{[x_r - x_u[n]]}{\sqrt{(x_r - x_u[n])^2 + (y_r - y_u[n])^2}} \quad (9)$$

$$\cos\varphi_{ur}[n] = \frac{[y_r - y_u[n]]}{\sqrt{(x_r - x_u[n])^2 + (y_r - y_u[n])^2}} \quad (10)$$

假设 UE-UAV 链路、AU 链路和 AR 链路信道由 Rician 信道链路主导。因此,

$$h_i[n] = \sqrt{\beta_0 d_i^{-\alpha_i}} \tilde{h}_i[n] \quad (11)$$

式中: β_0 是参考距离为 1 m 时的平均信道功率增益; $i \in \{u, k\}, \{ua\}, \{ra\}$ 是由 UE-UAV 链路、UAV-AP 链路和 RIS-AP 链路组成的索引集; $\alpha_i \in (2,4)$ 是路径损耗指数值。 $d_i[n]$ 是每个通道的距离。

$$d_{u,k}[n] = \sqrt{\|q[n] - w_k\|^2 + h_u^2} \quad (12)$$

$$d_{ua}[n] = \sqrt{\|q[n] - w_a\|^2 + h_u^2} \quad (13)$$

$$d_{ra}[n] = \sqrt{\|w_r - w_a\|^2 + h_r^2} \quad (14)$$

式中: $\tilde{h}_i[n]$ 代表 Rician fading。

$$\tilde{h}_i[n] = \sqrt{\frac{\beta_i}{1 + \beta_i}} \tilde{h}_i^{\text{LoS}}[n] + \sqrt{\frac{1}{1 + \beta_i}} \hat{h}_i^{\text{NLoS}}[n] \quad (15)$$

式中: β_i 代表 Rician 因子。 $\tilde{h}_i^{\text{LoS}}[n]$ 代表 LoS 元素。

$$\tilde{h}_{u,k}^{\text{LoS}}[n] = \left[1, e^{-j\frac{2\pi}{\lambda} d_{\phi_{u,k}}[n]}, \dots, e^{-j\frac{2\pi}{\lambda} (L-1) d_{\phi_{u,k}}[n]} \right]^T \quad (16)$$

$$\tilde{h}_{ua}^{\text{LoS}}[n] = \left[1, \int e^{-j\frac{2\pi}{\lambda} d_{\phi_{ua}}[n]}, \dots, e^{-j\frac{2\pi}{\lambda} (L-1) d_{\phi_{ua}}[n]} \right]^T \quad (17)$$

$$\tilde{h}_{ra}^{\text{LoS}}[n] = \left[1, e^{-j\frac{2\pi}{\lambda} d_{\phi_{ra}}[n]}, \dots, e^{-j\frac{2\pi}{\lambda} (M_r-1) d_{\phi_{ra}}[n]} \right]^T \otimes \left[1, e^{-j\frac{2\pi}{\lambda} d_{\phi_{ra}}[n]}, \dots, e^{-j\frac{2\pi}{\lambda} (M_r-1) d_{\phi_{ra}}[n]} \right]^T \quad (18)$$

式中: $\phi_{u,k}[n] = \frac{[x_u[n] - x_k]}{\sqrt{\|q[n] - w_r\|^2 + (h_u - h_r)^2}}$ 表示从 UE k 到 UAV

的信号 AoA 的余弦; $\phi_{ua}[n] = \frac{[x_a - x_u[n]]}{\sqrt{\|q[n] - w_a\|^2 + h_u^2}}$ 表示从

UAV 到 AP 的信号 AoD 的余弦; $\phi_{ra}[n] = \sin\theta_{ra}[n]\cos\varphi_{ra}[n]$, $\Omega_{ra}[n] = \sin\theta_{ra}[n]\sin\varphi_{ra}[n]$, $\theta_{ra}[n]$ 和 $\varphi_{ra}[n]$ 表示从 RIS 到 AP 的信号垂直和水平到达角 (AoD)。

$$\sin\theta_{ra}[n] = \frac{[h_r]}{\sqrt{\|w_a - w_r\|^2 + h_r^2}} \quad (19)$$

$$\sin\varphi_{ra}[n] = \frac{[x_a - x_r]}{\sqrt{(x_a - x_r)^2 + (y_a - y_r)^2}} \quad (20)$$

$$\cos\varphi_{ra}[n] = \frac{[y_a - y_r]}{\sqrt{(x_a - x_r)^2 + (y_a - y_r)^2}} \quad (21)$$

1.4 计算卸载模型

为了协调边缘计算的过程, 本文使用了一个专用的时隙划分协议, 将每个时隙划分为总共 $2K$ 个子时隙。前面的 K 个子时隙用于将 UE 的任务依次卸载给 UAV, 每个大小为 $\tau_k^O[n]$; 后面的 K 个子时隙用于 UAV 将 UE 接收到的任务的一部分传递给 AP, 每个大小为 $\tau_k^R[n]$ 。每个时隙的处理延迟要小于等于 δ_i , 即 $\sum_{k=1}^K \tau_k^O[n] + \sum_{k=1}^K \tau_k^R[n] \leq \delta_i$ 。其中, UAV 接收到的 UE k 任务分为两部分, 一部分用于 UAV 计算, 另一部分转发给 AP 执行。UAV 充当解码转发 (DF) 中继, 解码和重新编码的处理延迟是一个时隙, 即 δ_i , 因此 UAV 当前时隙中接收到的任务可以在本地计算和/或在下一个时隙交付。第一个时隙仅用于将任务从 UE 卸载到 UAV, 最后一个时隙仅用于 UAV 和 AP 的任务计算, 即 $\tau_k^R[1] = 0$, $\tau_k^R[N] = 0$, $\tau_k^O[N] = 0$ 。本文用 s_k 表示在 UE k 处具有零均值和单位方差的信息承载信号, p_i 表示 UE 的传输功率, $n \in \mathbb{C}^{L \times 1}$ 表示满足 $n \sim \mathcal{CN}(0, \sigma^2 \mathbb{I}_L)$ 的 UAV 处的信号, 其中 σ^2 是高斯白噪声功率。因此, 从 UE k 到 UAV 的接收信号 $y_k[n] \in \mathbb{C}^{L \times 1}$ 的计算公式为 $y_k[n] = \sqrt{p_i} h_{u,k}[n] s_k + n$ 。

本文用 $u_k^H[n] \in \mathbb{C}^{L \times 1}$ 表示检测向量, 也称为接收波束形成向量, 以检测和恢复 UE k 的信号。因此, UE k 的恢复信号为 $\hat{s}_k = u_k^H[n] y_k[n]$ 。因此, UAV 上 UE k 的接收信噪比为:

$$\gamma_{u,k}[n] = \frac{p_i |u_k^H[n] H_{u,k}[n]|^2}{\sigma^2 \|u_k^H[n]\|^2} \quad (22)$$

在从 UAV 到 AP 的信号传输过程中, 本文考虑 UAV 到 AP 的信号传输过程中 3 个组件的级联, 包括 UAV-AP 链路、UAV-RIS 链路和 RIS-AP 链路。本文推导出 UAV 到 AP 的 UE k 信号为 $Y_k[n] = (H_{ua}^H[n] + H_{ra}^H[n] \Theta_k[n] H_{ur}[n]) w_k[n] \tilde{s}_k + n$ 。其中, $\tilde{s}_k$ 表示具有零均值和单位方差的信息承载信号; $w_k[n] \in \mathbb{C}^{L \times 1}$ 表示相应的波束形成向量; $n \sim \mathcal{CN}(0, \sigma^2)$ 表示 AP 处的噪声。

UE k 在 AP 处的接收 SNR 的计算公式为

$$\gamma_{a,k}[n] = \frac{|(H_{ua}^H[n] + H_{ra}^H[n] \Theta_k[n] H_{ur}[n]) w_k[n]|^2}{\sigma^2 + \|H_{ua}[n] + u_k^H[n] H_{ra}[n]\|^2} \quad (23)$$

本文用 B 表示通信带宽。每个时隙中 UAV 从第 K 个 UE 接收的任务比特数为 $I_{u,k}^O[n] = \tau_k^O[n] B \log_2(1 + \gamma_{u,k}[n])$ 。

因此, 每个时隙中 UAV 传递给 AP 的 UE k 的任务比特数为 $I_{a,k}^R[n] = \tau_k^R[n] B \log_2(1 + \gamma_{a,k}[n])$ 。设 $I_{u,k}^{\text{comp}}[n]$ 表示满足最大计算能力约束的时隙 n 中 UE k 在 UAV 上的计算任务比特数, 即 $\sum_{k=1}^K I_{u,k}^{\text{comp}}[n] c_u \leq F_u^{\text{max}} \delta_i$, 其中 c_u 是 UAV 上一位计算所需的 CPU 周期, F_u^{max} 是 UAV 可以提供的最大 CPU 频率。设 c_a 和 F_a^{max} 为 AP 上一位计算所需的 CPU 周期和最大 CPU 频率。本文得到 AP 的计算约束为 $\sum_{k=1}^K I_{a,k}^R[n] c_a \leq F_a^{\text{max}} \delta_i$ 。为避

免 UAV 和 AP 的任务积压导致严重的延迟问题，本文假设当前时隙中接收到的任务比特数不会超过下一个时隙中可承受的执行能力。请注意，这项工作也可以很容易地扩展到延迟容忍的情况，该情况允许任务积压，只需稍作更改，而不会由于仿射约束而影响本文提出的算法。因此，本文获得约束为：

$$l_{u,k}^o[n] \leq l_{u,k}^R[n+1] + l_{u,k}^{\text{comp}}[n+1], n=1, \dots, N-1 \quad (24)$$

设 $l_{u,k}^{\text{loc}}[n]$ 是在第 K 个 UE 上本地计算的任务比特数，系统在每个时隙中完成的任务比特数由卸载比特数和在第 K 个 UE 在本地计算的任务比特数所组成。因此，在时隙 n 处，本文的目标是最大限度地提高 RIS 辅助的无人机 MEC 系统的能效，以增加系统完成的任务比特总量。比特总量可定义为：

$$L[n] = \sum_{k=1}^K (l_{u,k}^{\text{loc}}[n] + l_{u,k}^o[n]) \quad (25)$$

记任务期间第 K 个 UE 的最小任务比特数为 L_i ，则

$$\sum_{n=1}^N (l_{u,k}^{\text{loc}}[n] + l_{u,k}^o[n]) \geq L_i \quad (26)$$

1.5 能耗模型

在第 n 个时隙，第 K 个 UE 的任务卸载能耗可以表示为 $E_{u,k}[n] = p_{u,k}[n]t$ ，其中 $p_{u,k}[n]$ 表示 UE 的传输功率 UE k 消耗的局部计算能量可以表示为^[16]：

$$E_{u,k}^{\text{com}}[n] = \frac{k_{u,k} (l_{u,k}^{\text{loc}}[n])^3}{t^2} \quad (27)$$

$$E_U^{\text{com}} = \sum_{k=1}^K \frac{k_U (l_{u,k}^R[n])^3}{t^2} \quad (28)$$

UAV 的飞行能耗计算为 $E_U^{\text{fly}} = t \left(\vartheta_1 v^3 + \frac{\vartheta_2}{v[n]} \right)$ ，其中 ϑ_1 和 ϑ_2

是与 UAV 规格相关的两个参数。则总能耗包括所有 UE 的能耗和 UAV 的能耗为：

$$E[n] = \sum_{k=1}^K (E_{u,k}[n] + E_{u,k}^{\text{com}}[n]) + E_U^{\text{com}} + E_U^{\text{fly}} \quad (29)$$

定义能量效率 (energy efficiency, EE) 表示单位能量消耗下所完成的计算比特数，用于衡量计算能力和能耗的关系，其表达式为：

$$\text{EE} = \frac{\sum_{n=1}^N L[n]}{\sum_{n=1}^N E[n]} \quad (30)$$

2 算法设计

2.1 优化问题提出

现有的文献往往过于注重系统的 SE 却忽略了系统的 EE。然而忽略系统的 EE 会存在一些缺陷，这样会使得无人机的续航以及物联网设备所运行的时间受限^[17]。

传统的 RL 往往适用于动作空间和样本空间较小且离散的情况。然而类似 RIS 辅助的 UAV 通信的复杂任务往往有很大的状态空间和连续的动作空间，尤其是输入数据为图像

等高维度数据时，RL 很难处理此类任务^[18]。然而 DDPG (deep deterministic policy gradient) 算法是一种在连续动作空间中工作的强化学习算法，DDPG 通过引入经验回放缓冲区和目标网络，能够提高学习的稳定性和效率，特别适合本问题的求解。

基于上述的模型，本文将能源效率定义为任务周期内完成的任务比特总量与总能耗的比率，本文的目标是通过联合优化检测向量 $\mathbf{u}_k^H[n]$ ，波束形成向量 $\mathbf{w}_k[n]$ ，时隙划分 $\{\tau_k^O[n], \tau_k^R[n]\}$ ，相移 $\theta_{k,m}[n]$ ，UE 本地任务比特数 $l_{u,k}^{\text{loc}}[n]$ ，UAV 的计算比特数 $l_{u,k}^O[n]$ 和无人机的轨迹 $q[n]$ ，来最大化 UE 之间的能源效率。

因此，RIS 辅助的 UAV 系统的能效最大化问题定义的数学模型为：

$$\text{argmax EE} \quad (31)$$

s.t.

$$0 \leq \theta_{k,m}[n] < 2\pi, \forall k, m, n \quad (32)$$

$$\tau_k^O[N] = 0, \tau_k^R[1] = 0, \tau_k^R[N] = 0, \forall k \quad (33)$$

$$\tau_k^O[n] \geq 0, \tau_k^R[n] \geq 0, \forall k, n \quad (34)$$

$$\sum_{k=1}^K \tau_k^O[n] + \sum_{k=1}^K \tau_k^R[n] \leq \delta_t, \forall n \quad (35)$$

$$\|\mathbf{w}_k[n]\|^2 \leq p_u^{\text{max}}, \forall k, n \quad (36)$$

$$\frac{1}{T} \sum_{n=1}^N \sum_{k=1}^K \tau_k^R[n] \|\mathbf{w}_k[n]\|^2 \leq p_u^{\text{ave}} \quad (37)$$

$$l_{u,k}^O[n] \leq l_{u,k}^R[n+1] + l_{u,k}^{\text{comp}}[n+1], \forall k, n \quad (38)$$

$$\sum_{k=1}^K l_{u,k}^{\text{comp}}[n] c_u \leq F_u^{\text{max}} \delta_t \quad (39)$$

$$\sum_{k=1}^K l_{u,k}^R[n] c_a \leq F_a^{\text{max}} \delta_t \quad (40)$$

2.2 状态空间设计

在时隙 n ，系统状态 $s_n \in S$ 由 UAV 相对于其最终位置的相对位置 $q[n]$ ， $\theta_{k,m}[n]$ ， $w_k[n]$ 、UAV 的速度 $V[n]$ ，以及信号传输的变量 $l_{u,k}^{\text{loc}}[n]$ ， $l_{u,k}^R[n]$ ， $l_{u,k}^O[n]$ ， $\tau_k^R[n]$ 因此定义 s_t 为：

$$s_t = \{q[n], V[n], \theta_{k,m}[n], l_{u,k}^{\text{loc}}[n], l_{u,k}^R[n], l_{u,k}^O[n], \tau_k^R[n], w_k[n]\} \quad (41)$$

2.3 动作空间设计

在无人机的自主飞行中，为了实现高效准确的路径规划，需要收集所有可能的动作，并且对动作空间进行设置。动作空间 A 包含了所有可能的动作 a_n ，其中 a_n 可以通过一系列反射系数 $\theta_{k,m}[n]$ 和无人机的加速度 $A[n]$ 来描述。具体地说，动作 a_n 包括了一组反射系数 $\theta_{k,m}[n]$ ，其中每个反射系数 $\theta_{k,m}[n]$ 的取值范围在 $[0, 2\pi)$ 之间。

2.4 奖励函数设计

在每个时间步长 n ，奖励函数由两部分组成：通信部分和到最终位置的距离部分。通信部分由无人机的能量效率决定，能量效率表达为 $\sum_{n=1}^N L[n] - \alpha \sum_{n=1}^N E[n]$ ， α 是一个引入的

辅助参数。 $L[n]$ 也常用来表示 UAV 在执行任务时能够接收到的数据速率，数据速率越高，说明通信效率越高。 $E[n]$ 表示无人机执行行动所需要的能量消耗。该项的引入是为了平衡通信和能量消耗之间的权衡关系。到最终位置的距离部分由两个参数决定： d_N^{n-1} 和 d_N^n 。分别表示无人机在上一个时间步长和当前时间步长时与最终位置的距离。这个参数的作用是鼓励无人机尽快接近最终位置，以减少能量的消耗。

综上，奖励函数的具体计算公式为：

$$r_n = \sum_{n=1}^N L[n] - \alpha \sum_{n=1}^N E[n] + \varsigma(d_N^{n-1} - d_N^n) \quad (42)$$

式中： ς 是学习过程中的一个可调参数，作用是调整奖励函数对通信和距离的重视程度。

2.5 基于 DDPG 模型结构设计

具体的步骤如算法 1 所示。

算法 1：

随机初始化：critic 网络模型 $Q(s, a|\theta_Q)$ 和 actor 网络模型 $\pi(s|\theta_\pi)$

构建影子：critic 网络模型 $Q'(s, a|\theta_{Q'})$ 和 actor 网络模型 $\pi'(s|\theta_{\pi'})$

初始化经验回放池 D 的大小为 N

For episode=1:M Do

 初始化一个随机过程 $\mathcal{N}$ 用于动作探索

 即，产生一个随机噪声 $\mathcal{N}_t$ 施加到确定策略输出上；

 接受一个初始观测状态 s_1 ；

For $t=1:T$ Do

 依据当前策略及探索噪声随机选择一个动作

$$a_t = \pi(s_t|\theta_\pi) + \mathcal{N}_t$$

 执行动作 a_t ，环境反馈即时奖励 r_{t+1} 及新状态 s_{t+1}

 添加 transition $(s_t, a_t, r_{t+1}, s_{t+1})$ 到经验回放池 D 中

 从 D 中随机选取一小批 transitions $(s_j, a_j, r_{j+1}, s_{j+1})$

 计算 Q 参照值，记作 $y_j = r_{j+1} + \gamma Q'(s_{j+1}, \pi'(s_{j+1}|\theta_{\pi'})| \theta_{Q'})$

 基于均方损失 $\frac{1}{N} \sum_j (y_j - Q(s_j, a_j|\theta_Q))^2$ ，更新 critic 网络参数 θ_Q

 更新 actor 网络模型

$$\begin{aligned} \nabla_{\theta} J &\approx \mathbb{E}_{s_t \sim \rho^a} \left[\nabla_{\theta} Q(s, a|\theta_Q) \right]_{s=s_t, a=\pi(s_t|\cdot)} \\ &= \mathbb{E}_{s_t \sim \rho^s} \left[\nabla_a Q(s, a|\theta_Q) \right]_{s=s_t, a=\pi(s_t|\theta_\pi)} \nabla_{\theta} \pi(s_t|\theta_\pi) \Big|_{s=s_t} \\ &\approx \frac{1}{N} \sum_i \nabla_a Q(s, a|\theta_Q) \Big|_{s=s_t, a=\pi(s_t)} \nabla_{\theta} \pi(s|\theta_\pi) \Big|_{s_i} \end{aligned}$$

 平稳地更新影子 critic Q' 和影子 actor π' 的网络模型参数

$$\theta_{Q'} = \tau \theta_Q + (1 - \tau) \theta_{Q'}$$

$$\theta_{\pi'} = \tau \theta_\pi + (1 - \tau) \theta_{\pi'}$$

End For

DDPG 使用 4 个 DNN 来近似值，包括训练 Actor 网络和训练 Critic 网络，以及目标 Actor 网络和目标 Critic 网络。通

过使用确定性策略梯度定理来更新参数，并使用目标网络来防止不稳定的学习。

3 实验设计与结果分析

3.1 实验设计

考虑所有网络节点位于 RIS 反射侧的区域。具体而言，RIS 位于 yoz 平面内，其坐标为 (0 m, 200 m, 20 m)，AP 的坐标为 (200 m, 400 m, 0 m)，UEs 位于 xoy 平面内。UAV 的高度设置为 100 m，它从初始位置 (100 m, -800 m) 飞行到最终位置 (800 m, 400 m)，最大速度为 10 m/s，整个过程的总时长 T 为 60 s，每个时隙 δ_t 的时间长度为 1 s，UE 的设备个数为 5 个，天线 / 元件分离间隙分别为 $d = \lambda/4$ ， $d_y = d_z = \lambda/4$ ，反射元件 M_y 为 10 个， M_z 为 10 个，UAV 天线的数量 L 为 4 个，通信带宽 B 为 1 MHz，噪音功率 σ^2 为 -90 dBm，Rician 因子 $\beta_{u,k} = \beta_{ua} = \beta_{ra} = 10$ dB，路径损耗指数 $\alpha_{u,k} = 2.5$ ， $\alpha_{ua} = 3$ ， $\alpha_{ra} = 2.3$ ，UE 的发射功率 $P_t = 26$ dBm，UAV 峰值功率 $p_u^{\max} = 30$ dBm，平均功率预算 $p_u^{\text{ave}} = 28$ dBm，最大 CPU 频率 $F_u^{\max} = 1$ GHz， $F_u^{\max} = 15$ GHz，每比特计算的 CPU 周期 $c_u = c_a = 10^3$ cycles/bit，收敛精度 $\epsilon = 10^{-3}$ ，最大迭代次数 $j = 50$ 。奖励函数的可调参数 ς 设置为 0.32，训练 Actor 网络和训练 Critic 网络的学习率 η_a 和 η_c 为 10^{-4} ，训练 Actor 网络和训练 Critic 网络的衰减率 τ_a 和 τ_c 设置为 5×10^{-3} ，小批量 N_{batch} 的经验数量为 128，勘探噪声的方差 $\sigma_{\mathcal{N}}^2 = 0.2$ 。

设置两个基准实验：

基准实验 1：该基线将 DDPG 应用于在没有 RIS 的情况下优化，称为“无 RIS 的基线”或简单地说“无 RIS”。

基准实验 2：这个基线使用传统的基于凹凸过程 (CCCP) 算法在给定的 RIS 以及 UAV 配置条件下优化。CCCP 的算法将公式化问题分解为 3 个子问题来解决该问题。

3.2 实验结果分析

训练的收敛曲线如图 2 所示。随着无人机控制策略适应随机生成，平均奖励逐渐增加每一次任务的目标轨迹。DDPG 算法在第 417 轮迭代达到最大奖励。基准实验 1 在第 873 轮迭代达到最大奖励，基准实验 2 在第 1 424 轮迭代达到最大奖励。

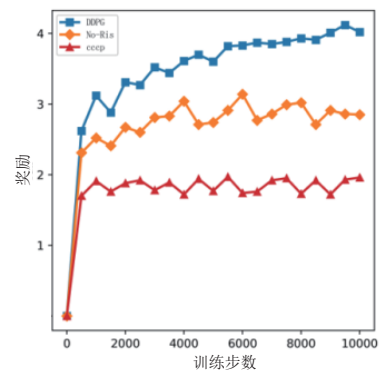


图 2 模型收敛曲线

由于所有涉及的信道的完美和瞬时 CSI 的不现实假设,DDPG 能够快速收敛是由于它的无人机加速度的动作空间大大更小。

图 3 (a) 显示了无人机接收数据率与任务结束时到目的地的完成距离之间的关系。在任务持续时间为 25~40 s 时,速率和完成距离之间的权衡较好。探索 DDPG 新动作时,目标网络和噪声增加了随机性,导致训练过程不稳定。图 3 (b) 展示了任务结束时无人机与预期目标之间距离的 CDF,任务持续时间为 40 s。结果表明,随着任务持续时间的增加,DDPG 算法更接近或到达目的地的频率更高,优于没有 RIS 的情况。尽管完美的 CSI 对轨迹规划重要,DDPG 算法通过使用更多 RIS 元素也能产生有效轨迹,但存在过拟合问题,导致策略更新噪声梯度和较低的数据率。

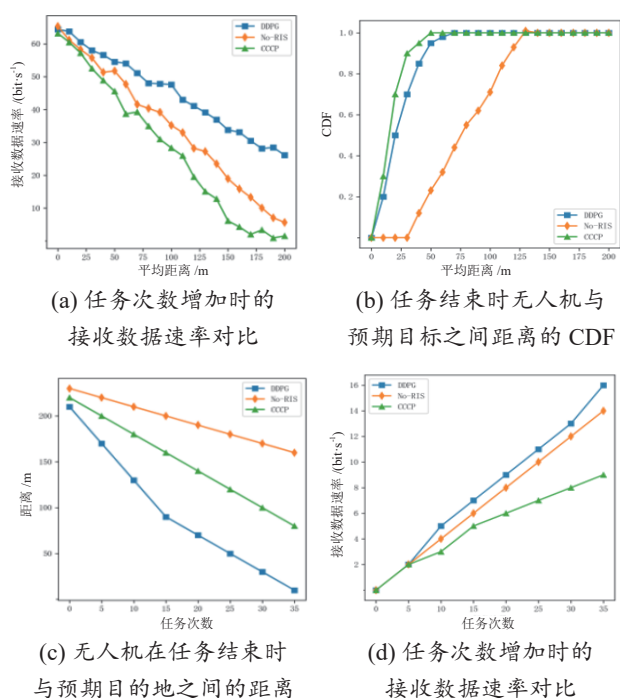


图 3 模型参数变化的性能指标差异

图 3 (c) 显示了任务结束时无人机与预期目的地之间的距离。随着任务持续时间增加,所有算法的距离减小。没有 RIS 时,无人机无法到达预期位置,因为必须接近 BS 以保持数据率。

图 3 (d) 显示了任务次数增加时的接收数据速率。所有算法下,接收数据速率随任务持续时间延长而增加,RIS 提高了数据速率。DDPG 算法的数据速率优于基准实验 1,证明了其适应系统变化的能力,特别是对于大量 RIS 元素。

4 结论

本文提出一种基于 DDPG 的 RIS、UAV 和 MEC 的波束赋形优化。通过优化波束形成向量、相移、时隙划分、无人机的轨迹,最大化系统的 EE。实验结果表明基于 DDPG 的方法比传统的 CCCP 算法收敛更快且更稳定。本文提出的基

于 DDPG 网络的 RIS 辅助无人机,在无人机轨迹控制和 RIS 配置方面表现优异,提高了系统的 EE。

参考文献:

- [1] SUN L, WAN L T, WANG X P. Learning-based resource allocation strategy for industrial IoT in UAV-enabled MEC systems[J].IEEE transactions on industrial informatics,2021,17(7):5031-5040.
- [2] LIU Y W, LIU X, MU X D, et al. Reconfigurable intelligent surfaces: principles and opportunities[J].IEEE communications surveys & tutorials, 2021,23(3):1546-1577.
- [3] LIU X, LIU Y W, CHEN Y, et al. RIS enhanced massive non-orthogonal multiple access networks: deployment and passive beamforming design[J].IEEE journal on selected areas in communications,2021,39(4): 1057-1071.
- [4] WU Q Q, ZHANG R. Joint active and passive beamforming optimization for intelligent reflecting surface assisted SWIPT under QoS constraints[J].IEEE journal on selected areas in communications, 2020,38(8):1735-1748.
- [5] WANG J, LIANG Y C, JOUNG J, et al. Joint beamforming and reconfigurable intelligent surface design for two-way relay networks[J].IEEE transactions on communications, 2021,69(8):5620-5633.
- [6] LUO J S, WANG F G, WANG S L, et al. Reconfigurable intelligent surface: reflection design against passive eavesdropping[J].IEEE transactions on wireless communications, 2021,20(5):3350-3364.
- [7] PANG X W, SHENG M, ZHAO N, et al. When UAV meets IRS: expanding air-ground networks via passive reflection[J].IEEE wireless communications,2021, 28(5):164-170.
- [8] LI S X, DUO B, RENZO M D, et al. Robust secure UAV communications with the aid of reconfigurable intelligent surfaces[J].IEEE transactions on wireless communications,2021, 20(10):6402-6417.
- [9] FANG S S, CHEN G J, LI Y H. Joint optimization for secure intelligent reflecting surface assisted UAV networks[J]. IEEE wireless communications letters, 2021,10(2):276-280.
- [10] LIU X, LIU Y W, CHEN Y. Machine learning empowered trajectory and passive Beamforming design in UAV-RIS wireless networks[J].IEEE journal on selected areas in communication,2021,39(7):2042-2055.
- [11] WEI Z Q, CAI Y X, SUN Z, et al. Sum-rate maximization for IRS-assisted UAV OFDMA communication systems[J].IEEE transactions on wireless communications, 2021,20(4):2530-2550.
- [12] LU H Q, ZENG Y, JIN S, et al. Aerial intelligent reflecting

基于 STM32 的光伏储能 DC-DC 变换器设计

刘一儒¹ 王 宇¹ 孟丽因¹
LIU Yiru WANG Yu MENG Linan

摘 要

为满足光伏储能系统中宽范围输入与稳定输出的电力转换需求，文章设计了一种基于 STM32F103C6T6 的高效光伏储能双向 DC-DC 变换器，采用 Buck-Boost 拓扑和全桥同步整流技术，结合 PID 闭环控制与 MPPT 扰动观察算法，实现了输入 25~50 V 宽范围、输出 30 V/5 A 的稳定转换。通过优化 IR2184S 驱动电路，系统效率超过 90%，动态响应速度达 80 μ s，输出电压波动小于 ± 0.03 V。测试结果表明，在负载变化和输入电压波动条件下，系统电压调整率仅为 0.033%，验证了设计方案的有效性和稳定性，为光伏储能系统提供了可靠的电力转换解决方案。

关键词 双向直流变换器；嵌入式系统；PID 算法；最大功率跟踪

doi: 10.3969/j.issn.1672-9528.2025.08.020

0 引言

随着可再生能源的普及，光伏发电系统在分布式能源领域愈发受瞩目。鉴于 DC-DC 变换器对提升光伏系统能量利用效率与稳定性举足轻重，其设计优化至关重要。本文聚焦基于 STM32 的光伏发电系统 DC-DC 变换器，以 STM-32F103C6T6 单片机为核心，借双向 DC-DC 电路实现电能转换，利用单片机生成的 PWM 信号驱动电路以保障负载电压稳定。设计融入 PID 与 MPPT 算法，依电流、电压反馈达成

恒压闭环控制、最大功率点跟踪及电池组充放电管理。经模拟光伏发电实验，系统直流电压输出稳定，效率超 90%，通过优化电路与算法，为光伏系统提供了高效可靠方案。

1 总体设计

本系统以 STM32 单片机为核心，构建了基于 PWM 控制的高效 DC-DC 变换器。引入 PID 与 MPPT 算法，实现恒压输出、最大功率点跟踪及电池充放电管理。系统由双向 DC-DC、辅助电源、驱动、电流采样与显示模块组成，在模拟光伏条件下运行稳定、效率超 90%，整体结构如图 1 所示。

1. 辽宁工业大学电子与信息工程学院 辽宁锦州 121000

surface: joint placement and passive beamforming design with 3D beam flattening[J].IEEE transactions on wireless communications, 2021,20(7):4128-4143.

[13] KHALILI A, MONFARED E M, ZARGARI S, et al.Resource management for transmit power minimization in UAV-assisted RIS HetNets supported by dual connectivity[J]. IEEE transactions on wireless communications,2022,21(3): 1806-1822.

[14] PAN Y J, WANG K Z, PAN C H, et al.UAV-assisted and intelligent reflecting surfaces-supported terahertz communications[J].IEEE wireless communications letters, 2021,10(6):1256-1260.

[15] HUA M, YANG L X, WU Q Q, C. et al.UAV-assisted intelligent reflecting surface symbiotic radio system[J].IEEE transactions on wireless communications, 2021, 20(9): 5769-5785.

[16] XU Y, ZHANG T K, LIU Y W, et al. Computation capacity enhancement by joint UAV and RIS design in IoT[J]. IEEE internet of things journal, 2022, 9(20): 20590-20603.

[17] GE M X, WANG L Y, ZHANG G L, et al. Joint optimization of task offloading and resource allocation in UAV-assisted mec networks[C]//2021 China Automation Congress (CAC). Piscataway:IEEE, 2021: 1879-1884.

[18] LIU C J, ZHONG Y L, WU R L, et al. Deep reinforcement learning based 3D-trajectory design and task offloading in UAV-enabled MEC system[J]. IEEE transactions on vehicular technology, 2024,74(2):3185-3195.

【作者简介】
高航(2000—)，男，辽宁大连人，硕士研究生，研究方向：无线与移动通信技术。
(收稿日期：2025-02-28 修回日期：2025-07-29)