中国科学:信息科学 2026 年 第 56 卷 第 11 期: 2917–2934
SCIENTIA SINICA Informationis
Agentic AI 赋能的 6G 融合通信专题 • 论文

Agentic AI 赋能的低空智联网传输优化方法

马张枫[1,3], 张锐晨[2*], 廖柏林[1], 朱秋明[3], 吴启晖[3]

1. 邵阳学院信息科学与工程学院, 邵阳 422000, 中国
2. College of Computing and Data Science, Nanyang Technological University, Singapore 639798, Singapore
3. 南京航空航天大学电子信息工程学院, 南京 211106, 中国
* 通信作者. E-mail: ruichen.zhang@ntu.edu.sg

收稿日期: 2026–03–31; 修回日期: 2026–08–13; 接受日期: 2026–10–16; 网络出版日期: 2026–11–09

国家自然科学基金 (批准号: 62401371)、江苏省科技重大专项 (批准号: BG2025039) 和中国博士后科学基金 (批准号: 2025M774279) 资助项目

摘要 为助力低空经济智能化发展, 基于蜂窝网络的无人机 (unmanned aerial vehicle, UAV) 通信系统成为空地协同连接的核心方案. 本文利用前沿技术赋能 UAV 蜂窝网络智能建模与优化, 开展了 Agentic AI 赋能的低空智联网传输优化方法研究. 融合检索增强生成与语义路由技术, 构建了具备规划、工具调用与反馈修正能力的生成式 AI 代理框架, 精准刻画了传播场景、部署方式、中继协议与优化目标四个维度下的建模需求; 基于代理生成的系统模型, 利用混合专家 (mixture of experts, MoE) 与近端策略优化 (proximal policy optimization, PPO) 相结合的算法, 对放大转发和解码转发两类协议下的功率分配进行了求解; 在获得最优功率配置后, 从可达速率与约束违反率两方面对算法性能进行了评估. 结果表明: 所提 MoE-PPO 算法的可达速率约为 4.56 bit/s/Hz, 较 PPO 算法提升 12.3%, 约束违反率由 2.9% 减小至 0.8%; 等参数量对照下 MoE 结构本身贡献 9.8% 的增益; 代理输出与人工建模结果的一致率为 79.5%, 移除检索增强生成模块后减小至 55.4%. 因此, 所提框架能够显著提升 UAV 蜂窝网络定制化建模的自动化程度, 缓解优化变量强耦合对求解性能的制约, 为低空智联网传输优化的设计与实现提供了理论基础和数据支撑.

关键词 低空智联网, 大语言模型, 生成式 AI 代理, 混合专家, 近端策略优化

1 引言

近年来, 世界各国纷纷将低空经济纳入国家顶层布局, 低空产业智能化发展已成为全球经济竞争的新前沿[1]. 2023 年, 中国正式将低空经济列为新兴支柱产业, 赋予其突出的优先发展地位; 此后, 2024 年和 2025 年政府工作报告及党的二十届三中全会相继作出专项部署, 持续强化政策引领, 推动市场规模向万亿量级跨越. 与此同时, 空域管理体制的渐进式开放催生出载人出行、低空物流、空中交通等多元化应用场景, 高密度、高频次、异构化的飞行运营模式逐步成为低空空域的典型特征[2]. 然而, 现有通信基础设施在覆盖连续性、多节点协同接入及大容量并发传输等方面存在明显短板, 难以有效支撑上述复杂运营需求, 亟需构建新型的低空智联网, 以夯实低空经济高质量发展的技术底座[3].

引用格式: 马张枫, 张锐晨, 廖柏林, 等. Agentic AI 赋能的低空智联网传输优化方法. 中国科学: 信息科学, 2026, 56: 2917–2934, doi: 10.1360/SSI-2026-0120
Ma Z F, Zhang R C, Liao B L, et al. Agentic AI-enabled transmission optimization for low-altitude intelligent networks. Sci Sin Inform, 2026, 56: 2917–2934, doi: 10.1360/SSI-2026-0120

© 2026《中国科学》杂志社

作为有人驾驶与无人驾驶航空器密集交汇处, 低空空域具有节点分布立体化、业务需求异构化、传播环境复杂多变等显著特点, 属于高动态、高密度的通信保障核心区域 [4]. 目前, 低空通信网络普遍依托地面蜂窝基站 (base station, BS) 向上延伸覆盖, 借助无人机 (unmanned aerial vehicle, UAV) 承担空中中继与接入服务 [5]. 然而, 随着应用场景的持续扩展, UAV 蜂窝通信系统面临日益突出的技术瓶颈. 空间部署复杂度高, 协议参数强耦合, 网络配置高度依赖领域专家经验, 难以满足低空运营的多样化需求. 因此, 仅依赖人工建模与参数调优这种被动适应信道变化的传统方法已难以为继, 亟需引入智能化技术, 使建模与求解两个环节均主动感知场景特征、自动优化配置, 从而实现低空通信网络的高效优化 [6].

作为 AI 领域最具变革潜力的方向之一, Agentic AI 能够实现对复杂任务的自主感知、动态规划、工具调用与闭环决策, 并根据环境反馈持续调整执行策略, 从而按需完成高度复杂的端到端自动化任务 [7~9]. 目前, 国内外学者已针对不同场景下如何借助 Agentic AI 提升系统智能化开展了广泛研究. 借助大语言模型 (large language model, LLM) 与混合专家 (mixture of experts, MoE) , 文献 [10] 提出面向卫星网络传输的生成式 AI 代理框架, 通过多专家协同实现复杂任务的智能调度与自适应优化. 文献 [11] 进一步拓展至更普适的网络场景, 验证了该方法在多样化网络任务中的泛化能力与性能优势. 文献 [12] 提出了知识图谱与检索增强生成 (retrieval-augmented generation, RAG) 相融合的无线网络智能设计框架, 显著提升 LLM 对专业领域知识的推理精度. 文献 [13] 基于 LLM 实现无线信道预测, 有效降低了信道建模对领域先验知识的学习门槛. 文献 [14] 系统探讨了 LLM 赋能下一代无线网络的基础理论、关键挑战与发展愿景, 进一步验证了其在复杂无线通信系统智能化设计中的广阔应用. 然而, 上述工作的辅助对象均为单一维度的建模, 缺乏对传播场景、部署方式、中继协议与优化目标等多维度联合建模的系统性框架, 也未考虑语义路由机制对检索效率的影响, 难以支撑 UAV 蜂窝网络这类系统级定制化建模需求.

近年来, 深度强化学习 (deep reinforcement learning, DRL) 成为低空网络无线资源管理的有效手段之一. 文献 [15] 采用深度 Q 网络 (deep Q-network, DQN) 对异构任务与服务质量感知的 UAV 中继通信进行联合优化, 但 DQN 的离散动作空间限制了其对连续变量的优化精度. 文献 [16] 采用近端策略优化 (proximal policy optimization, PPO) 算法求解 UAV 网络的资源分配问题, 其策略网络可直接输出连续取值的资源变量, 避免了动作离散化带来的精度损失. 然而, 标准 PPO 算法由单一策略网络同时输出全部优化变量, 在多变量联合优化中探索效率低, 易收敛至次优解. 文献 [17] 系统综述了 MoE 在无线网络生成式 AI 与 DRL 中的应用, 其门控网络依据输入为各专家分配权重, 使不同专家分别适配不同类型的任务. 然而, 现有工作中的专家分工均发生在不同任务之间, 尚未见将其用于同一优化问题中相互耦合的决策变量. 更重要的是, 上述建模与优化两类工作彼此独立开展, 建模所得的系统模型与约束条件须经人工整理方能构成求解算法的输入, 二者之间缺乏统一的框架予以衔接, 难以适应低空场景快速变化的组网需求.

针对上述问题, 本文从传播场景、部署方式、中继协议与优化目标四个维度构建了面向 UAV 蜂窝网络的领域知识库, 为各维度的知识块赋予统一的组织结构, 并设计两层语义路由机制与 RAG 流程, 实现对专业知识的精准召回. 其次, 依据召回的领域知识, 设计了具备规划、工具调用与反馈修正能力的生成式 AI 代理, 以及 MoE-PPO 算法, 并对放大转发 (amplify-and-forward, AF) 和解码转发 (decode-and-forward, DF) 两类协议下的功率分配方案进行了建模、求解和分析. 研究结果表明, 借助生成式 AI 代理可以对复杂的 UAV 蜂窝网络进行多维度的自动化建模. 通过将 MoE 结构嵌入演员网络, 可显著提升系统可达速率并降低约束违反率, 以此来缓解优化变量强耦合对求解性能的制约, 为低空智联网的传输优化提供了新思路.

符号说明. 非加粗变量表示标量, 加粗小写变量表示向量, 加粗大写变量表示矩阵. $\Xi_{N_{\mathrm{B}}}(x)$ 表示 N_{B} 阶狄利克雷核, 定义为 $\Xi_{N_{\mathrm{B}}}(x)=\sin(N_{\mathrm{B}}x/2)/(N_{\mathrm{B}}\sin(x/2))$; 当 $x=2k\pi$ ($k\in\mathbb{Z}$) 时, $\Xi_{N_{\mathrm{B}}}(2k\pi)=(-1)^{k(N_{\mathrm{B}}-1)}$. $k_0=2\pi/\lambda$ 表示自由空间波数, $\lambda=c_0/f_c$ 表示波长, c_0 表示光速, f_c 表示载频. $\exp(\cdot)$ 表示自然指数函数; $\mathrm{softmax}(\mathbf{x})_i=\exp(x_i)/\sum_j\exp(x_j)$ 表示 softmax 函数; $\sigma(x)=1/(1+\exp(-x))$ 表示 sigmoid 函数; $\tanh(x)=(\exp(x)-\exp(-x))/(\exp(x)+\exp(-x))$ 表示双曲正切函数; $\mathrm{ReLU}(x)=\max(0,x)$ 表示线性整流函数; $\mathrm{clip}(x,a,b)=\max(a,\min(b,x))$ 表示截断函数; $\mathrm{KL}(p\|q)=\sum_x p(x)\log(p(x)/q(x))$ 表示 Kullback-Leibler 散度; $\Gamma(s)=\gamma(s,\infty)$ 表示伽马函数, $\gamma(s,x)=\int_0^x t^{s-1}\mathrm{e}^{-t}\,\mathrm{d}t$ 表示不完全伽马函数; $\mathrm{Bernoulli}(p)$ 表示参数为 p 的伯努利分布; $\mathbf{1}_{\{\cdot\}}$ 表示

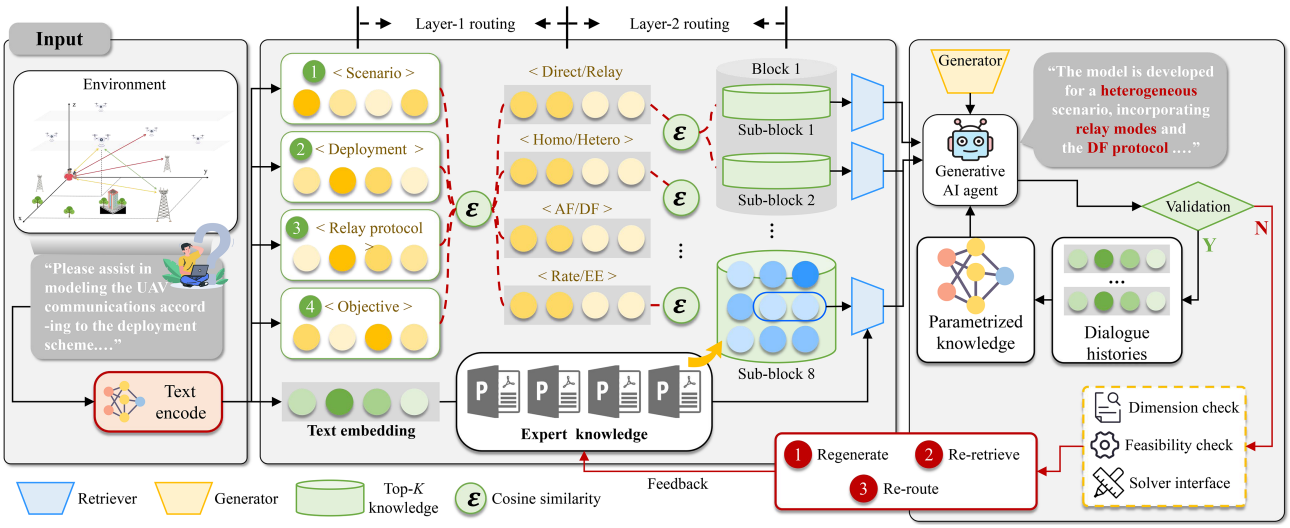

图 1　(网络版彩图) 面向 UAV 蜂窝网络建模的生成式 AI 代理框架.

Figure 1　(Color online) The generative AI agent framework for cellular-connected UAV network modeling.

示性函数; $\mathbb{E}[\cdot]$ 表示期望算子; $|\cdot|$ 表示集合基数; $\oplus$ 表示上下文拼接算子.

2　生成式 AI 代理框架

本节构建面向 UAV 蜂窝网络定制化建模的生成式 AI 代理框架, 该框架以 LLM 为核心推理引擎, 融合 RAG 技术调用 UAV 蜂窝网络专用知识库, 实现建模任务的自动化, 其总体结构如图 1 所示.

2.1　系统建模

UAV 蜂窝网络受三维空间部署、空对地信道特性差异、多类型中继传输等约束影响, 数学建模复杂度远高于传统地面通信系统. 为支撑生成式 AI 代理的定制化建模, 本小节首先从传播场景、部署方式、中继协议与优化目标四个核心维度, 给出 UAV 蜂窝网络的建模要素及其解析表达式.

2.1.1　传播场景

BS 通过光纤接入核心网, UAV 经无线回传链路与 BS 连接, 二者协同为坐标原点的用户设备 (user equipment, UE) 提供服务 [18], 如图 2 所示. 本文提出自适应混合服务方案, UE 可按需选择两种通信模式. (i) 直连模式: UE 通过单跳接入链路直接接入 BS; (ii) 中继模式: 以 UAV 为中继, 经空中接入链路与无线回传链路构成的双跳路径接入 BS. 采用随机几何理论完成系统建模, 结合选择合并策略, 对 UE 在不同信道条件下的信干噪比 (signal-to-interference-plus-noise ratio, SINR) 进行统一表征

$$\mathrm{SINR} = \max\{\mathrm{SINR}_{\mathrm{BU}}, \mathrm{SINR}_{\mathrm{e2e}}\}, \tag{1}$$

其中, $\mathrm{SINR}_{\mathrm{BU}}$ 为直连链路 SINR, $\mathrm{SINR}_{\mathrm{e2e}}$ 为中继链路端到端 SINR. 假设 BS 高度为 h_{B}, 其水平投影位置服从密度 λ_{B} 的二维齐次泊松点过程 (homogeneous Poisson point process, HPPP) Φ_{B}; 其天线阵列为均匀线性阵列, 阵元数和间距分别为 N_{B} 和 d_{B}, 则全向天线增益为

$$G_{\mathrm{B}}^{\mathrm{OmniD}}(\theta,\theta_{\mathrm{B}}) = G_{\max} + G_{\mathrm{E,V}}(\theta) + 20\log_{10}(|f_{\mathrm{A}}(\theta,\theta_{\mathrm{B}})|), \quad \theta_{\mathrm{B}} \in \left(\frac{\pi}{2},\pi\right), \tag{2}$$

其中, θ 为以 z 轴正方向为参考的天顶角, θ_{B} 为天线主瓣下倾角, $G_{\max}$ 表示阵元的最大增益, $f_{\mathrm{A}}(\theta,\theta_{\mathrm{B}}) = \Xi_{N_{\mathrm{B}}}(k_0 d_{\mathrm{B}}(\cos\theta - \cos\theta_{\mathrm{B}}))$ 表示阵列因子, $G_{\mathrm{E,V}}(\theta) = -\min\{12\left((\theta-\pi/2)/\theta_{\mathrm{3dB}}^{\mathrm{AC}}\right)^2, \mathrm{SLA}_{\mathrm{V}}\}$ 表示垂直辐射方向图,

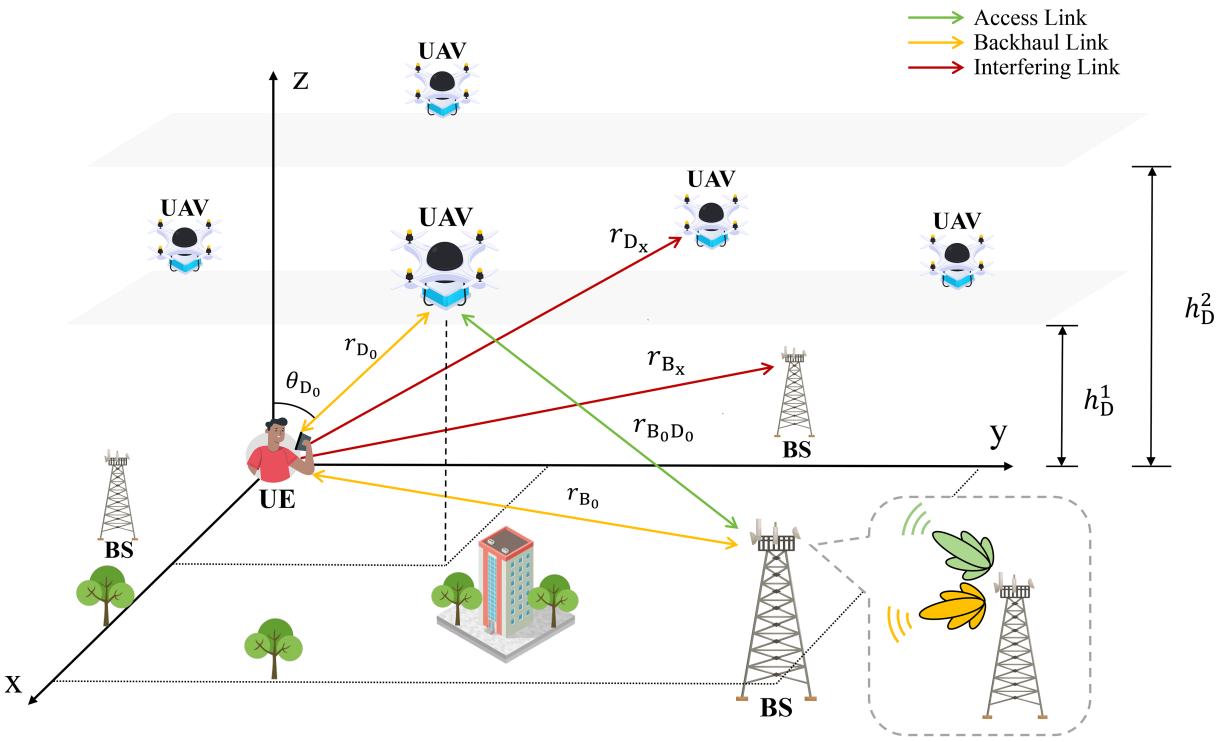

图 2　(网络版彩图) 系统模型.

Figure 2　(Color online) An illustration of the system model.

$\theta_{3\mathrm{dB}}^{\mathrm{AC}}$ 表示接入天线的半功率波束宽度, $\mathrm{SLA_V}$ 表示旁瓣电平衰减限值. 针对全向天线通过旁瓣覆盖 UAV 引发的通信中断, 本文在 BS 侧额外配置定向天线, 与全向天线共享射频前端总功率, 其增益为

$$G_{\mathrm{B}}^{\mathrm{DirU}}(\theta,\phi,\theta_0,\phi_0)=G_{\max}-\min\{-G_{\mathrm{V}}(\theta,\theta_0)-G_{\mathrm{H}}(\phi,\phi_0),A_{\mathrm{m}}\},\quad \theta_0\in\left(0,\frac{\pi}{2}\right), \tag{3}$$

其中, $G_{\mathrm{V}}(\theta,\theta_0)$ 和 $G_{\mathrm{H}}(\phi,\phi_0)$ 分别表示天线垂直与水平增益分量, A_{m} 表示天线前后向比. 此外, UAV 侧配置两套互补天线子系统. 其一, 回传天线复用 BS 上倾定向天线方向图, 增益为

$$G_{\mathrm{D}}^{\mathrm{BH}}(\theta,\phi,\theta_0,\phi_0)=G_{\mathrm{B}}^{\mathrm{DirU}}(\theta,\phi,\theta_0,\phi_0). \tag{4}$$

其二, 接入天线采用主瓣垂直向下的下倾全向天线, 增益为

$$G_{\mathrm{D}}^{\mathrm{AC}}(\theta)=G_{\max}-\min\left\{12\left(\frac{\theta-\pi}{\theta_{3\mathrm{dB}}^{\mathrm{AC}}}\right)^2,\mathrm{SLA_V}\right\}. \tag{5}$$

最后, UE 配备增益为 G_{U} 的各向同性天线.

2.1.2　部署方式

UAV 与地面节点的视距 (line-of-sight, LoS) 概率可表示为 [18]

$$\mathbb{P}_{\mathrm{L}}(\theta)=\frac{1}{1+c_1\exp\left(-c_2\left(90^\circ-\theta-c_1\right)\right)}, \tag{6}$$

其中, c_1、c_2 为环境相关常数. 本文构建两类典型 UAV 部署场景, 适配不同应用需求.

(1) 同构部署. 由单一类型 UAV 构成空中中继层, 与地面 BS 协同服务 UE, 适用于广域覆盖、应急通信等场景. UAV 空间位置在区域 $\mathcal{V}_{\mathrm{D}}=\mathbb{R}^2\times[h_{\mathrm{D}}^1,h_{\mathrm{D}}^2]$ 内服从密度为 λ_{D} 的三维 HPPP Φ_{D}, 并满足 $h_{\mathrm{D}}^1>h_{\mathrm{B}}$.

(2) 异构部署. 由多类功能互补 UAV 构建多层空中中继体系, 适用于强遮挡、高容量等复杂场景. 本文将其建模为两层独立 UAV 集群, 均服从三维 HPPP. (i) 高空 UAV 层 $\Phi_{\mathrm{D,high}}$: 密度为 $\lambda_{\mathrm{D,high}}$, 悬停高度为 $[h^1_{\mathrm{D,high}}, h^2_{\mathrm{D,high}}]$, 负责广域覆盖与远距离回传; (ii) 低空 UAV 层 $\Phi_{\mathrm{D,low}}$: 密度为 $\lambda_{\mathrm{D,low}}$, 悬停高度为 $[h^1_{\mathrm{D,low}}, h^2_{\mathrm{D,low}}]$, 负责热点区域容量增强. 异构场景的 UAV 记为 $\Phi_{\mathrm{D}}^{\mathrm{het}} = \Phi_{\mathrm{D,high}} \cup \Phi_{\mathrm{D,low}}$. 进一步, 两层高度满足 $h_{\mathrm{B}} < h^1_{\mathrm{D,low}} < h^2_{\mathrm{D,low}} < h^1_{\mathrm{D,high}} < h^2_{\mathrm{D,high}}$.

2.1.3　中继协议

为提升传输可靠性, 本文在 UAV 侧引入 AF 和 DF 两类协议. 基于选择合并策略, 两类协议下 UE 端到端 SINR 分别为

$$\mathrm{SINR}^{\mathrm{AF}} = \max\left\{\mathrm{SINR}_{\mathrm{BU}}, \mathrm{SINR}^{\mathrm{AF}}_{\mathrm{e2e}}\right\}, \tag{7}$$

$$\mathrm{SINR}^{\mathrm{DF}} = \max\left\{\mathrm{SINR}_{\mathrm{BU}}, \mathrm{SINR}^{\mathrm{DF}}_{\mathrm{e2e}}\right\}, \tag{8}$$

其中, $\mathrm{SINR}^{\mathrm{AF}}_{\mathrm{e2e}}$ 和 $\mathrm{SINR}^{\mathrm{DF}}_{\mathrm{e2e}}$ 分别表示为

$$\mathrm{SINR}^{\mathrm{AF}}_{\mathrm{e2e}} = \frac{\mathrm{SINR}_{\mathrm{BD}} \cdot \mathrm{SINR}_{\mathrm{DU}}}{\mathrm{SINR}_{\mathrm{BD}} + \mathrm{SINR}_{\mathrm{DU}} + 1}, \tag{9}$$

$$\mathrm{SINR}^{\mathrm{DF}}_{\mathrm{e2e}} = \min\{\mathrm{SINR}_{\mathrm{BD}}, \mathrm{SINR}_{\mathrm{DU}}\}, \tag{10}$$

其中, $\mathrm{SINR}_{\mathrm{BU}}$、$\mathrm{SINR}_{\mathrm{BD}}$ 与 $\mathrm{SINR}_{\mathrm{DU}}$ 分别为直连链路、回传链路与空中接入链路的 SINR.

假设 BS-UE 链路为非视距 (non-line-of-sight, NLoS) 传播, BS-UAV 链路为 LoS 传播, UAV-UE 链路为 LoS 与 NLoS 混合传播. 任意 $\mathrm{B}_x \in \Phi_{\mathrm{B}}$ 与 $\mathrm{D}_x \in \Phi_{\mathrm{D}}$ 同 UE 的距离分别为 $r_{\mathrm{B}_x} = \sqrt{u^2_{\mathrm{B}_x} + h^2_{\mathrm{B}}}$ 与 $r_{\mathrm{D}_x} = \sqrt{u^2_{\mathrm{D}_x} + h^2_{\mathrm{D}_x}}$, u_{B_x}、u_{D_x} 为地面水平投影距离; 第 $q \in \{\mathrm{L}, \mathrm{N}\}$ 类链路的大尺度衰落为 $\ell_q(r) = r^{-\alpha_q}\eta_q^{-1}$, α_q 和 η_q 分别为路径损耗指数和额外平均损耗; 小尺度衰落由 Nakagami-m 模型刻画, 其概率密度函数为 $f_X(x) = m^m x^{m-1}\exp(-mx)/\Gamma(m)$. 基于最大平均接收功率准则, 服务 BS 与服务 UAV 的选取规则分别为 $\mathrm{B}_0 = \arg\max_{\mathrm{B}_x \in \Phi_{\mathrm{B}}} P^{\mathrm{Rx}}_{\mathrm{B}_x}$ 与 $\mathrm{D}_0 = \arg\max_{\mathrm{D}_x \in \Phi_{\mathrm{D}}} P^{\mathrm{Rx}}_{\mathrm{D}_x}$, 对应的服务链路接收功率为 $P^{\mathrm{Rx}}_{\mathrm{B}_0} = P^{\mathrm{AC}}_{\mathrm{B}} G^{\mathrm{AC}}_{\mathrm{B}_0} G_{\mathrm{U}} h_{\mathrm{B}_0} \ell_{\mathrm{N}}(r_{\mathrm{B}_0})$、$P^{\mathrm{Rx}}_{\mathrm{B}_0\mathrm{D}_0} = P^{\mathrm{BH}}_{\mathrm{B}} G^{\mathrm{BH}}_{\mathrm{B}_0} G^{\mathrm{BH}}_{\mathrm{D}_0} h_{\mathrm{B}_0\mathrm{D}_0} \ell_{\mathrm{L}}(r_{\mathrm{B}_0\mathrm{D}_0})$ 与 $P^{\mathrm{Rx}}_{\mathrm{D}_0} = P_{\mathrm{D}} G^{\mathrm{AC}}_{\mathrm{D}_0} G_{\mathrm{U}} h_{\mathrm{D}_0} \ell_{q_0}(r_{\mathrm{D}_0})$, 其中 h_{B_0}、$h_{\mathrm{B}_0\mathrm{D}_0}$、$h_{\mathrm{D}_0}$ 与 r_{B_0}、$r_{\mathrm{B}_0\mathrm{D}_0}$、$r_{\mathrm{D}_0}$ 分别为三条链路的小尺度衰落与传播距离. 服务节点之外的节点为潜在干扰源, 记非服务 BS 集合为 $\Phi'_{\mathrm{B}} = \Phi_{\mathrm{B}} \backslash \mathrm{B}_0$、非服务 UAV 集合为 $\Phi'_{\mathrm{D}} = \Phi_{\mathrm{D}} \backslash \mathrm{D}_0$.

受频率复用与资源调度的影响, 非服务节点并不总与服务链路占用同一频段. 为此, 引入相互独立且与节点位置无关的同频激活变量 $a_{\mathrm{B}_x} \sim \mathrm{Bernoulli}(p_{\mathrm{B,act}})$ 与 $a_{\mathrm{D}_x} \sim \mathrm{Bernoulli}(p_{\mathrm{D,act}})$, $a_{\mathrm{B}_x} = 1$ 与 $a_{\mathrm{D}_x} = 1$ 分别表示非服务 BS B_x 与非服务 UAV D_x 在当前频段发射, $p_{\mathrm{B,act}}$ 与 $p_{\mathrm{D,act}}$ 为相应的同频激活概率. 非服务 BS 与非服务 UAV 对 UE 的累计干扰分别为 $I_{\mathrm{BU}} = \sum_{\mathrm{B}_x \in \Phi'_{\mathrm{B}}} a_{\mathrm{B}_x} P^{\mathrm{AC}}_{\mathrm{B}_x} G^{\mathrm{AC}}_{\mathrm{B}_x} G_{\mathrm{U}} h_{\mathrm{B}_x} \ell_{\mathrm{N}}(r_{\mathrm{B}_x})$ 与 $I_{\mathrm{DU}} = \sum_{\mathrm{D}_x \in \Phi'_{\mathrm{D}}} a_{\mathrm{D}_x} P_{\mathrm{D}_x} G^{\mathrm{AC}}_{\mathrm{D}_x} G_{\mathrm{U}} h_{\mathrm{D}_x} \ell_{q_x}(r_{\mathrm{D}_x})$. 引入上述激活变量, 等价于对干扰点过程 Φ'_{B}、Φ'_{D} 分别以概率 $p_{\mathrm{B,act}}$、$p_{\mathrm{D,act}}$ 进行独立稀释. 对于回传链路, 当其两端的定向天线均采用极窄波束时, 旁瓣泄漏增益极低, 其他 BS 与 UAV 的干扰可以忽略, 即 $I_{\mathrm{BD}} \approx 0$、$I_{\mathrm{DD}} \approx 0$, 故 $\mathrm{SINR}_{\mathrm{BD}}$ 不受同频激活变量的影响. 由此, 相应 SINR 分别为

$$\mathrm{SINR}_{\mathrm{BU}} = \frac{P^{\mathrm{Rx}}_{\mathrm{B}_0}}{I_{\mathrm{BU}} + I_{\mathrm{DU}} + P^{\mathrm{Rx}}_{\mathrm{D}_0} + N_0}, \tag{11}$$

$$\mathrm{SINR}_{\mathrm{BD}} = \frac{P^{\mathrm{Rx}}_{\mathrm{B}_0\mathrm{D}_0}}{I_{\mathrm{BD}} + I_{\mathrm{DD}} + N_0}, \tag{12}$$

$$\mathrm{SINR}_{\mathrm{DU}} = \frac{P^{\mathrm{Rx}}_{\mathrm{D}_0}}{I_{\mathrm{BU}} + I_{\mathrm{DU}} + P^{\mathrm{Rx}}_{\mathrm{B}_0} + N_0}, \tag{13}$$

其中, N_0 为加性高斯白噪声功率.

2.1.4　优化目标

完成系统建模与干扰分析后, 本文针对 UAV 蜂窝网络核心性能指标开展优化设计, 为后续生成式 AI 代理灵活适配提供理论支撑, 重点考察两类核心性能指标, 如下所示.

(1) 可达速率. 基于选择合并后的最优 SINR, UE 在 AF 与 DF 协议下的可达速率分别为

$$R^{\mathrm{AF}}(\boldsymbol{\xi}) = \log_2\big(1+\mathrm{SINR}^{\mathrm{AF}}(\boldsymbol{\xi})\big), \tag{14}$$

$$R^{\mathrm{DF}}(\boldsymbol{\xi}) = \log_2\big(1+\mathrm{SINR}^{\mathrm{DF}}(\boldsymbol{\xi})\big), \tag{15}$$

其中, $\boldsymbol{\xi} = (P_{\mathrm{B}}^{\mathrm{AC}}, P_{\mathrm{B}}^{\mathrm{BH}}, P_{\mathrm{D}})$ 表示功率向量.

(2) 能量效率. 能量效率定义为系统可达速率与总功耗之比, 总功耗包括 BS 发射功率 $P_{\mathrm{B}}^{\mathrm{AC}}$ 和 $P_{\mathrm{B}}^{\mathrm{BH}}$、UAV 发射功率 P_{D} 及 UAV 悬停功率 P_{hov}. AF 与 DF 协议下能量效率分别为

$$\mathrm{EE}^{\mathrm{AF}}(\boldsymbol{\xi}) = \frac{R^{\mathrm{AF}}(\boldsymbol{\xi})}{\mu_{\mathrm{B}}\left(P_{\mathrm{B}}^{\mathrm{AC}}+P_{\mathrm{B}}^{\mathrm{BH}}\right)+\mu_{\mathrm{D}}P_{\mathrm{D}}+P_{\mathrm{hov}}}, \tag{16}$$

$$\mathrm{EE}^{\mathrm{DF}}(\boldsymbol{\xi}) = \frac{R^{\mathrm{DF}}(\boldsymbol{\xi})}{\mu_{\mathrm{B}}\left(P_{\mathrm{B}}^{\mathrm{AC}}+P_{\mathrm{B}}^{\mathrm{BH}}\right)+\mu_{\mathrm{D}}P_{\mathrm{D}}+P_{\mathrm{hov}}}, \tag{17}$$

其中, μ_{B}、μ_{D} 分别为 BS、UAV 功放效率因子的倒数; $P_{\mathrm{hov}} = \sqrt{(M_{\mathrm{UAV}}g)^3/(2\pi n_{\mathrm{r}}\rho_{\mathrm{air}}r_{\mathrm{r}}^2)}$ 为 UAV 悬停机械功率, M_{UAV} 为 UAV 总质量, g 为重力加速度, n_{r} 为旋翼数, r_{r} 为旋翼半径, ρ_{air} 为海平面空气密度.

可见, UAV 蜂窝网络建模与优化需多维度协同, 手动建模高度依赖领域专家经验, 且场景动态、参数强耦合, 易产生建模偏差.

2.2　知识库构建

生成式 AI 代理若无可检索的领域知识库, 建模只能依赖 LLM 的参数化知识, 正确性与可溯源性均无保证. 为此, 本文按 2.1 小节的四个维度构建 UAV 蜂窝网络领域知识库. 每个维度对应一个数据块, 并按该维度的两个选项分为两个子数据块, 即直连模式与中继模式、同构部署与异构部署、AF 协议与 DF 协议、可达速率与能量效率, 共 4 个数据块、8 个子数据块. 知识块是知识库的基本单元, 每个知识块对应一条可独立引用的建模事实, 由一条公式或结论及其符号定义、成立条件与适用域构成, 长度控制在 120 ∼ 400 个词元, 过长会使单个知识块覆盖多个建模要素, 过短则会割裂公式与其成立条件. 知识块取自同行评审文献、3GPP TR 36.777 [19] 中的空对地信道模型、TR 38.901 [20] 中的天线方向图模型, 以及 2.1 小节各维度的推导结论. 入库时按统一模板填写结论陈述、符号定义、成立条件、适用域与来源标注, 缺项者不予入库, 含公式的知识块另经符号计算工具作量纲与符号一致性检查. 最终知识库共含 1090 个知识块, 平均长度 218 个词元.

2.3　语义路由器

代理要生成准确的系统模型, 首先要找到正确的子数据块, 再从中检索知识. 语义路由器负责把 UE 的自然语言描述转化为子数据块调用指令, 采用两层结构. 第一层把 UE 描述映射到数据块, 第二层在该数据块内确定子数据块, 两层都按语义距离判别. 将 UE 查询记为 $q_{\mathrm{usr}} \in \mathcal{Q}_{\mathrm{usr}}$. 编码器 $\mathcal{E}(\cdot)$ 把查询、数据块标签与知识块编码为 ζ 维非零嵌入向量. 本文沿用文献 [10] 的做法, 采用预训练的 text-embedding-ada-002 模型生成嵌入向量, 并以余弦距离度量语义相似度

$$d_{\cos}(x,y) = 1 - \frac{\mathcal{E}(x)^{\mathrm{T}}\mathcal{E}(y)}{\|\mathcal{E}(x)\|_2\,\|\mathcal{E}(y)\|_2}, \tag{18}$$

数据块与子数据块的标签分别记为 b_i 与 s_j^i, 其中 $i \in \{1,2,3,4\}$、$j \in \{1,2\}$. 例如, b_3 = “中继协议”、s_1^3 = “AF”、s_2^3 = “DF”. 查询 q_{usr} 的两层路由结果为

$$i^\star = \underset{i\in\{1,2,3,4\}}{\arg\min}\ d_{\cos}(q_{\mathrm{usr}}, b_i), \qquad j^\star = \underset{j\in\{1,2\}}{\arg\min}\, d_{\cos}(q_{\mathrm{usr}}, s_j^{i^\star}), \tag{19}$$

即查询依次被路由到语义距离最小的数据块和子数据块.

2.4　检索增强生成

RAG 从目标子数据块中检索相关知识块, 作为辅助上下文供 LLM 生成建模答案 $m=(m_1,\dots,m_L)$, 本文采用序列级检索增强生成. 同一知识块 z 用于预测整条目标序列的全部词元, 只在完整序列层对知识块作边缘化, 这与在每个词元处重新求和的词元级方案不同. 设目标子数据块非空且 $1\leqslant K\leqslant|\mathcal{K}_{i^\star,j^\star}|$, 知识块的检索分数与 Top-$K$ 集合定义为

$$s_\eta(q_{\mathrm{usr}},z)=1-d_{\cos}(q_{\mathrm{usr}},z),\qquad \mathcal{Z}_K(q_{\mathrm{usr}})=\text{Top-}K_{z\in\mathcal{K}_{i^\star,j^\star}}\, s_\eta(q_{\mathrm{usr}},z). \tag{20}$$

令 $\tau_{\mathrm{r}}>0$ 为检索温度. 在 Top-K 集合内重新归一化, 得到截断检索概率

$$p_\eta^{(K)}(z\mid q_{\mathrm{usr}})=\frac{\exp(s_\eta(q_{\mathrm{usr}},z)/\tau_{\mathrm{r}})}{\displaystyle\sum_{z'\in\mathcal{Z}_K(q_{\mathrm{usr}})}\exp(s_\eta(q_{\mathrm{usr}},z')/\tau_{\mathrm{r}})},\quad z\in\mathcal{Z}_K(q_{\mathrm{usr}}), \tag{21}$$

对 $z\notin\mathcal{Z}_K(q_{\mathrm{usr}})$ 令 $p_\eta^{(K)}(z\mid q_{\mathrm{usr}})=0$. 由此, 条件生成概率为

$$p_{\mathrm{Seq}}^{(K)}(m\mid q_{\mathrm{usr}})=\sum_{z\in\mathcal{Z}_K(q_{\mathrm{usr}})}p_\eta^{(K)}(z\mid q_{\mathrm{usr}})\prod_{\ell=1}^{L}p_\theta(m_\ell\mid q_{\mathrm{usr}},z,m_{<\ell}). \tag{22}$$

2.5　问题构建

基于生成式 AI 代理框架, UE 可根据场景需求与性能目标, 灵活定制 UAV 蜂窝网络约束优化问题. 四个建模维度各有两个选项, 代理按 UE 给定的选项自动生成对应的问题实例. 本小节以中继模式、异构部署、DF 协议与可达速率最大化为例, 定义优化变量为 $\boldsymbol{\xi}=(P_{\mathrm{B}}^{\mathrm{AC}},P_{\mathrm{B}}^{\mathrm{BH}},P_{\mathrm{D}})$, 所构建的优化问题如下:

$$\begin{aligned}
\max_{\boldsymbol{\xi}}\quad & R^{\mathrm{DF}}(\boldsymbol{\xi}) && \text{(23a)}\\
\text{s.t.}\quad & P_{\mathrm{B}}^{\mathrm{AC}}+P_{\mathrm{B}}^{\mathrm{BH}}\leqslant P_{\mathrm{B}}^{\max}, && \text{(23b)}\\
& P_{\mathrm{D}}^{\min}\leqslant P_{\mathrm{D}}\leqslant P_{\mathrm{D}}^{\max}, && \text{(23c)}\\
& \mathrm{SINR}_{\mathrm{BD}}\big(P_{\mathrm{B}}^{\mathrm{BH}},P_{\mathrm{B}}^{\mathrm{AC}},P_{\mathrm{D}}\big)\geqslant\gamma_{\mathrm{th}}, && \text{(23d)}\\
& P_{\mathrm{B}}^{\mathrm{AC}}\geqslant P_{\mathrm{B}}^{\mathrm{AC,min}}, && \text{(23e)}\\
& P_{\mathrm{B}}^{\mathrm{BH}}\geqslant 0, && \text{(23f)}
\end{aligned}$$

其中, 式 (23b) 为 BS 总功率约束, $P_{\mathrm{B}}^{\max}$ 为下倾全向接入天线与上倾定向回传天线共享的射频前端最大输出功率; 式 (23c) 为 UAV 发射功率范围约束; 式 (23d) 为 DF 译码可行性约束, 要求回传链路接收 SINR 不低于译码门限 γ_{th}, 即 UAV 须成功译码服务 BS 的回传信号后方可转发; 式 (23e) 为 BS 地面覆盖保障约束, 要求接入天线发射功率不低于 $P_{\mathrm{B}}^{\mathrm{AC,min}}$, 与式 (23b) 共同构成回传优化与地面覆盖之间的功率权衡; 式 (23f) 为 BS 回传天线发射功率非负约束.

以能量效率为优化目标时, 约束集合不变, 仅将目标函数改为速率与总功耗之比, 相应的优化问题为

$$\begin{aligned}
\max_{\boldsymbol{\xi}}\quad & \mathrm{EE}^{\mathrm{DF}}(\boldsymbol{\xi})\\
\text{s.t.}\quad & \text{(23b)}\sim\text{(23f)}.
\end{aligned} \tag{24}$$

2.6　规划、工具调用与反馈修正

2.3 和 2.4 小节的路由与检索为开环流程, 生成结果无论对错都直接输出. 本小节引入规划、工具调用与反馈修正三个环节, 使错误可被发现并纠正.

(1) 规划. 传播场景决定 UE 经直连还是中继接入, 即参与传输的链路. 部署方式随后决定 UAV 的空间分布与悬停高度区间, 进而决定服务 UAV. 中继协议决定端到端 SINR 的形式, 以及是否引入译码可行性约束. 优化目标则由前三者的表达式组合而成. 因此, 目标函数在其余三项确定之前无法写出. 据此, 将复合请求 q_{req} 分解为四个原子查询, 并按依赖关系组织为有向无环图

$$\mathcal{G}_{\mathrm{plan}} = \Big(\{b_1, b_2, b_3, b_4\},\ \big\{(b_1, b_2), (b_2, b_3), (b_3, b_4)\big\}\Big), \tag{25}$$

其拓扑序即为执行顺序

$$\sigma = \mathrm{TopoSort}\big(\mathcal{G}_{\mathrm{plan}}\big) = (b_1, b_2, b_3, b_4). \tag{26}$$

第 i 个原子查询由两部分拼接而成. 一是该维度的查询模板 $\kappa_{\sigma(i)}$, 二是前序各维度已通过校验的答案 $m^{\star}_{\sigma(j)}$ $(j < i)$, 即

$$q^{\mathrm{at}}_{\sigma(i)} = \kappa_{\sigma(i)} \oplus \Big(\oplus_{j<i} m^{\star}_{\sigma(j)}\Big). \tag{27}$$

路由与检索时, 该原子查询即为式 (19)∼(22) 中的 q_{usr}. 规划器只负责执行顺序与上下文前缀. 原子查询属于哪个数据块与子数据块, 仍由两层语义路由判断.

(2) 工具调用. 第 i 个维度在第 t 轮得到候选答案 $m^{(t)}_{\sigma(i)}$ 后, 代理将其交给三个外部工具检查. 量纲校验判断量纲是否齐次, 可行性检查判断约束是否可行, 求解器接口判断能否实例化. 结果分别记为二值变量 $d^{(t)}_i$、$f^{(t)}_i$ 与 $s^{(t)}_i$, 校验通过时取 1. 三项均为 1 时才算校验通过, 即

$$\chi^{(t)}_i = \mathbf{1}\big\{d^{(t)}_i = 1\big\} \cdot \mathbf{1}\big\{f^{(t)}_i = 1\big\} \cdot \mathbf{1}\big\{s^{(t)}_i = 1\big\}. \tag{28}$$

检查均由程序按符号计算与数值判定完成, 结果可复现. 检查逐项进行, 失败时可指出错误的建模要素.

(3) 反馈修正. 校验未通过时, 依据失败类型 $\nu^{(t)}_i$ 选择修正算子

$$\mathfrak{R}^{(t)} = \begin{cases} \mathfrak{R}_{\mathrm{gen}}, & \nu^{(t)}_i = \text{量纲或表达式错误}, \\ \mathfrak{R}_{\mathrm{ret}}, & \nu^{(t)}_i = \text{证据缺失}, \\ \mathfrak{R}_{\mathrm{route}}, & \nu^{(t)}_i = \text{维度误判}. \end{cases} \tag{29}$$

三者分别对生成器、检索器与语义路由器执行重生成、重检索与重路由. 每轮修正都会用到此前未用过的知识块或未试过的子数据块, 各轮不会重复. 修正最多进行 $T_{\max}$ 轮, 流程必然终止. 达到上限仍未通过时, 输出当前最优候选并标明失败. 由此, 单条复合请求的 LLM 调用次数为 $C_{\mathrm{llm}} = 4 + \sum_{i=1}^{4} t_i$. 其中 t_i 为第 i 个维度的修正轮数, $0 \leqslant t_i \leqslant T_{\max}$. 四个维度处理完后, 代理汇总通过校验的子答案. 随后检查其符号、量纲、链路时序与约束是否相互一致, 并经求解器接口实例化为优化问题. 算法 1 给出了完整流程.

3　基于 MoE 的 PPO 算法

由于式 (23) 的优化变量相互耦合, 常规方法难以高效求得最优解, 这已成为制约系统性能提升的瓶颈之一. 为此, 本节提出 MoE-PPO 算法, 求解生成式 AI 代理所生成的问题, 即中继模式、异构部署、DF 协议下的非凸优化问题.

3.1　MoE 辅助的 PPO 框架

MoE 是一种基于中央门控网络的集成学习架构, 由 I 个独立专家网络 $(E_1, E_2, \ldots, E_I)$ 与 1 个中央门控网络 G_σ 组成. 每个专家网络拥有独立可学习参数且同步运行, 可独立处理输入并输出相同维度的特征; 门控网络依据输入特征的语义与任务特性生成权重向量, 将其分配至最适配的专家网络, 实现专家的选择性激活. MoE 模型的核心函数可表示为

$$y = \sum_{i=1}^{I} G_\sigma(x)_i E_i(x), \tag{30}$$

算法 1 生成式 AI 代理.

输入: UE 请求 q_{req}, 嵌入模型 $\mathcal{E}$, 知识库 $\mathcal{K}$, 生成器, 检索数 K, 迭代上限 $T_{\max}$.
输出: 可直接实例化的 UAV 蜂窝网络约束优化问题.

按式 (25) 将 q_{req} 的四个建模维度组织为有向无环图, 并按式 (26) 取其拓扑序 σ;
for 拓扑序中的每个维度 $\sigma(i)$ **do**
　按式 (27) 以前序维度已通过校验的答案为前缀, 构造原子查询 $q^{\text{at}}_{\sigma(i)}$; 置 $t \leftarrow 0$;
　repeat
　　计算 $\mathcal{E}(q^{\text{at}}_{\sigma(i)})$, 并按式 (19) 得到 $(i^\star, j^\star)$;
　　按式 (20) 计算检索分数并取 $\mathcal{Z}_K$, 按式 (21) 计算截断概率;
　　按式 (22) 生成候选答案 $m^{(t)}_{\sigma(i)}$;
　　调用三个确定性工具, 按式 (28) 给出校验指示 $\chi^{(t)}_{\sigma(i)}$;
　　若未通过, 则按式 (29) 依失败类型选择重生成、重检索或重路由; $t \leftarrow t+1$;
　until $\chi^{(t)}_{\sigma(i)} = 1$ 或 $t > T_{\max}$
　若达上限仍未通过校验, 则输出当前最优候选并标记为失败;
end for
汇总全部已通过校验的子答案, 校验符号、量纲、链路时序与约束的一致性, 经求解器接口实例化为优化问题.

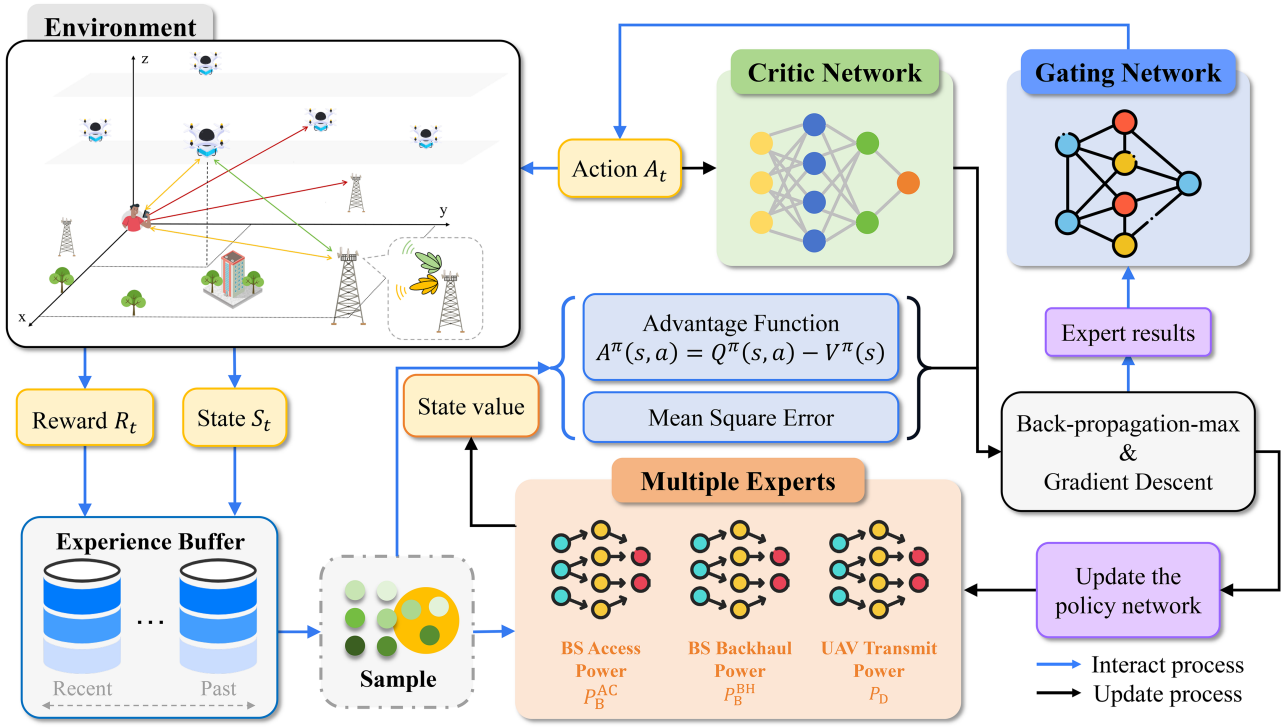

图 3　(网络版彩图) MoE-PPO 算法框架.

Figure 3　(Color online) Framework of the MoE-PPO algorithm.

其中, $E_i(x)$ 为第 i 个专家网络对输入 x 的输出, 门控权重 $G_\sigma(x)_i = \text{softmax}(x \cdot W_g)_i$, W_g 为权重矩阵.

本文所提 MoE-PPO 框架如图 3 所示, 假设每个专家输出的候选动作 a_t^i 都是一组完整的功率配置, 与动作空间同维. 演员网络集成多个专家策略, 负责输出优化动作; 评论家网络评估状态值函数, 为演员网络的策略迭代提供指导. 记第 i 个专家的参数为 ψ_i, 门控网络的参数为 θ_g, 演员网络的全部参数为 $\theta_A = \{\{\psi_i\}_{i=1}^I, \theta_g\}$; 评论家网络的参数记为 θ_C, 其给出的优势函数记为 $A_{\theta_C}(\cdot)$.

(1) 演员网络. 演员网络采用 MoE 设计, 通过多专家协同决策提升复杂场景适配能力, 其策略输出定义为

$$\pi_{\theta_A}(a_t|s_t)=\sum_{i=1}^{I} w_{\theta_i}(s_t)\,\pi_{\psi_i}(a_t|s_t), \tag{31}$$

其中, π_{ψ_i} 为单个专家策略, $w_{\theta_i}(s_t)$ 为状态 s_t 下第 i 个专家的动态权重, 由门控网络统一调度管理.

(2) 门控函数. 门控函数用于根据当前系统状态 s_t 动态选择并加权聚合最优专家策略. 然而, 门控机制涉及离散空间下的专家选择过程, 标准反向传播算法难以直接对离散决策操作进行梯度传递, 导致演员网络的梯度估计面临较大挑战. 为此, 本文引入一种基于优势函数的反向传播最大化策略梯度方法, 实现门控网络与专家网络参数的联合优化. 具体而言, 在给定状态 s_t 下, 对每个专家策略 $\pi_{\psi_i}(\cdot|s_t)$ 采样得到候选动作 a_t^i 并计算其优势值 $A_{\theta_C}(s_t,a_t^i)$, 该值反映在状态 s_t 下执行动作 a_t^i 所带来的相对收益, 优势值最大的专家策略即为当前状态下的最优策略. 在此基础上, 演员网络针对最优专家进行梯度聚焦更新, 与专家 i 对应的梯度可表示为

$$\mathrm{grad}_{\psi_i}=\delta_i\,\nabla_{\psi_i}\log\pi_{\psi_i}(a_t^i|s_t), \tag{32}$$

其中, 示性项在专家 i 的候选动作取得最大优势值时为 1, 否则为 0, δ_i 为该示性项与专家 i 当前门控权重之差.

(3) 目标函数. 为适配多专家协同决策, 对目标函数进行调整, 如下所示:

$$J(\theta_A)=\mathbb{E}_t\Big[\rho_t(\theta_A)\,A_t^{\pi_{\theta_A^{\mathrm{old}}}}\Big], \tag{33}$$

其中, $\rho_t(\theta_A)=\pi_{\theta_A}(a_t|s_t)/\pi_{\theta_A^{\mathrm{old}}}(a_t|s_t)$ 为当前策略与旧策略的动作概率比; $A_t^{\pi_{\theta_A^{\mathrm{old}}}}$ 为优势函数, 通过聚合各专家对奖励的加权贡献, 凸显每个专家在集体策略中的作用, 如下所示:

$$A_t^{\pi_{\theta_A^{\mathrm{old}}}}=r(s_t,a_t)+\gamma V_{\theta_C^{\mathrm{old}}}(s_{t+1})-V_{\theta_C^{\mathrm{old}}}(s_t). \tag{34}$$

基于信任域策略优化的单调改进理论, 为保证策略迭代稳定且收敛, 引入 KL 散度约束限制新旧策略的差异, 即满足

$$\mathbb{E}_t\Big[\mathrm{KL}\Big(\pi_{\theta_A}(a_t|s_t)\,\|\,\pi_{\theta_A^{\mathrm{old}}}(a_t|s_t)\Big)\Big]\leqslant\varepsilon, \tag{35}$$

其中, ε 为控制信任域范围的正超参数. 为降低计算复杂度, 满足式 (35) 约束的目标函数 $J(\theta_A)$ 可近似为裁剪代理目标函数, 如下所示:

$$J^{\mathrm{CLIP}}(\theta_A)=\mathbb{E}\Big[\min\Big(\rho_t(\theta_A)\,A_t^{\pi_{\theta_A^{\mathrm{old}}}},\ \mathrm{clip}(\rho_t(\theta_A),1-\varepsilon,1+\varepsilon)\,A_t^{\pi_{\theta_A^{\mathrm{old}}}}\Big)\Big]. \tag{36}$$

(4) 更新机制. 演员网络与评论家网络的参数均由 Adam 优化器通过环境采样进行梯度更新, 如下所示:

$$\theta_A=\theta_A^{\mathrm{old}}+\alpha_A\frac{1}{B}\sum_{t=1}^{B}\Big(\nabla_{\theta_A}\tilde{J}_t^{\mathrm{CLIP}}(\theta_A)\Big), \tag{37}$$

$$\theta_C=\theta_C^{\mathrm{old}}-\alpha_C\frac{1}{B}\sum_{t=1}^{B}\nabla_{\theta_C}\big(V_{\theta_C}(s_t)-V_{\mathrm{tar}}(s_t)\big)^2, \tag{38}$$

其中, $\tilde{J}_t^{\mathrm{CLIP}}(\theta_A)$ 为裁剪代理目标函数的实际值, α_A、α_C 分别为演员网络与评论家网络的学习率; 基于 n 步时序差分方法, 状态 s_t 的目标值函数 $V_{\mathrm{tar}}(s_t)$ 为

$$V_{\mathrm{tar}}(s_t)=\sum_{k=0}^{n}\gamma^k r(s_{t+k},a_{t+k})+\gamma^{n+1}V_{\theta_C^{\mathrm{old}}}(s_{t+n+1}), \tag{39}$$

其中, γ 为奖励折扣因子. 综上, 算法 2 给出了所提 MoE-PPO 算法的完整流程.

算法 2 MoE-PPO 算法.

输入: 训练轮次数, 探索步数, 更新轮次数, 学习率, 信道状态信息, 专家数量 I.
输出: 最优功率配置 $\boldsymbol{\xi}^* = \big((P_{\mathrm{B}}^{\mathrm{AC}})^*, (P_{\mathrm{B}}^{\mathrm{BH}})^*, P_{\mathrm{D}}^*\big)$.

初始化环境, 观测初始状态 s_0;
for 每次迭代 **do**
  for 每个探索步 t **do**
    每个专家 i 基于子策略 π_{ψ_i} 独立采样与动作空间同维的候选动作 a_t^i, 并按式 (34) 计算对应优势值 $A_{\theta_C}(s_t, a_t^i)$;
    门控网络输出权重 $w_{\theta_i}(s_t)$, 以概率 $w_{\theta_i}(s_t)$ 选取对应专家动作, 按式 (31) 聚合得最终动作 $a_t = \{P_{\mathrm{B}}^{\mathrm{AC}}, P_{\mathrm{B}}^{\mathrm{BH}}, P_{\mathrm{D}}\}$;
    按式 (40) 获取当前奖励 r_t, 观测下一状态 s_{t+1};
    将转移样本 (s_t, a_t, r_t, s_{t+1}) 存入经验池 $\mathcal{D}$;
  end for
  for 每个更新步 **do**
    从经验池 $\mathcal{D}$ 中采样 B 个小批量转移样本;
    按式 (37) 与 (38) 更新演员网络 θ_A 与评论家网络 θ_C;
    按式 (34) 计算所有专家的优势函数, 并按式 (32) 更新门控参数;
    更新状态 $s_t \leftarrow s_{t+1}$;
  end for
end for

3.2　决策过程

本文将决策过程建模为马尔可夫决策过程, 通过三元组 $\langle \mathcal{S}, \mathcal{A}, \mathcal{R} \rangle$ 对框架进行形式化定义, $\mathcal{S}$、$\mathcal{A}$ 与 $\mathcal{R}$ 分别对应状态空间、动作空间与奖励函数. 各分量的具体设计如下.

(1) 动作空间. 动作空间的设计核心是实现优化变量的精准输出与约束满足, 对每个变量采用双曲正切函数进行映射, 实现动作输出与约束可行域的匹配, 具体设计如下: (i) 对于 BS 接入天线发射功率 $P_{\mathrm{B}}^{\mathrm{AC}}$, 为满足约束式 (23e) 与 (23b), 映射关系为 $P_{\mathrm{B}}^{\mathrm{AC}} = 0.5(P_{\mathrm{B}}^{\max} + P_{\mathrm{B}}^{\mathrm{AC,min}}) + 0.5(P_{\mathrm{B}}^{\max} - P_{\mathrm{B}}^{\mathrm{AC,min}}) \tanh(x_{P_{\mathrm{B}}^{\mathrm{AC}}}) \in [P_{\mathrm{B}}^{\mathrm{AC,min}}, P_{\mathrm{B}}^{\max}]$, $x_{P_{\mathrm{B}}^{\mathrm{AC}}}$ 为对应神经网络激活函数的输出值; (ii) 对于 BS 回传天线发射功率 $P_{\mathrm{B}}^{\mathrm{BH}}$, 其可行域由约束式 (23b) 隐式确定为 $[0, P_{\mathrm{B}}^{\max} - P_{\mathrm{B}}^{\mathrm{AC}}]$, 映射关系为 $P_{\mathrm{B}}^{\mathrm{BH}} = 0.5(P_{\mathrm{B}}^{\max} - P_{\mathrm{B}}^{\mathrm{AC}})(\tanh(x_{P_{\mathrm{B}}^{\mathrm{BH}}}) + 1)$, 该设计从结构上保约束式 (23b) 与 (23e) 始终得到满足, 无需额外迭代校验; (iii) 对于 UAV 发射功率 P_{D}, 为满足约束式 (23c), 映射关系为 $P_{\mathrm{D}} = 0.5(P_{\mathrm{D}}^{\max} + P_{\mathrm{D}}^{\min}) + 0.5(P_{\mathrm{D}}^{\max} - P_{\mathrm{D}}^{\min}) \tanh(x_{P_{\mathrm{D}}}) \in [P_{\mathrm{D}}^{\min}, P_{\mathrm{D}}^{\max}]$. 综上, 动作空间定义为 $\mathcal{A} = \{P_{\mathrm{B}}^{\mathrm{AC}}, P_{\mathrm{B}}^{\mathrm{BH}}, P_{\mathrm{D}}\}$, 其基数为 3, 与专家数 I 一致.

(2) 状态空间. 状态空间用于刻画决策时刻的系统运行状态, 由 UE 当前信息向量 $\mathbf{u}$、上一时间步所选动作向量 $\mathbf{a}$ 及即时奖励 r 三部分构成: (i) $\mathbf{u} = \{\mathrm{SINR}_{\mathrm{BU}}, \mathrm{SINR}_{\mathrm{BD}}, \mathrm{SINR}_{\mathrm{DU}}\}$, 分别对应上一时间步直连链路、回传链路与空地接入链路的 SINR, 直接反映链路传输效能; (ii) 动作向量 $\mathbf{a}$ 取自动作空间 $\mathcal{A}$, 记录上一时间步执行的优化策略, 为当前策略迭代提供历史参考; (iii) 即时奖励 r 由当前状态与所选动作共同决定. 综上, 状态空间定义为 $\mathcal{S} = \{\mathbf{u}, \mathbf{a}, r\}$, 其基数为 $|\mathcal{S}| = |\mathbf{u}| + 3 + 1$.

(3) 奖励函数. 奖励函数遵循直接奖励项与约束惩罚项的复合结构, 在最大化目标函数的同时满足问题式 (23) 的各项约束, 如下所示:

$$r = R^{\mathrm{DF}}(\boldsymbol{\xi}) \times \underbrace{\Omega_{P_{\mathrm{B}}} \times \Omega_{P_{\mathrm{D}}} \times \Omega_{\mathrm{BD}}}_{\text{惩罚项}}, \tag{40}$$

其中, $\Omega_{P_{\mathrm{B}}}$ 对应约束式 (23b) 与 (23e), 当 $P_{\mathrm{B}}^{\mathrm{AC}} + P_{\mathrm{B}}^{\mathrm{BH}} \leqslant P_{\mathrm{B}}^{\max}$ 且 $P_{\mathrm{B}}^{\mathrm{AC}} \geqslant P_{\mathrm{B}}^{\mathrm{AC,min}}$ 时取值为 1, 否则为 0; $\Omega_{P_{\mathrm{D}}}$ 对应约束式 (23c), 当 $P_{\mathrm{D}} \in [P_{\mathrm{D}}^{\min}, P_{\mathrm{D}}^{\max}]$ 时取值为 1, 否则为 0; Ω_{BD} 对应约束式 (23d), 当 $\mathrm{SINR}_{\mathrm{BD}} \geqslant \gamma_{\mathrm{th}}$ 时取值为 1, 否则为 0.

4　仿真分析

4.1　仿真设置

本小节通过数值仿真验证生成式 AI 代理的建模正确性与 MoE-PPO 算法的求解性能. 系统参数设置如下. (i) 功率与部署参数: $P_{\mathrm{B}}^{\max}=20$ W, $P_{\mathrm{D}}^{\max}=2$ W, $P_{\mathrm{D}}^{\min}=0$ W, $P_{\mathrm{B}}^{\mathrm{AC,min}}=5$ W, $h_{\mathrm{B}}=20$ m, $\lambda_{\mathrm{B}}=10^{-6}$, $h_{\mathrm{D,high}}^{1}=300$ m, $h_{\mathrm{D,high}}^{2}=500$ m, $\lambda_{\mathrm{D,high}}=5\times10^{-5}$, $h_{\mathrm{D,low}}^{1}=50$ m, $h_{\mathrm{D,low}}^{2}=150$ m, $\lambda_{\mathrm{D,low}}=10^{-4}$, $p_{\mathrm{B,act}}=0.2$, $p_{\mathrm{D,act}}=0.1$; (ii) 天线参数: $A_{\mathrm{m}}=30$ dB, $N_{\mathrm{B}}=8$, $d_{\mathrm{B}}=0.5\lambda$, $\theta_{\mathrm{B}}=100^{\circ}$, $\theta_{0}=30^{\circ}$, $\theta_{3\mathrm{dB}}^{\mathrm{AC}}=65^{\circ}$, $\mathrm{SLA_V}=30$ dB, $G_{\max}=8$ dBi, $G_{\mathrm{U}}=0$ dBi; (iii) 信道与协议参数: $\alpha_{\mathrm{L}}=2.5$, $\alpha_{\mathrm{N}}=4$, $m=1$, $N_0=10^{-8}$ W, $\gamma_{\mathrm{th}}=0$ dB; (iv) UAV 机体参数: $M_{\mathrm{UAV}}=1.5$ kg, $n_{\mathrm{r}}=4$, $r_{\mathrm{r}}=0.15$ m, $g=9.8$ m/s^2, $\rho_{\mathrm{air}}=1.225$ kg/m^3, $\mu_{\mathrm{B}}=1.2$, $\mu_{\mathrm{D}}=1.1$. 为考察城区密集程度对系统性能的影响, 仿真覆盖郊区 ($c_1=4.88$, $c_2=0.43$, η_{L}=0.1 dB, $\eta_{\mathrm{N}}=21$ dB)、城区 ($c_1=9.61$, $c_2=0.16$, $\eta_{\mathrm{L}}=1$ dB, $\eta_{\mathrm{N}}=20$ dB)、密集城区 ($c_1=12.08$, $c_2=0.11$, $\eta_{\mathrm{L}}=1.6$ dB, $\eta_{\mathrm{N}}=23$ dB) 与超密集城区 ($c_1=27.23$, $c_2=0.08$, $\eta_{\mathrm{L}}=2.3$ dB, $\eta_{\mathrm{N}}=34$ dB) 四类典型场景.

生成式 AI 代理的推理引擎采用 GPT-4o, 解码温度为 0.2, top-p 为 0.9, 单次输出上限为 2048 个词元, 并设固定的 seed 参数以提高仿真的可重复性; 嵌入模型采用 text-embedding-ada-002, 输出维度 $\zeta=1536$. 向量检索由 FAISS 内积索引实现, 嵌入向量经 ℓ_2 归一化后检索, 检索温度取 $\tau_{\mathrm{r}}=0.1$, 检索数取 $K=8$, 检索与一致性校验在本地执行. MoE-PPO 算法的超参数如下: 神经网络层数为 4 层, $\varepsilon=0.2$, 优化器选用 Adam, 激活函数组合为 ReLU 与 tanh, 演员网络总参数量约 1.34×10^5, 其中单个专家的隐藏层宽度为 148、门控网络与评论家网络的隐藏层宽度为 256, 学习率取 3×10^{-4}, 折扣因子 $\gamma=0.95$, n 步时序差分的步数 $n=10$, 经验池容量 40960, 批处理大小 2048, 每回合含 100 个环境步. 值得注意的是, 本小节以可达速率为优化目标; 以能量效率为目标的优化问题式 (24), 可由同一代理生成并由同一求解器处理, 求解流程不变, 限于篇幅不再单独给出结果.

4.2　生成式 AI 代理性能评估

本小节按语义路由、知识检索与生成校验的顺序, 从三个方面评估代理. 评测集由四个建模维度的选项组合生成, 共 400 条复合测试查询. 每条查询按维度分解为 4 条原子查询, 共 1600 条, 每个子数据块 200 条, 其正确归属在生成时即已确定. 生成质量以 2.1 小节的人工建模结果为基准, 判定工具与 2.6 小节的校验器为同一实现.

如图 4 所示, 混淆矩阵中 S-Dir 与 S-Rel 为传播场景的直连与中继模式, D-Hom 与 D-Het 为同构与异构部署, P-AF 与 P-DF 为 AF 与 DF 协议, O-Rate 与 O-EE 为可达速率与能量效率目标. 第一层路由正确率为 $1270/1600=79.4\%$, 第一层正确的条件下第二层正确率为 $1108/1270=87.2\%$, 联合正确率为 $1108/1600=69.3\%$. 各子数据块的正确率介于 $0.55\sim0.80$, 以中继模式子块最低. 单项混淆比例最大为 0.14, 出现在同一维度内语义相近的子块之间, 以及中继模式与 DF 协议之间. 这说明路由错误集中于语义距离最近的相邻类, 与余弦相似度路由的判别机制相符. 按维度聚合, 第一层正确率依次为传播场景 0.72、部署方式 0.85、中继协议 0.79 与优化目标 0.82. 传播场景位于拓扑序首位, 没有前序上下文可用, 因此最低.

子数据块确定后, 检索质量决定生成所依据的知识. 如图 5 所示, 在 $K=8$ 处, 经两层路由的 Recall@K 为 0.91, 同等 K 下全库检索仅为 0.75, 相差 16 个百分点. nDCG@K 为 0.69、MRR@K 为 0.62, 表明目标知识块不仅能被召回, 而且在候选集中排序靠前. 移除语义路由后, 检索时延由 79 ms 增至 317 ms, 单条请求的输入词元由 6.4×10^3 增至 19.6×10^3. 这是因为两层分块把检索范围限定在目标子数据块内, 在提升召回率的同时压缩了上下文开销.

图 6 给出代理输出与人工建模的一致率, 整体为 79.5%. 移除语义路由后降至 75.6%, 移除 RAG 模块后降至 55.4%, 其中约束集合一项下降最为显著. 典型失效集中于量纲错误、遗漏约束与表达式缺项三类, 恰好对应知识块所承载的符号定义、成立条件与公式本身. 由此可见, RAG 决定建模的正确性, 语义路由决定检索的效率与召回, 二者职能不同, 不能相互替代.

一致率不反映结果是否有据可查. 如图 7 所示, 随修正轮次增加, 含校验器的代理源据支撑度由 22.0% 提

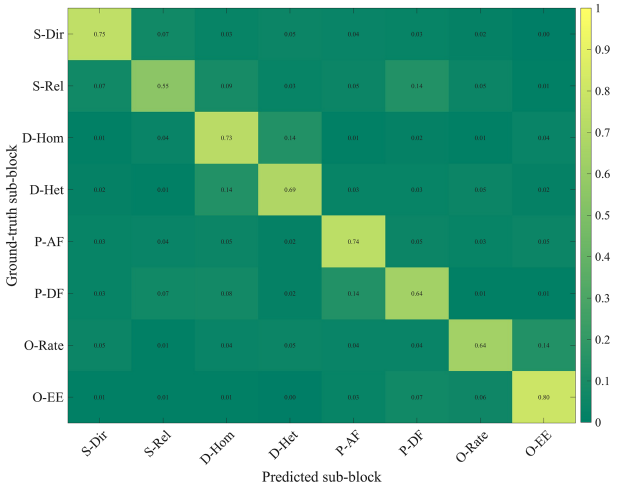

图 4　(网络版彩图) 两层语义路由的子数据块混淆矩阵.

Figure 4　(Color online) Confusion matrix of the two-level semantic router at the sub-block level.

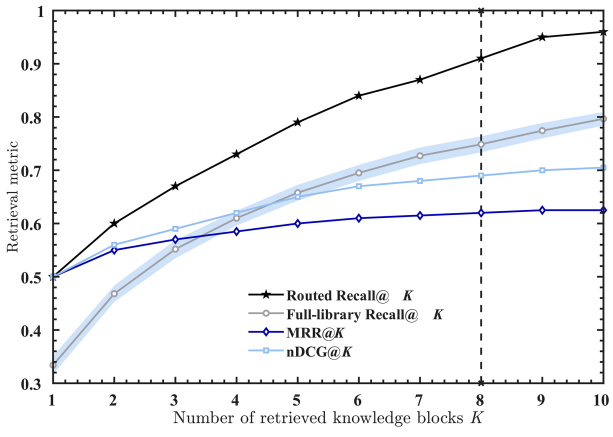

图 5　(网络版彩图) 检索质量随召回知识块数 K 的变化.

Figure 5　(Color online) Retrieval quality versus the number of retrieved knowledge blocks K.

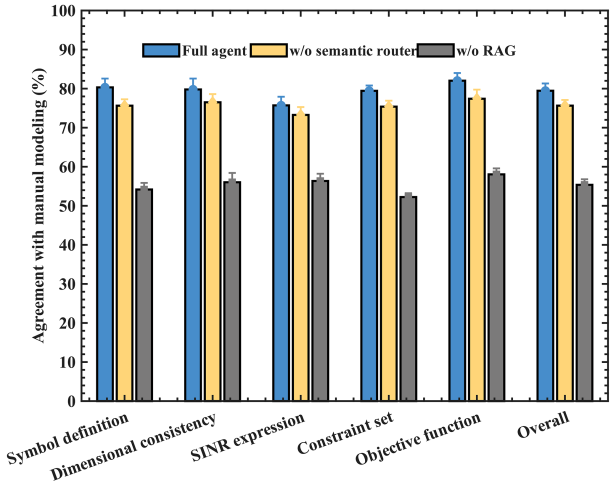

图 6　(网络版彩图) 自动建模与人工建模的一致率.

Figure 6　(Color online) Agreement between automated and manual modeling.

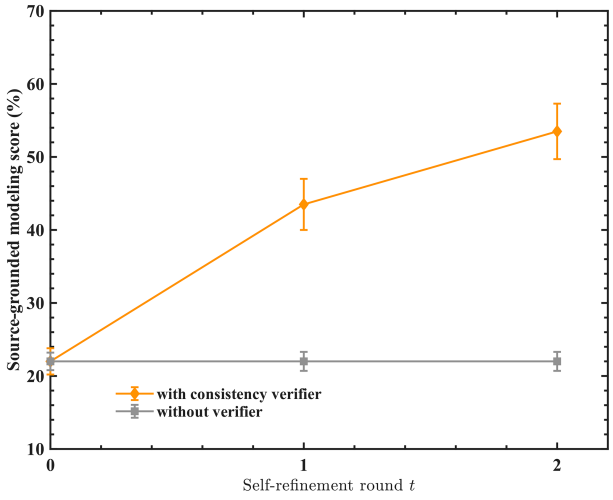

图 7　(网络版彩图) 反馈修正对源据支撑度的提升.

Figure 7　(Color online) Improvement of the source-grounded score by feedback correction.

升至 53.5%, 而不含校验器的开环组合始终为 22.0%. 由此可见, 源据支撑度的提升来自校验与反馈修正本身, 而非多轮重复采样.

4.3　MoE-PPO 算法性能评估

对比算法包括软演员 – 评论家 (soft actor-critic, SAC)、双延迟深度确定性策略梯度 (twin delayed deep deterministic policy gradient, TD3)、PPO、深度确定性策略梯度 (deep deterministic policy gradient, DDPG) 四种 DRL 算法, 以及贪婪与随机两种启发式基线. 图 8 给出各算法的平均收敛曲线, 表 1 给出其在独立最终测试集上的统计结果. 可以看出, 所提算法的可达速率为 (4.56 ± 0.12) bit/s/Hz, 较 PPO 算法提升 12.3%, 较贪婪与随机算法分别提升 37.8% 与 106.3%. 约束违反率方面, 所提算法为 0.8%, 低于 SAC 的 1.4%、TD3 的 2.0%、PPO 的 2.9% 与 DDPG 的 4.8%. 因此, 所提算法在速率与约束满足两方面均优于其余四种 DRL 算法. 贪婪与随机算法只在满足约束式 (23d) 的可行动作集合内取值, 因而不产生约束违反. 需要指出, 所提算法较 SAC 算法仅

表 1　各算法在独立最终测试集上的统计对比.

Table 1　Statistical comparison of the algorithms on the independent final test set.

Algorithm	Achievable rate (bit/s/Hz)	95% CI	Gain over PPO (%)	Welch t-test p	Cohen's d	Constraint violation (%)
MoE-PPO	4.56 ± 0.12	[4.47, 4.65]	+12.3	< 0.001	3.27	0.8
SAC	4.43 ± 0.14	[4.33, 4.53]	+9.1	< 0.001	2.29	1.4
TD3	4.24 ± 0.16	[4.13, 4.35]	+4.4	0.030	1.06	2.0
PPO	4.06 ± 0.18	[3.93, 4.19]	–	–	–	2.9
DDPG	3.66 ± 0.21	[3.51, 3.81]	−9.9	< 0.001	2.05	4.8
Greedy	3.31 ± 0.06	[3.27, 3.35]	−18.5	< 0.001	5.59	0.0
Random	2.21 ± 0.05	[2.17, 2.25]	−45.6	< 0.001	14.00	0.0

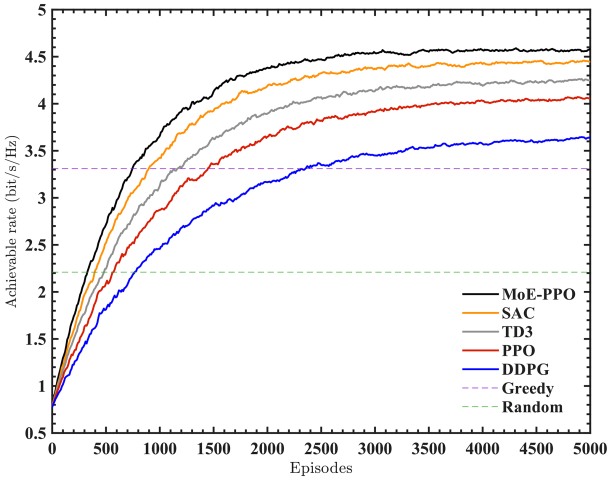

图 8　(网络版彩图) 各算法的平均收敛曲线.

Figure 8　(Color online) Average convergence curves of the compared algorithms.

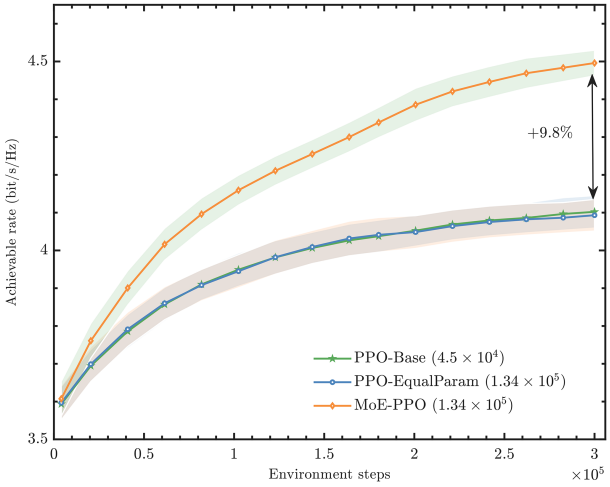

图 9　(网络版彩图) 参数量匹配对照下的训练收敛曲线.

Figure 9　(Color online) Training curves under parameter-matched comparison.

高 2.9%, 两者置信区间部分重叠, 经 Bonferroni 校正后差异不显著, 故在可达速率上尚不能确定优于 SAC 算法.

为区分增益的来源, 在同一工作点、评论家网络与超参数均相同的条件下, 依次考察参数量、专家数量与门控机制三个因素. 如图 9 所示, 所提算法由 3 个专家与 1 个门控网络构成, 可训练参数量 $N_\theta \approx 1.34 \times 10^5$. 等参数量 PPO 算法采用单一演员网络, 隐藏层宽度为 256, 参数量与之相同. 基础 PPO 算法同为单一演员网络, 隐藏层宽度缩至 $\lceil 256/\sqrt{3} \rceil = 148$, $N_\theta \approx 4.5 \times 10^4$. 三者收敛后的可达速率依次为 4.496、4.093 与 4.102 bit/s/Hz. 可以看出, 单一网络下参数量由 4.5×10^4 增至 1.34×10^5 未带来增益, 两者置信区间完全重叠; 而参数量相同时, 所提算法较等参数量 PPO 算法提升 9.8%, 两者置信区间不相交. 因此, 增益来自变量级的专家分工, 而非参数容量.

如图 10 所示, 考察 $I = 1, 2, 3, 4, 6$ 五种配置, 单专家隐藏层宽度按 $\lceil 256/\sqrt{I} \rceil$ 缩放, 使总参数量保持在 $1.34 \times 10^5 \pm 5\%$ 以内. 可达速率依次为 4.093、4.297、4.496、4.304 与 4.318 bit/s/Hz, 在 $I = 3$ 处最高; 达到终值 95% 所需回合数依次为 735、1456、1507、1756 与 1574. $I = 4$ 与 $I = 6$ 时可达速率分别降低 4.3% 与 4.0%. 这是因为 $I > 3$ 时新增专家与既有专家职责重叠, 式 (32) 中的示性项在几个近似等价的专家之间频繁切换, 门控权重被稀释, 单个专家的有效样本减少. 因此, 专家数取优化变量数是速率与训练开销之间的折中. 当代理生成的问题变量较少时, 专家数也相应减少.

前两组消融说明增益来自变量级的专家分工, 图 11 进一步给出这一分工的直接观测. 训练结束后固定策略, 仅改变回传链路 SINR, 记录三个专家的归一化门控权重, 其中 w_1、w_2 与 w_3 分别对应 BS 接入功率、BS 回传

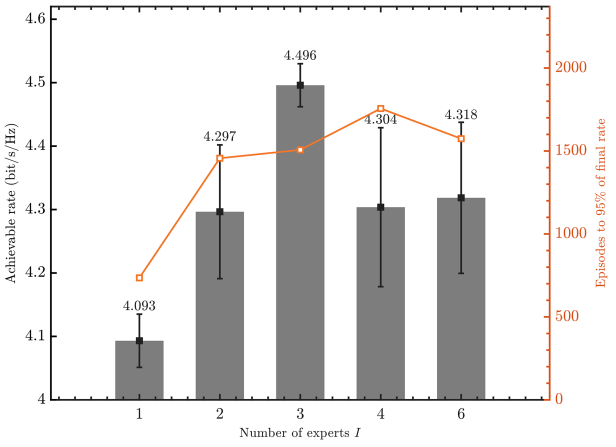

图 10　(网络版彩图) 专家数量对速率与收敛的影响.

Figure 10　(Color online) Impact of the number of experts on achievable rate and convergence speed.

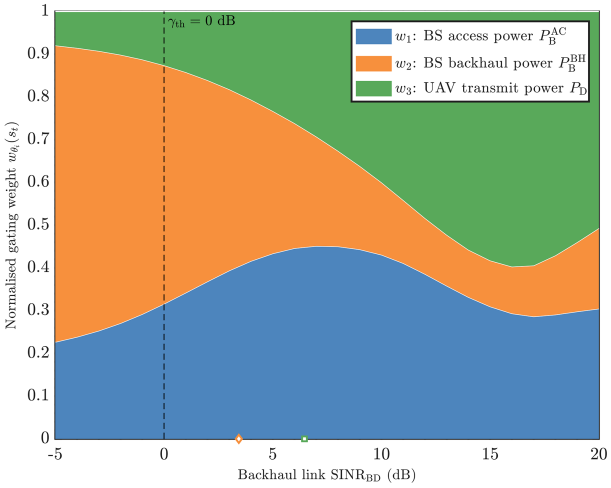

图 11　(网络版彩图) 门控权重随回传链路 SINR 的变化.

Figure 11　(Color online) Migration of gating weights with the backhaul SINR.

功率与 UAV 发射功率. 当 SINR_{BD} 逼近译码门限 γ_{th} 时, 即接近约束式 (23d) 边界, 由式 (10) 可知端到端性能取决于回传, 此时回传功率专家的权重为 0.558, 占据主导. 随 SINR_{BD} 升高, 式 (10) 的取小运算逐步切换至接入跳, 该权重于 3.4 dB 处不再占主导, 在 15 dB 处降至 0.108; UAV 发射功率专家的权重则由 0.08 单调升至 16 dB 处的 0.597, 两者于 6.5 dB 处交叉. 当 SINR_{BD} 升至 20 dB 时, 回传功率专家的权重小幅回升至 0.188. 这是因为约束式 (23d) 已不再为紧约束, 门控机制转而利用该专家降低 P_{B}^{BH}, 以便在总功率约束式 (23b) 下分配给接入天线. 可见, 门控依据回传链路状态在三个专家之间动态调度, 各专家在训练后分别对应一个功率变量. 结合图 9 和 10 的结果, 增益不在于引入多个专家, 而在于变量级分工配合状态相关的调度.

4.4　关键参数敏感性分析

本小节考察关键参数变化时, 所提算法的优势是否仍然成立. 如图 12 所示, BS 功率预算 $P_{\text{B}}^{\max}$ 在 10∼30 W 范围内增大时, 可达速率单调增加而增速递减. 各考察点上所提算法均高于同协议下的 PPO 算法, DF 协议也始终优于 AF 协议.

如图 13 所示, 译码门限 γ_{th} 由 -5 dB 增大至 15 dB 时, 可行解比例降至约 78%, 两种算法的可达速率随之下降. 所提算法下降更为平缓, 在门限最高处仍保持优势. 这是因为约束式 (23d) 接近紧约束时, 门控机制增大了回传链路功率分配的专家权重. 该现象与图 11 所示的权重迁移相互印证.

如图 14 所示, 最低部署高度 $h_{\text{D,low}}^{1}$ 在 50 ∼ 250 m 范围内变化时, 各类传播环境下的可达速率均呈单峰特性. 部署高度较低时, 空地链路受建筑物遮挡, 视距概率较低. 部署高度较高时, 路径损耗增大. 最优部署高度随环境密集程度增大, 郊区与城区约为 150 m, 密集城区与超密集城区约为 200 m. 这表明建筑物遮挡加剧时, 需要更大的仰角才能建立视距链路.

如图 15 所示, $\gamma = 0$ 时策略退化为逐时隙贪婪决策, 长期平均速率最低而约束违反率最高; 随 γ 增大二者同时改善, 并在 $\gamma \geqslant 0.9$ 后趋于平稳, 这表明性能提升并不依赖超参数调节.

5　结论

本文在低空智联网场景下, 采用生成式 AI 代理框架, 针对多类型 UAV 部署与中继协议组合场景, 进行了定制化建模与功率优化分析. 通过语义路由与 RAG 知识库精准召回, 选取了与 UE 需求相匹配的建模维度, 并设计了适配复杂约束优化问题的 MoE-PPO 求解算法. 与 PPO 算法的对比表明, 所提算法的可达速率提升 12.3%,

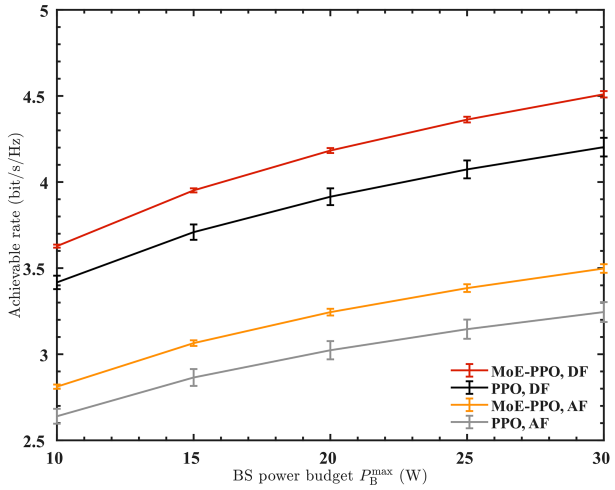

图 12　(网络版彩图) 可达速率随基站功率预算的变化.

Figure 12　(Color online) Achievable rate versus the BS power budget.

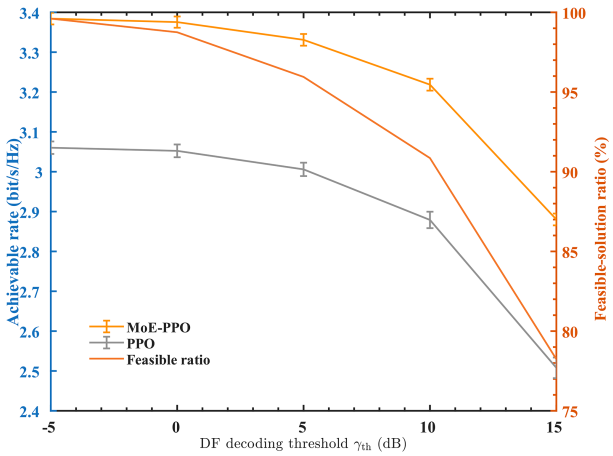

图 13　(网络版彩图) 速率与可行解比例随 γ_{th} 的变化.

Figure 13　(Color online) Achievable rate and feasible-solution ratio versus the DF decoding threshold.

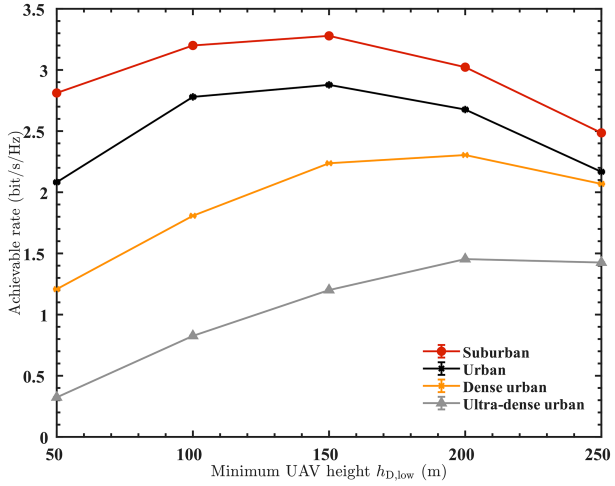

图 14　(网络版彩图) 可达速率随 UAV 最低高度的变化.

Figure 14　(Color online) Achievable rate versus the minimum UAV height.

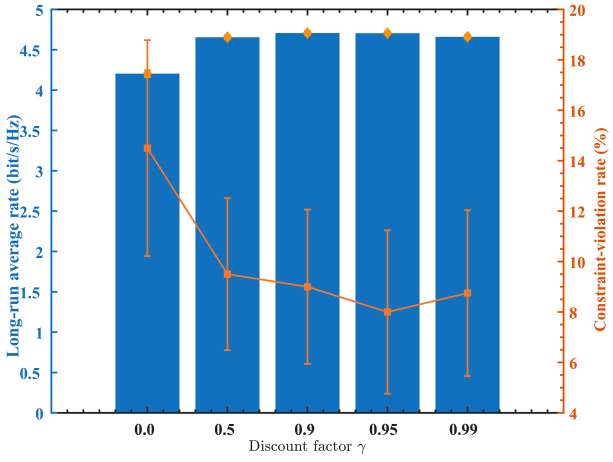

图 15　(网络版彩图) 平均速率与约束违反率随折扣因子的变化.

Figure 15　(Color online) Long-run average rate and constraint-violation ratio versus the discount factor.

约束违反率由 2.9% 降至 0.8%, 明显增强了对译码可行性约束的满足程度. 在未来的研究中, 可将该框架推广至多 UE 接入、UAV 移动轨迹优化以及非理想信道状态信息等更为复杂的低空场景, 并从时延、能效等多个维度分析生成式 AI 代理的引入给低空智联网带来的影响.

参考文献

1　Wu Q H, Dong P H, Zhou F H. Fundamental theory and key technologies for information system and security management of low-altitude economy. BNSFC, 2026, 40: 2–18 [吴启晖, 董培浩, 周福辉. 低空经济信息系统与安全管控基础理论与关键技术. 中国科学基金, 2026, 40: 2–18]

2　Feng Z Y, Zhou Y, Wang L, et al. Theory and key technologies for large-scale unmanned aerial vehicle networked control in low-altitude digital intelligence networks. BNSFC, 2026, 40: 19–34 [冯志勇, 周映, 王琳, 等. 低空数智网的大规模无人机网络化管控理论与关键技术. 中国科学基金, 2026, 40: 19–34]

3 Dong C, Cui C, Jia Z Y, et al. Survey of unified representation technology of multi-dimensional information for low altitude intelligent network. J Electron Inf Technol, 2025, 47: 1215–1229 [董超, 崔灿, 贾子晔, 等. 面向低空智联网的多维信息统一表征技术综述. 电子与信息学报, 2025, 47: 1215–1229]
4 Zhou S, Xie B, Shen D, et al. Agile coverage for low-altitude aerial intelligent networks: a blended hyper-cellular solution. China Commun, 2025, 22: 22–36
5 Song Y, Zeng Y, Yang Y, et al. An overview of cellular ISAC for low-altitude UAV: new opportunities and challenges. IEEE Commun Mag, 2025, 63: 88–95
6 Vaezi M, Lin X, Zhang H, et al. Deep reinforcement learning for interference management in UAV-based 3D networks: potentials and challenges. IEEE Commun Mag, 2024, 62: 134–140
7 Xiao Y, Shi G, Zhang P. Toward agentic AI networking in 6G: a generative foundation model-as-agent approach. IEEE Commun Mag, 2025, 63: 68–74
8 Liu G, Liu Y, Zhang R, et al. Wireless agentic AI with retrieval-augmented multimodal semantic perception. IEEE Commun Mag, 2026, 64: 230–236
9 Zhang R, Tang S, Liu Y, et al. Toward agentic AI: generative information retrieval inspired intelligent communications and networking. IEEE Commun Mag, 2026, 64: 197–204
10 Zhang R, Du H, Liu Y, et al. Generative AI agents with large language model for satellite networks via a mixture of experts transmission. IEEE J Sel Areas Commun, 2024, 42: 3581–3596
11 Du H Y, Liu G Y, Lin Y J, et al. Mixture of experts for intelligent networks: a large language model-enabled approach. In: Proceedings of the 20th International Wireless Communications and Mobile Computing Conference, 2024. 531–536
12 Xiong Y, Zhang R, Liu Y, et al. When knowledge graph meets retrieval augmented generation for wireless networks: a tutorial and case study. IEEE Wireless Commun, 2026, 33: 134–142
13 Liu B, Liu X, Gao S, et al. LLM4CP: adapting large language models for channel prediction. J Commun Inf Netw, 2024, 9: 113–125
14 Wang Z, Zhang J, Du H, et al. Large language model empowered next-generation MIMO networks: fundamentals, challenges, and visions. Digital Commun Netw, 2026, 12: 894–902
15 Chen S, Peng C, Jiang L, et al. A robust deep Q-network (DQN) for heterogeneous tasks and QoS-aware UAV relay communication optimization. IEEE Internet Things J, 2025, 12: 33980–33994
16 Sun K, Yang J, Li J, et al. Proximal policy optimization-based hierarchical decision-making mechanism for resource allocation optimization in UAV networks. Electronics, 2025, 14: 747
17 Xu Y, Wang J, Zhang R, et al. Mixture of experts for decentralized generative AI and reinforcement learning in wireless networks: a comprehensive survey. IEEE Commun Surv Tutorials, 2026, 28: 4051–4085
18 Banagar M, Dhillon H S. Fundamentals of 3D two-hop cellular networks analysis with wireless backhauled UAV: modeling and fundamentals. IEEE Trans Wirel Commun, 2022, 21: 6419–6433
19 3GPP. Study on Enhanced LTE Support for Aerial Vehicles. TR 36.777 V15.0.0. 2017
20 3GPP. Study on Channel Model for Frequencies from 0.5 to 100 GHz. TR 38.901 V16.1.0. 2020

Agentic AI-enabled transmission optimization for low-altitude intelligent networks

Zhangfeng MA[1,3], Ruichen ZHANG[2*], Bolin LIAO[1], Qiuming ZHU[3] & Qihui WU[3]

1. *School of Information Science and Engineering, Shaoyang University, Shaoyang 422000, China*
2. *College of Computing and Data Science, Nanyang Technological University, Singapore 639798, Singapore*
3. *College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China*
* Corresponding author. E-mail: ruichen.zhang@ntu.edu.sg

Abstract To facilitate the intelligent development of the low-altitude economy, the cellular-connected unmanned aerial vehicle (UAV) communication system has become the core solution for air-ground cooperative connections. This paper employs advanced technologies to promote the intelligent modeling and optimization of UAV cellular networks, and conducts research on Agentic AI-enabled transmission optimization for low-altitude intelligent networks. In this paper, retrieval-augmented generation and semantic routing are combined to construct a generative AI agent framework with planning, tool calling and feedback correction capabilities, which accurately characterizes the modeling requirements in four dimensions of propagation scenario, deployment mode, relaying protocol and optimization objective. Based on the system model generated by the agent, an algorithm combining mixture of experts (MoE) with proximal policy optimization (PPO) is adopted to solve the power allocation under the amplify-and-forward and decode-and-forward protocols. After obtaining the optimal power allocation, the algorithm performance is evaluated in terms of the achievable rate and the constraint violation ratio. The outcomes indicate that the proposed MoE-PPO algorithm achieves an achievable rate of about 4.56 bit/s/Hz, which is 12.3% higher than that of the PPO algorithm, and the constraint violation ratio reduces from 2.9% to 0.8%. Under the equal-parameter comparison, the MoE structure itself contributes a gain of 9.8%. The agreement between the agent outputs and the manual modeling results is 79.5%, and it reduces to 55.4% after removing the retrieval-augmented generation module. Consequently, the proposed framework can significantly improve the automation of customized modeling for UAV cellular networks, and alleviate the restriction of the strong coupling among optimization variables on the solving performance. The research can provide theoretical basis and data support for the design and implementation of transmission optimization in low-altitude intelligent networks.

Keywords low-altitude intelligent networks, large language model, generative AI agent, mixture of experts, proximal policy optimization