

流态天线辅助的无人机 ISAC 系统多维资源协同优化策略

邓丹昊, 杨俊, 李嘉悦, 刘大维

(华北电力大学电子与通信工程系, 河北 保定 071003)

摘要: 针对流态天线辅助无人机通感一体化 (ISAC) 系统, 聚焦轨迹规划与通感任务的高效协同问题, 提出了一种基于深度强化学习 (DRL) 的多维资源协同优化策略。首先, 构建基于收发两端流态天线的无人机通感模型, 联合优化无人机轨迹、用户与目标调度、收发天线位置选择; 然后, 将通信速率最大化问题拆分, 引入基于多工作节点的优势演员评论家 (MW-A2C) 算法, 并利用概率强制动作辅助无人机完成位置到达任务; 最后采用贪婪算法求解调度子问题, 设计即时性能反馈机制以更新 DRL 网络权重。仿真结果表明, 所提策略在灵活调度无人机的同时, 能够有效提升系统的总和通感性能, 同时相比于基线策略, 具备更优的可扩展性与鲁棒性。

关键词: 流态天线; 无人机网络; 通感一体化; 多维资源; 深度强化学习

中图分类号: TP309

文献标志码: A

doi: 10.11959/j.issn.1000

Multidimensional resource co-optimization strategy in unmanned aerial vehicle ISAC system assisted by fluid antenna

Deng Danhao, Yang Jun, Li Jiayue, Liu Dawei

Department of Electronic and Communication Engineering, North China Electric Power University, Baoding 071003, China

Abstract: To address the collaboration issue of trajectory planning and communication-sensing tasks in unmanned aerial vehicle integrated sensing and communication (ISAC) system assisted by fluid antenna, a deep reinforcement learning (DRL) based multidimensional resource co-optimization strategy was proposed. Firstly, an unmanned aerial vehicle ISAC model was constructed where fluid antennas were equipped at both transceivers. The unmanned aerial vehicle trajectory, user and target scheduling, and transceivers antenna location selection were jointly optimized. Then the communication rate maximization problem was divided into two sub-problems. A multi-worker advantage actor-critic (MW-A2C) algorithm was introduced, which employed the probabilistic mandatory action to assist the unmanned aerial vehicle in reaching the destination. Finally, the greedy algorithm was adopted to solve the scheduling sub-problem, and an immediately performance feedback mechanism was designed to update the weights of the DRL networks. Simulation results demonstrate that the proposed strategy enables flexible adjustment of the unmanned aerial vehicle trajectory, and can effectively enhance the overall ISAC performance. Moreover, the MW-A2C strategy outperforms the baseline strategies in term of the scalability and robustness.

Key words: Fluid antenna, unmanned aerial vehicle networks, ISAC, multidimensional resource, DRL

0 引言

随着物联网、智能驾驶、智慧城市等垂直应用

的蓬勃发展, 第六代移动通信系统 (6-th Generation wireless systems, 6G) 对数据传输速率和环境

收稿日期: XXXX-XX-XX; 修回日期: XXXX-XX-XX

通信作者: 邓丹昊, dengdanhao@ncepu.edu.cn

基金项目: 中央高校基本科研业务费专项资金资助 (No.2024MS115)

Foundation Items: The Fundamental Research Funds for the Central Universities (No.2024MS115)

感知精度等指标提出了更高的要求,由此通信感知一体化(Integrated Sensing and Communication, ISAC)技术应运而生^[1-3]。但地面 ISAC 网络受制于固定节点,通信链路易被障碍物遮挡,在感知远距离或高动态目标时,面临感知精度下降问题^[4]。无人机凭借高机动性与视距链路传输优势,能够主动调整自身飞行轨迹,通过与 ISAC 平台的融合,可以提供更有利的通信与感知条件,在灾后应急响应、动态目标追踪与临时热点覆盖等领域发挥着越来越重要的作用^[5-6]。然而,无人机面临尺寸与功耗的双重限制,难以通过增加天线数量来增强通感性能。因此,如何在不牺牲有效载荷与续航能力的前提下,提升资源利用率,已成为当前无人机 ISAC 系统研究中亟待解决的关键问题。

近年来,流态天线技术吸引了国内外研究者的广泛关注,通过在预设空间内切换天线的物理位置,可以获得额外的空间自由度^[7-8]。这种新兴可重构天线技术的引入,在释放更多性能增益潜力的同时,也扩展了原有的资源维度,显著增加了资源优化的复杂度。文献[9]分析了流态天线在增强覆盖、能效、频谱容量等关键性能指标方面的潜力,验证了流态天线与非地面网络等 6G 使能技术的融合优势。文献[10]研究流态天线在多用户下行场景中的时延问题,在传统功率分配方案的基础上,将端口选择纳入优化变量,基于迭代算法显著提升有效容量。文献[11]考虑流态天线辅助的超大规模 MIMO 系统,聚焦信道的空间非平稳性,通过调整天线位置与预编码向量最大化系统总和速率。文献[12]引入语义通信范式,联合优化发射波束、天线端口和语义压缩率,在满足功率阈值的前提下,最大化等效语义速率。文献[13]针对流态天线辅助的协作式非正交多址接入网络,提出了一种基于交替优化的功率分配策略,以提升边缘用户的最小中断概率。

作为通用天线技术,流态天线能够赋能 ISAC 场景,在传统的波束赋形基础上,通过改变天线位置调整波束方向,从而满足通信和感知对天线波束的不同需求。文献[14]旨在平衡通信与感知性能,设计了一种基于稀疏优化的迭代策略,从而降低系统的资源开销。文献[15]则面向多用户多目标 ISAC 网络,通过优化发射接入点的波束赋形矩阵与天线位置向量,实现了多节点协作性能的有效增强。文

献[16]推导了流态天线在无源 ISAC 系统的可达性能上下界,并验证了接收孔径与感知分辨率间的相关性。文献[17]引入全双工技术,考虑收发端自干扰对通感性能的影响,联合发射与接收两端天线优化,提出一种基于连续凸逼近的速率最优方案。

针对无人机与流态天线的结合场景,其需要在考虑无线资源管理问题的基础上,综合无人机飞行过程面临的各类挑战,但相关研究仍处于起步阶段。文献[18]针对单无人机场景和预设线性飞行轨迹,研究建模精度和信道容量等关键性能指标,证明了流态天线对通信性能的显著影响。文献[19]面向悬停式无人机,基于最大增益选择策略,自适应的调整天线位置,旨在提升系统的能量效率。文献[20]扩展至多无人机协作场景,在满足信息传输时延约束的前提下最小化移动目标的定位误差,提出基于多智能体强化学习的无人机轨迹与天线端口优化方案。文献[21]结合流态天线与可重构智能表面技术,采用基于二阶锥规划的逐次凸逼近方法,联合优化端口位置和反射相位。与前述论文聚焦于通信传输不同,文献[22]则关注无人机感知方面,通过构造目标估计的平均克拉美-罗界,在满足起止位置和速度约束的前提下,交替优化无人机位置与资源。然而,现有的多数论文均假设无人机部署在固定位置或基于预设轨迹,未能充分利用无人机的灵活调度能力。

针对流态天线应用于无人机 ISAC 系统的研究相对较少。文献[23]设定基站承担 ISAC 任务,用于感知固定位置的目标,并与地面和空中用户进行通信。文献[24]则假设基站由于遮挡无法感知地面目标,无人机接收感知回波后,将部分计算任务卸载到远端基站,通过调整无人机轨迹与资源,最小化感知数据处理任务的平均完成时延。文献[25]考虑单无人机与地面用户和目标间的通感任务,提出三时间尺度优化框架,基于不同周期更新无人机位置、天线端口和波束赋形向量,从而平衡通信速率与感知精度。上述论文同样将视角着重于流态天线所引发的资源管理问题,对无人机本身续航问题的研究较少。

针对上述问题,本文面向流态天线辅助的无人机 ISAC 系统,综合考虑无人机飞行状态与资源管理效能,在满足各目标感知性能的前提下,最大化系统的长期通信和速率。本文的主要贡献如下:

(1) 考虑无人机受限的通感算资源, 提出基于流态天线的无人机时分ISAC模型, 在无人机起止位置约束和最大距离约束的限制下, 研究基于无人机轨迹设计、用户与目标调度、发射与接收天线位置选择的多维资源协同优化问题。

(2) 将具备非凸性与序列决策性的问题拆分求解, 并提出基于即时奖励反馈的马尔可夫决策过程(Markov Decision Process, MDP)模型, 针对无人机轨迹的稀疏奖励问题, 通过概率强制动作确保无人机端到端的任务闭环。

(3) 第一阶段针对轨迹与天线位置优化, 引入基于多工作节点的优势演员评论家(Multi-Worker Advantage Actor-Critic, MW-A2C)算法, 利用并行经验采集提升训练效率。第二阶段则利用贪婪算法求解调度问题, 以减轻无人机计算负担。

1 系统模型

1.1 系统概述

本文构建了一个流态天线辅助的无人机ISAC系统, 如图1所示, 该系统由一架无人机、索引为 $k \in \mathcal{K} = \{1, \dots, K\}$ 的 K 个地面用户以及索引为 $m \in \mathcal{M} = \{1, \dots, M\}$ 的 M 个感知目标组成。无人机在发射端与接收端分别配置 L_t 与 L_r 个流态天线, 而地面用户和感知目标均为单天线。

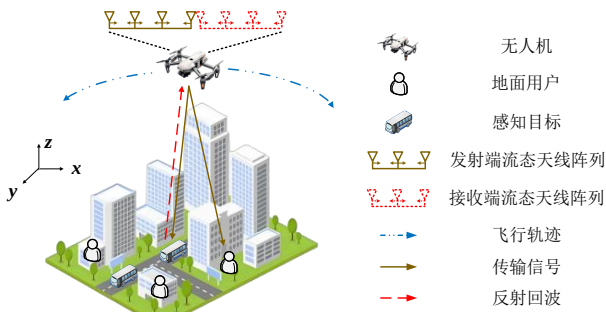


图1 流态天线辅助的无人机ISAC系统

无人机的飞行周期 T 被均分为 N 个时隙, 时隙的持续时间满足 $\Delta = T/N$ 。假设 Δ 被划分的足够小, 使得无人机的位置在每个时隙内保持不变。基于三维笛卡尔坐标系统, 无人机在时隙 n 的位置可以表示为 $\mathbf{q}[n] = [x_U[n], y_U[n], H_U]$, 其中 H_U 为无人机飞行高度。发射端的流态天线采用多天线共轨滑动架构, 在运行过程中, 所有天线被限制在同一条线性导轨上, 并沿导轨按顺序排列, 水平位置可

以表示为 $\mathbf{x}[n] = \{x_1[n], \dots, x_{L_t}[n]\}$, 其中 $x_{l_t}[n] \in [0, D_{\max}]$ 为第 l_t 根发射天线在时隙 n 的位置, $D_{\max}$ 为预设导轨长度。为防止相邻流态天线间的耦合, 最小间距约束可以表示为:

$$|x_{l_t}[n] - x_{l_t-1}[n]| \geq D_{\min}, \forall l_t \in \{2, \dots, L_t\} \quad (1)$$

其中 $D_{\min}$ 表示相邻天线的最小间距。同理, 接收端流态天线的水平位置可以表示为 $\mathbf{y}[n] = \{y_1[n], \dots, y_{L_r}[n]\}$, $y_{l_r}[n] \in [0, D_{\max}]$, 且需满足约束:

$$|y_{l_r}[n] - y_{l_r-1}[n]| \geq D_{\min}, \forall l_r \in \{2, \dots, L_r\} \quad (2)$$

此外, 用户和目标均为固定位置, 分别表示为 $\mathbf{q}_k = [x_k, y_k, 0]$ 和 $\mathbf{q}_m = [x_m, y_m, 0]$ 。

与地面基站不同, 无人机受机载能量和计算能力的限制。因此, 本文所提的ISAC系统采用时分多址接入方式。规定在每时隙中, 无人机最多只能与一个用户进行通信, 同时最多仅能感知一个目标, 从而避免用户间干扰导致的性能降低, 以及多目标感知在接收端处理时造成的高复杂度^[26]。设定0-1变量 $\alpha_k[n]$ 表征无人机与用户 k 的关联关系, 其中 $\alpha_k[n] = 1$ 表示无人机在时隙 n 向用户 k 发射信号, 否则 $\alpha_k[n] = 0$; 同理, 设定0-1变量 $\beta_m[n]$ 表征无人机与目标 m 的关联关系, 其中 $\beta_m[n] = 1$ 表示无人机在时隙 n 感知目标 m , 否则 $\beta_m[n] = 0$ 。基于上述模型, 存在以下约束条件:

$$\sum_{k \in \mathcal{K}} \alpha_k[n] \leq 1, \forall n \quad (3)$$

$$\sum_{m \in \mathcal{M}} \beta_m[n] \leq 1, \forall n \quad (4)$$

1.2 无人机移动模型

考虑飞行过程中的能量消耗, 无人机需要利用充电站周期性补充能量, 以提供持续性通感服务。同时假设在单位周期内, 机载总能量充足, 能够支撑无人机以最大推进功率完成通感任务。因此, 无人机的起止位置满足以下约束:

$$\mathbf{q}[0] = \mathbf{q}_I \quad (5)$$

$$\mathbf{q}[N] = \mathbf{q}_F \quad (6)$$

其中 $\mathbf{q}_I$ 和 $\mathbf{q}_F$ 分别表示无人机的起始位置坐标和终止位置坐标。

受限于无人机最大飞行速度 $V_{\max}$, 在单位时隙内, 无人机的最远飞行距离不得超过 $d_{\max} = V_{\max} \Delta$:

$$\|q[n] - q[n-1]\| \leq d_{\max}, \forall n \in \{1, \dots, N\} \quad (7)$$

1.3 通信模型

由于天线阵列尺寸远小于信号传播距离，因此本文采用远场信道模型，设定发射端所有的流态天线具有相同的离开角（Angle of Departure, AoD）和路径增益幅度^[24]，仅相位会随其位置变化而改变。

无人机与用户 k 间的 AoD 可以表示为：

$$\theta_k[n] = \arcsin \left(\frac{H_U}{\sqrt{\|q[n] - q_k\|^2}} \right) \quad (8)$$

则从无人机到用户 k 的发射导向矢量可以表示为：

$$\mathbf{a}_1(\mathbf{x}[n], \theta_k[n]) = \left[e^{j \frac{2\pi}{\lambda} x_1[n] \sin(\theta_k[n])}, \dots, e^{j \frac{2\pi}{\lambda} x_{L_t}[n] \sin(\theta_k[n])} \right]^T \quad (9)$$

其中 λ 表示载波波长。

当无人机具备足够高的飞行高度时，其与地面用户间的信道可视作视距（Line-of-Sight, LoS）链路，且视距概率接近 100 %^[27]。因此，无人机与用户 k 间的信道向量可以表示为：

$$\mathbf{h}_k[n] = \sqrt{\frac{\beta_0}{d_k^2[n]}} \mathbf{a}_1(\mathbf{x}[n], \theta_k[n]) \quad (10)$$

其中 β_0 表示单位距离 1 米时的信道功率增益， $d_k[n]$ 表示无人机与用户 k 间的距离。

¹ 波束赋形方案及其优化策略可以进而提升流态天线的通感性能，但此并非本文讨论的重点，留待未来工作。

$$\mathbf{w}_t = \mathbf{1}_{L_t} / \sqrt{L_t} \quad n$$

发射端基于等幅同相激励，波束赋形向量固定为全 1 向量，即。¹在时隙，发射信号可以表示为：

$$\mathbf{x}_k[n] = \sqrt{\frac{p_c}{L_t}} \mathbf{1}_{L_t} s[n] \quad (11)$$

其中 p_c 表示无人机的通信发射功率， $s[n]$ 为传输信号，满足 $E[|s[n]|^2] = 1$ 。则用户 k 的接收信号为：

$$y_k[n] = \alpha_k[n] \mathbf{h}_k[n] \mathbf{x}_k[n] + n_k[n]$$

$$= \alpha_k[n] \sqrt{\frac{p_c}{L_t}} \sqrt{\frac{\beta_0}{d_k^2[n]}} \mathbf{a}_1(\mathbf{x}[n], \theta_k[n]) \mathbf{1}_{L_t} s[n] + n_k[n] \quad (12)$$

其中 $n_k[n]$ 表示加性高斯白噪声（Additive White Gaussian Noise, AWGN）。用户 k 的信噪比（Signal to Noise Ratio, SNR）可以表示为：

$$\gamma_k[n] = \alpha_k[n] \frac{p_c}{L_t} \frac{\beta_0}{d_k^2[n]} \frac{|\mathbf{a}_1(\mathbf{x}[n], \theta_k[n]) \mathbf{1}_{L_t}|^2}{\sigma_k^2} \quad (13)$$

其中 σ_k^2 为 AWGN 功率。用户 k 的可达速率表示为：

$$R_k[n] = \log_2(1 + \gamma_k[n]) \quad (14)$$

1.4 感知模型

无人机利用信号 $s[n]$ 的反射回波感知目标。无人机到目标 m 的发射和接收导向矢量分别表示为：

$$\mathbf{a}_2(\mathbf{x}[n], \theta_m[n]) = \left[e^{j \frac{2\pi}{\lambda} x_1[n] \sin(\theta_m[n])}, \dots, e^{j \frac{2\pi}{\lambda} x_{L_t}[n] \sin(\theta_m[n])} \right]^T \quad (15)$$

$$\mathbf{a}_3(\mathbf{y}[n], \theta_m[n]) = \left[e^{j \frac{2\pi}{\lambda} y_1[n] \sin(\theta_m[n])}, \dots, e^{j \frac{2\pi}{\lambda} y_{L_r}[n] \sin(\theta_m[n])} \right]^T \quad (16)$$

其中 $\theta_m[n]$ 表示无人机与目标 m 间的 AoD。

由无人机到目标 m ，再反射回无人机的感知信道矩阵可以表示为：

$$\mathbf{h}_m[n] = \sqrt{\frac{G_t G_r \lambda^2 \sigma_r}{(4\pi)^3 d_m^4[n]}} \mathbf{a}_3(\mathbf{y}[n], \theta_m[n]) \mathbf{a}_2^H(\mathbf{x}[n], \theta_m[n]) \quad (17)$$

其中 G_t 和 G_r 分别表示无人机发射和接收天线增益， σ_r 为雷达横截面， $d_m[n]$ 表示无人机与目标 m 间的距离。无人机接收的目标 m 回波信号为：

$$y_m[n] = \beta_m[n] \sqrt{\frac{p_c}{L_r L_t}} \mathbf{1}_{L_r} \mathbf{h}_m[n] \mathbf{1}_{L_t} s[n] + c_m[n] + \sqrt{\rho \frac{p_c}{L_r L_t}} \mathbf{1}_{L_r} \mathbf{h}_{\text{sl}}[n] \mathbf{1}_{L_t} s[n] + n_r[n] \quad (18)$$

其中 $c_m[n]$ 为目标 m 对应的环境杂波信号， $\mathbf{h}_{\text{sl}}[n]$ 表示自干扰信道矩阵， $\rho \in [0, 1]$ 表示自干扰消除后的残余系数， $n_r[n]$ 表示无人机接收端感知 AWGN。经匹配滤波后，目标 m 的感知信干噪比（Signal to

Interference plus Noise Ratio, SINR) 可以表示为:

$$\gamma_m[n] = \beta_m[n] \frac{p_c}{L_r L_t} \frac{G_t G_r \lambda^2 \sigma_r}{(4\pi)^3 d_m^4[n]} \times \frac{\left| 1_{L_r} \mathbf{a}_3(\mathbf{y}[n], \theta_m[n]) \mathbf{a}_2^H(\mathbf{x}[n], \theta_m[n]) 1_{L_t} \right|^2}{P_{\text{clu}}[n] + P_{\text{sl}}[n] + \sigma_r^2} \quad (19)$$

其中 $P_{\text{clu}}[n]$ 表示杂波干扰功率, 为简化分析, 假设地面散射特性均匀, 同一时隙内不同感知目标对应的统计杂波功率近似相同。 $P_{\text{sl}}[n]$ 表示自干扰功率, σ_r^2 为无人机接收端感知 AWGN 功率。目标 m 的感知速率表示为:

$$R_m[n] = \log_2(1 + \gamma_m[n]) \quad (20)$$

2 问题描述

本文的优化目标是在保证各目标感知性能的前提下, 最大化系统的长期通信和速率。基于此, 本文提出了一种基于无人机轨迹设计、用户与目标调度、发射与接收天线位置选择的资源优化问题, 其数学模型可以表示为:

$$\begin{aligned} \max_{\mathcal{Q}, \mathcal{P}, \mathcal{B}, \mathcal{A}, \mathcal{Y}} \quad & \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} R_k[n] \\ \text{s.t.} \quad & \text{C1: } \sum_{n \in \mathcal{N}} R_k[n] \geq R_{\text{th}}^c, \forall k \\ & \text{C2: } \sum_{n \in \mathcal{N}} R_m[n] \geq R_{\text{th}}^s, \forall m \\ & \text{C3: } x_{\min} \leq x_U[n] \leq x_{\max}, \forall n \\ & \text{C4: } y_{\min} \leq y_U[n] \leq y_{\max}, \forall n \\ & \text{C5: } \alpha_k[n], \beta_m[n] \in \{0, 1\}, \forall k, m, n \\ & \text{C6: (1) - (7)} \end{aligned} \quad (21)$$

其中时隙集合 $\mathcal{N} = \{1, \dots, N\}$, $\mathcal{Q} = \{\mathbf{q}[n]\}_{n \in \mathcal{N}}$ 表示无人机位置集合, $\mathcal{P} = \{\alpha_k[n]\}_{k \in \mathcal{K}, n \in \mathcal{N}}$ 表示地面用户调度集合, $\mathcal{B} = \{\beta_m[n]\}_{m \in \mathcal{M}, n \in \mathcal{N}}$ 表示感知目标调度集合, $\mathcal{A} = \{\mathbf{x}[n]\}_{n \in \mathcal{N}}$ 表示发射端流态天线的天线位置选择集合, $\mathcal{Y} = \{\mathbf{y}[n]\}_{n \in \mathcal{N}}$ 为接收端流态天线的天线位置选择集合。此外, 约束 C1 表示各用户的通信和速率需满足最低阈值 R_{th}^c ; 约束 C2 表示各目标的感知和速率需满足最低阈值 R_{th}^s ; C3 和 C4 表示无人机的飞行范围限制; C5 为 0-1 变量约束。

问题 (21) 的优化目标和优化变量均具备非凸性, 且优化变量包含连续与离散两种类型, 求解较为困难。同时无人机轨迹优化具有时间相关性, 因

此问题 (21) 还属于序列决策问题, 需要联合求解 N 个时隙的所有优化变量, 以获取最优解, 计算复杂度较高。为此, 本文首先基于变量性质, 聚焦具体时隙, 将问题 (21) 拆解为 N 个联合无人机位置与天线位置选择子问题 P1, 以及 N 个用户与目标调度子问题 P2。需要注意的是, 每个时隙 P1 子问题的优化目标仍然是系统的长期通信和速率, 而非当前时刻的即时通信和速率:

$$\begin{aligned} (P1.n) \quad & \max_{\mathbf{q}[n], \mathbf{x}[n], \mathbf{y}[n]} \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} R_k[n] \\ \text{s.t.} \quad & \text{C1: } \sum_{n \in \mathcal{N}} R_k[n] \geq R_{\text{th}}^c, \forall k \\ & \text{C2: } \sum_{n \in \mathcal{N}} R_m[n] \geq R_{\text{th}}^s, \forall m \quad (22) \\ & \text{C3: } x_{\min} \leq x_U[n] \leq x_{\max} \\ & \text{C4: } y_{\min} \leq y_U[n] \leq y_{\max} \\ & \text{C5: (1), (2), (5) - (7)} \end{aligned}$$

N 个 P2 子问题则是在对应时隙完成无人机位置与天线位置决策后优化, 仅需考虑当前时隙的即时通信和速率最优, 并舍弃无关约束:

$$\begin{aligned} (P2.n) \quad & \max_{\{\alpha_k[n]\}_{k \in \mathcal{K}}, \{\beta_m[n]\}_{m \in \mathcal{M}}} \sum_{k \in \mathcal{K}} R_k[n] \\ \text{s.t.} \quad & \text{C1: } \sum_{n \in \mathcal{N}} R_k[n] \geq R_{\text{th}}^c, \forall k \\ & \text{C2: } \sum_{n \in \mathcal{N}} R_m[n] \geq R_{\text{th}}^s, \forall m \quad (23) \\ & \text{C3: } \alpha_k[n] \in \{0, 1\}, \forall k \\ & \text{C4: } \beta_m[n] \in \{0, 1\}, \forall m \\ & \text{C5: (3), (4)} \end{aligned}$$

如图 2 所示, 本文将前者 P1 时序联合建模为 MDP, 并引入深度强化学习 (Deep Reinforcement Learning, DRL) 策略进行求解, 由此关联时序性决策与潜在效益。后者 P2 采用贪婪算法求解, 并将即时性能反馈至 DRL 模型, 以调整相应轨迹设计与天线位置选择策略。

3 基于 MW-A2C 的多维资源协同优化策略

3.1 MDP 模型

在本文建模中, 整个无人机 ISAC 系统充当动态环境, 搭载 DRL 算法的无人机被视作智能体, 用以完成资源决策, 智能体与环境间的交互过程由 MDP 模型刻画。具体来说, MDP 将 N 个优化子问题 P1 转化为由状态 s , 动作 a 和奖励 r 构成的经验

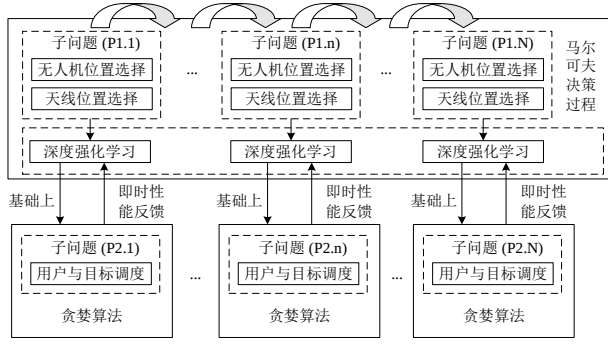


图2 子问题求解逻辑

序列，而资源决策过程被建模为动作选择策略 π ，通过优化该策略以实现长期累积回报的最大化。基于第2章中的优化问题，状态空间、动作空间、奖励函数的具体定义如下：

状态空间 $\mathcal{S}$ ：状态 $s_n \in \mathcal{S}$ 表示时隙 n 所有环境变量的集合，用以指导策略来选择资源动作。针对子问题P1，资源决策受节点类型、位置信息、AoD角度、信道信息、飞行情况、累计速率等多重因素的影响。节点类型表示为 $\mathcal{F} = \{\varphi_i\}_{i \in \mathcal{K} \cup \mathcal{M}}$ ，用来区分地面用户 $\varphi_i = 1$ 与感知目标 $\varphi_i = 0$ ，并表明节点数量。位置信息旨在提供完整的空间拓扑，包括当前时刻的无人机位置 $\mathbf{q}[n]$ 、用户位置与目标位置 $\mathcal{Q}_g = \{\mathbf{q}_i\}_{i \in \mathcal{K} \cup \mathcal{M}}$ ，并利用起止位置 $\mathbf{q}_I$ 和 $\mathbf{q}_F$ 引导无人机飞行。当前AoD角度 $\mathcal{E}[n] = \{\theta_i[n]\}_{i \in \mathcal{K} \cup \mathcal{M}}$ 的显式纳入可提供信号传播方向的高层次特征，降低DRL算法的学习负担。信道信息是资源选择决策的核心依据，分为当前时刻的通信信道向量集合和感知信道向量集合 $\mathcal{H}[n] = \{\mathbf{h}_i[n]\}_{i \in \mathcal{K} \cup \mathcal{M}}$ 。飞行情况包括时间和距离两个维度，在时间上，利用飞行周期 N 和当前时隙 n 指示任务完成进度，从而辅助智能体在不同阶段采取差异化的决策策略；在距离上，利用当前时隙无人机与终止位置间的距离 $d_F[n]$ 来引导无人机平衡能耗与通感性能。累计速率为各目标的当前感知和速率 $\mathcal{R}[n] = \{\sum_{\tau=1}^n R_m[\tau]\}_{m \in \mathcal{M}}$ 及感知速率阈值 R_{th}^s ，用以表征约束C2的满足情况，辅助智能体平衡通信与感知。

综上所述，当前状态 s_n 可以表示为：

$$s_n = \{\mathcal{F}, \mathbf{q}[n], \mathcal{Q}_g, \mathbf{q}_I, \mathbf{q}_F, \mathcal{E}[n], \mathcal{H}[n], n, N, d_F[n], \mathcal{R}[n], R_{th}^s\} \quad (22)$$

动作空间 $\mathcal{A}$ ：动作 $a_n \in \mathcal{A}$ 为针对当前状态 s_n 执行的资源选择，包括下一时刻的无人机位置 $\mathbf{q}[n+1]$ ，以及下一时刻的收发端天线位置选择 $\mathbf{x}[n+1]$ 和 $\mathbf{y}[n+1]$ 。针对无人机位置 $\mathbf{q}[n+1]$ ，其受制于飞行距离约束(7)，因此多数的动作变量是不可行的。同时无人机位置属于连续变量，智能体难以引导无人机在有限步数 N 内从固定起点精确到达指定终点。为解决上述问题，本文假设在单位时隙内无人机沿固定的方向和速度飞行，利用当前时隙的无人机运动指令来替换下一时隙的无人机位置。定义 $f[n] \in \{1, \dots, F\}$ 为无人机飞行方向， $v[n] \in [0, V_{max}]$ 为无人机移动速度，其中 F 表示可行飞行方向数。则当前时隙 n 的动作 a_n 可以表示为：

$$a_n = \{f[n], v[n], \mathbf{x}[n+1], \mathbf{y}[n+1]\} \quad (23)$$

强制作动：当无人机的机载能量耗尽时，其存在坠毁的风险，因此终止位置约束(6)须在每个飞行周期内始终成立。然而，即使动作(23)将无人机飞行固化为有限的离散方向，该约束仍然难以精确满足。本文在普通决策动作的基础上引入概率强制作动，其触发时隙满足以下约束：

$$(N-n)\zeta d_{max} \leq d_F[n] \leq (N-n)d_{max} \quad (24)$$

其中 $\zeta \in (0, 1)$ 用于控制触发范围。强制作动的触发概率为 $p_{man} \in [0, 1]$ ，当强制作动被触发后，无人机以最大速度飞行，且飞行方向在决策方向 $f[n]$ 的基础上，叠加与终止位置连线方向的矢量。

奖励函数：系统的总体奖励为时间的累计值，利用折扣因子 γ 控制未来奖励的影响程度，可以表示为 $R_n = \sum_{\tau=0}^{N-n} \gamma^\tau r_{n+\tau}$ ，其中 r_n 为当前时隙 n 的即时奖励。而即时奖励的设计需要考虑每个状态-动作对的可行性，并且需要吻合子问题P1的优化目标。为此，本文将奖励函数建模为：

$$r_n = w_r \xi_r[n] \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} R_k[n] + \sum_i w_i \xi_i[n] \mu_i[n] \quad (25)$$

其中 w_r 和 $\{w_i\}_i$ 均表示奖励权重。 $\xi_r[n]$ 为0-1变量表示无人机在时隙 N 是否到达终止位置， $\xi_r[n] = 1$ 表示当前时隙为 N 且无人机成功到达，否则 $\xi_r[n] = 0$ 。 $\xi_i[n]$ 同样为0-1变量，用以表征当前动作 a_n 是否满足第 i 个约束， $i \in \{1, \dots, 9\}$ ， $\xi_i[n] = 1$ 表示动作 a_n 违反该约束，需要被赋予惩罚项，否则 $\xi_i[n] = 0$ 。 $\mu_i[n]$ 表示第 i 个约束的负奖励，其中约

束 C1 和 C2 的负奖励可以分别表示为:

$$\mu_1[n] = -\frac{1}{K} \sum_{k \in \mathcal{K}} \max \left(0, \frac{R_{th}^c - \sum_{v=1}^n R_k[v]}{R_{th}^c} \right) \quad (26)$$

$$\mu_2[n] = -\frac{1}{M} \sum_{m \in \mathcal{M}} \max \left(0, \frac{R_{th}^s - \sum_{v=1}^n R_m[v]}{R_{th}^s} \right) \quad (27)$$

其他项均为负值常数。

3.2 MW-A2C 算法求解

根据 3.1 中的 MDP 模型, 状态和动作中都存在连续变量, 属于无限状态-动作空间, 由此产生的维度灾难使得基础的 DRL 算法面临收敛难题。为了应对该挑战, 本文提出了基于 MW-A2C 的轨迹与天线位置协同优化策略。

如图 3 所示, MW-A2C 框架包括一个全局网络和 J 个 worker。全局网络包含权重为 θ_μ 的 actor 网络, 用以表征动作选择策略 $\pi(a_n | s_n; \theta_\mu)$; 同时囊括权重为 θ_Q 的 critic 网络, 用以估计状态的期望累积回报, 即值函数 $V(s_n; \theta_Q)$ 。所有 worker 均采用与全局网络相同的网络结构, 各自独立运行环境副本, 以完成并行 (s_n, a_n, r_n, s_{n+1}) 样本采集。权重更新任务由全局网络负责, 各 worker 在初始化阶段复制全局网络权重, 并在每次更新后从全局网络同步最新权重, 以保证本地策略与全局策略的一致性。

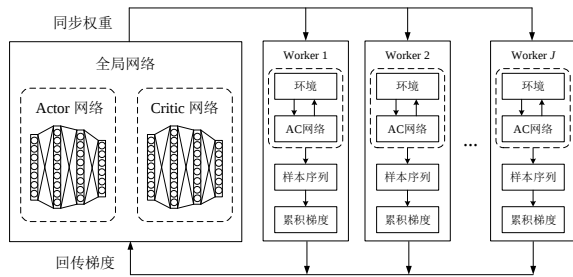


图3 MW-A2C算法框架

每次更新, worker 基于 l 步采样原则, 首先与环境交互 l 步, 获得 l 个经验样本。然后维护局部状态, 记录当前的状态与已执行步数, 确保下一轮交互从上一轮的终止状态继续推进, 直至完成完整周期 N 步。随后计算每个状态-动作对的优势函数:

$$A(s_n, a_n; \theta_\mu, \theta_Q) = r_n + \gamma r_{n+1} \dots + \gamma^{L-1} r_{n+L-1} + \gamma^L V(s_{n+L}; \theta_Q) - V(s_n; \theta_Q) \quad (28)$$

其中 L 表示剩余采样步数。

Actor 网络通过梯度上升调整权重, 使得高优势动作的选取概率增加, 以提升策略的期望累积奖励, 因此对应的损失函数可以表示为:

$$f_n(\theta_\mu) = \log \pi(a_n | s_n; \theta_\mu) A(s_n, a_n; \theta_\mu, \theta_Q) \quad (29)$$

Critic 网络则通过最小化输出值函数与累积奖励之间的时分 (Temporal-Difference, TD) 误差来更新权重, 以提升值函数的准确性。对应的损失函数可以表示为:

$$f_n(\theta_Q) = (V(s_n; \theta_Q) - R_n)^2 \quad (30)$$

其中 R_n 同样基于 l 步采样原则。

最后, 每个 worker 计算各自的累积梯度, 并将梯度值回传给全局网络, 由全局网络同步进行权重更新。第 j 个 worker 的累积 actor 梯度与累积 critic 梯度可以分别表示为:

$$d\mu^j = d\mu^j + \sum_{t=0}^{l-1} \nabla_{\theta_\mu} f_{n_{in}+t}^j(\theta_\mu) \quad (31)$$

$$dQ^j = dQ^j + \sum_{t=0}^{l-1} \nabla_{\theta_Q} f_{n_{in}+t}^j(\theta_Q) \quad (32)$$

其中 n_{in} 表示当前更新的初始时隙。

上述 MW-A2C 框架经过充分交互与迭代训练后, 当 actor 与 critic 网络的损失函数均收敛至平稳, 且终止时隙 N 的奖励值不再出现明显波动时, 训练过程终止, 此时获得的动作选择策略即为所求的轨迹与天线位置协同优化策略。

3.3 贪婪算法求解

子问题 P2 可以拆分为: 基于用户调度的通信速率最优与阈值满足问题和基于目标调度的感知速率阈值满足问题, 两者相互独立。由于两者的约束均存在长期性能, 无法依据瞬时指标评估。可基于贪婪算法, 寻求当前时隙的即时最优, 作为近似最优决策, 以减轻无人机计算负担, 并通过即时性能反馈机制调整 DRL 决策, 协同无人机位置和天线位置优化实现阈值约束满足。因此, 用户调度与目标调度策略可以分别表示为:

$$\alpha_k[n] = \begin{cases} 1, & |h_k[n]1_{L_i}|^2 > |h_k[n]1_{L_i}|^2, \forall k' \neq k \\ 0, & \text{其他} \end{cases} \quad (33)$$

$$\beta_m[n] = \begin{cases} 1, & |1_{L_r} \mathbf{h}_m[n] 1_{L_t}|^2 > |1_{L_r} \mathbf{h}_{m'}[n] 1_{L_t}|^2, \forall m' \neq m \\ 0, & \text{其他} \end{cases} \quad (34)$$

3.4 算法总体流程

基于上述建模与训练求解流程, 本文提出的基于 MW-A2C 的多维资源协同优化策略如算法 1 所示。

算法 1 基于 MW-A2C 的多维资源协同优化策略

输入 Actor 和 critic 网络权重、并行 worker 数 J 、训练更新次数 E 、用户和目标位置、无人机起止位置、飞行周期 N 、学习率等

输出 Actor 网络动作选择策略、用户调度和目标调度策略

- 1) 初始化 Actor 和 critic 网络权重
- 2) 初始化 J 个 worker, 并复制全局网络权重
- 3) for $e \leq E$ do
- 4) 同步运行 J 个 worker
- 5) for $n \leq N$ do
- 6) 每个 worker 并行执行
- 7) 根据策略 $\pi(a_n | s_n; \theta_\mu)$ 选择动作 a_n
- 8) 通过贪婪算法 (公式(33)和(34)) 求解用户调度和目标调度
- 9) 计算即时奖励 r_n
- 10) 获得并储存样本 (s_n, a_n, r_n, s_{n+1})
- 11) $n \leftarrow n + 1$
- 12) if $n \bmod l = 0$ 或者 $n = N$
- 13) $e \leftarrow e + 1$
- 14) 根据(28)计算所有样本的优势函数
- 15) 根据(31)和(32)计算累积梯度
- 16) 回传梯度 $d\mu^j$ 和 dQ^j 至全局网络
- 17) 全局网络: 待收集所有 worker 的回传梯度后, 更新权重 θ_μ 和 θ_Q
- 18) worker 同步全局网络权重
- 19) $d\mu^j = 0, dQ^j = 0$
- 20) 丢弃所有经验样本
- 21) end if
- 22) end for
- 23) end for

4 仿真结果与分析

4.1 仿真参数设置

仿真设置无人机飞行区域固定在 400×400 平方米的正方形区域内, 起始位置位于 $(0, 200)$, 终止位置为 $(400, 200)$, 用户位置和目标位置在该区域内随机生成。网络仿真参数如表 1 所示。

表 1 网络仿真参数

参数名称	参数值
无人机飞行高度 H_U/m	100
无人机最大飞行速度 $V_{\max}/(\text{m/s})$	40
无人机发射功率 p_e/W	1
时隙长度 Δ/s	1
信道功率增益 β_0/dB	-60

表 1 网络仿真参数(续表)

参数名称	参数值
无人机发射天线增益 G_t	127.4
无人机接收天线增益 G_r	127.4
雷达横截面 σ_r	1
载波波长 λ/mm	10
预设导轨长度 $D_{\max}$	202
相邻天线的最小间距 $D_{\min}$	0.52
感知速率阈值 $R_{\text{th}}^s/(\text{bps/Hz})$	5
通信速率阈值 $R_{\text{th}}^c/(\text{bps/Hz})$	10
发射端流态天线数 L_t	3
接收端流态天线数 L_r	3

Actor 和 critic 网络都是四层网络结构, 包括三个隐藏层, 神经元数分别为: 512、256、128, 并采用 Relu6 激活函数。Actor 网络的输出层表征动作的均值与方差, 其神经元个数均等于动作维度, 分别采用 tanh 和 softplus 激活函数, 利用高斯分布生成动作向量。Critic 网络的输出层仅包含 1 个神经元, 采用线性激活函数来估计价值函数。基于不同权重组合的预仿真与对比, 设置奖励函数权重 w_r 为 10, 其余权重 $\{w_i\}_i$ 均为 0.5, 负值奖励常数均设置为 -1。算法仿真参数如表 2 所示。

针对本文所提出的基于 MW-A2C 的多维资源协同优化策略, 采用两类基线策略进行比较。

第一类策略从模型的角度出发。1) A2C-T:

表2 算法仿真参数

参数名称	参数值
Worker 数量	4
折扣因子 γ	0.95
Actor 网络学习率	5e-4
Critic 网络学习率	1e-3
训练周期	1000
训练优化器	Adam

仅发射端配置流态天线,接收端为传统多天线,发射端天线位置参与优化;2) A2C-N:收发两端均为传统多天线,仅优化无人机轨迹以及用户与目标调度;3) Random:收发端均配置流态天线,但不执行优化策略;4) Random-T:仅发射端配置流态天线,但不执行优化策略;5) Random-N:收发两端均为传统多天线,且不执行优化策略。

第二类策略从无人机轨迹与天线位置优化算法的角度出发,包括:Q学习(Q-Learning, QL)、深度Q网络(Deep Q Network, DQN)、深度确定性策略梯度(Deep Deterministic Policy Gradient, DDPG)、双延迟深度确定性策略梯度(Twin Delayed Deep Deterministic Policy Gradient, TD3),此时的用户与目标调度仍采用贪婪算法。

4.2 仿真结果

图4展示了所提MW-A2C算法在训练过程中,长期通信和速率随训练周期的变化曲线,用以表征收敛性能。其中浅色曲线为原始训练数据,深色曲线为20周期滑动平均结果。可以看出,随着训练的进行,原始曲线呈现先上升后平稳的趋势,但是该曲线始终存在明显波动。这是因为在训练中,智能体通过高斯分布采样动作,以增强对状态-动作空间的探索能力,使得动作组合具有随机性。尤其是在前400个训练周期内,算法偏重于随机探索,奖励值存在较大差异,因此波动相对剧烈;而随着探索率逐渐衰减,后期性能波动逐渐减小,但由于随机动作采样策略的存在,不同周期的动作序列依旧存在差异,从而导致速率数据在收敛阶段仍表现出一定程度的波动。经20周期滑动平均后,探索策略的随机影响受到削弱,曲线呈现出稳定上升趋势,后逐步收敛至350 bps/Hz附近,由此表明所提算法具备良好的收敛性,能够逐步学习到更优的策略。

图4所提MW-A2C算法的收敛曲线

表3比较了不同奖励权重下,算法收敛后的速率性能。可以看出,奖励权重 w_r 过大或者过小都会引发速率性能的降低。该原因在于,过大的奖励缩放会导致优势函数估计方差增大,从而降低算法稳定性并引发策略震荡;而过小的奖励缩放则会导致策略难以有效区分不同动作的优劣。因此,基于上述仿真对比,本文中采用权重 $w_r = 10$ 。此外,为验证所提MW-A2C算法的收敛稳定性,已在不同随机种子下进行重复仿真,且算法均能实现稳定收敛,收敛特性与图4所示结果基本一致。

基于所提算法,图5比较了不同飞行周期数下,无人机在最后5个训练周期的飞行轨迹。结果表明,MW-A2C策略能够根据用户与目标分布情况,自适应的调整无人机飞行轨迹,从而接近用户并获取更高速率性能。同时,在所有情况下,无人机均能在飞行周期结束时成功达到目的地。在图5(a)中,由于飞行时隙有限,无人机仅沿起止位置沿线附近飞行,存在向用户方向飞行的趋势。当飞行周期 $N=30$,无人机具备充分的时间遍历上半区域的用户和目标,但是缺乏对用户2和用户3的调度。这是因为用户1为近用户,无人机在探索过程中容易接近该用户,从而得到正向反馈。具体来说,当无人机飞临用户1上方时,该用户可获得更高的信道增益,由此具有高即时速率与高累积奖励。而针对远用户的搜寻需要耗费大量的时隙与容错概率,且触发几率和累积奖励都需要维持在高水平,才能获得更充分地策略学习。因此,算法容易陷入对局部状态-动作区域的过度学习,使得无人机面临局部范围驻留问题。当飞行周期扩展至 $N=35$ 和 $N=40$ 时,从图5(c)和图5(d)中可以看出,无人机逐渐向用户2和用户3的方向靠近,但是轨迹仍偏重于上半部分区域,说明算法逐渐学习到了更广泛的状态-动作空间,但策略更新中近用户场景的样本占比依然较高。在 $N=45$ 和 $N=50$ 的情况下,无人机在飞行过程中,其接近远用户2和远用户3的概率逐渐上升,算法能够积累到多样化的经验样本,经训练后可以学习到更优的飞行策略。因此图5(e)和图5(f)中,无人机从初始位置起飞,可依次在三个用户附近悬停,而后到达终点。上述结果表明,随着飞行周期 N 的增加,所提MW-A2C协同优化策略可以获得轨迹层面的更优性能,从而兼顾

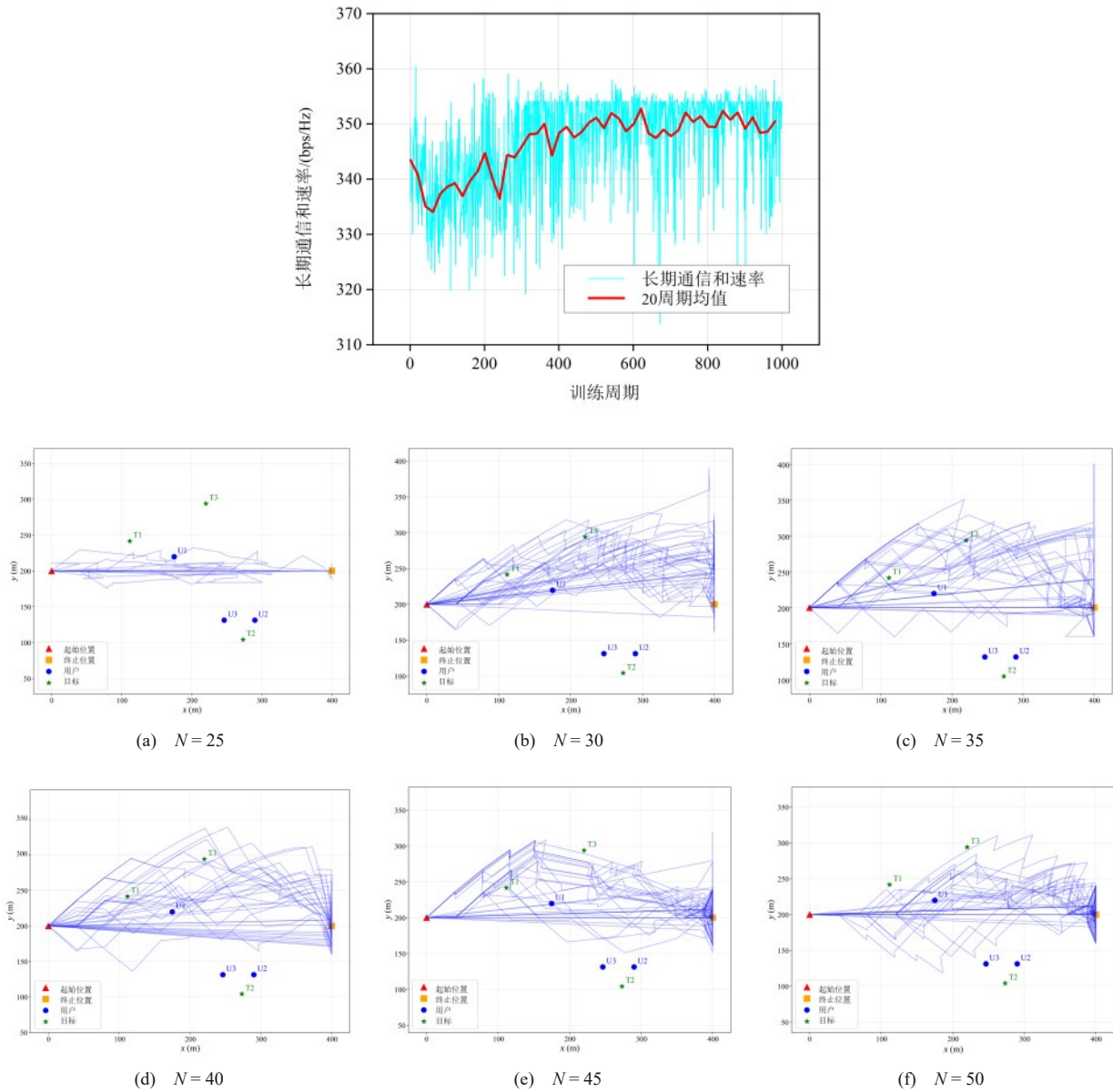


图5 不同飞行周期下无人机的飞行轨迹(5训练周期)

表3 算法收敛后的速率性能对比/(bps/Hz)

奖励权重	$w_r = 1$	$w_r = 5$	$w_r = 10$	$w_r = 15$	$w_r = 20$
长期通信和速率	334.69	338.56	348.89	345.25	335.15

更多的用户。

图6展示了不同的模型策略,其长期通信和速率随飞行周期的变化情况。无论是所提策略还是随机策略,具备流态天线的模型,即A2C和Random,其性能均优于对应的-T策略和-N策略。随着飞行周期 N 的增加,A2C针对于Random的性能增益不断增加,由25时隙的27.22 bps/Hz提升至50

时隙的52.21 bps/Hz。上述原因在于,飞行周期的增加直接导致优化变量数量成比例增长。随机方案在面对更高维的变量空间时性能提升有限,而A2C算法能利用增加的变量进行更精细的策略搜索,从而逐步拉大与随机方案之间的性能差距。此外,在各飞行周期下,相较于-T策略和-N策略的性能增益,流态天线模式引入额外的天线位置变量,同样扩展了A2C的优化空间,因此性能增益最大。上述结果表明,收发端同时部署流态天线可带来显著的性能增益,由此证明了所提模型架构的有效性。

图7展示了不同的模型策略,其长期通信和速

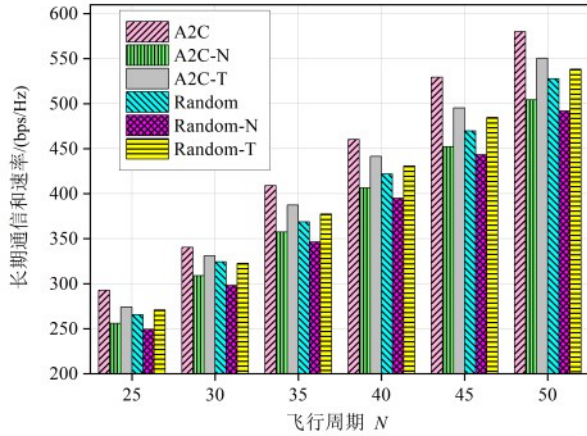


图6 不同模型策略下长期通信和速率与飞行周期间的关系

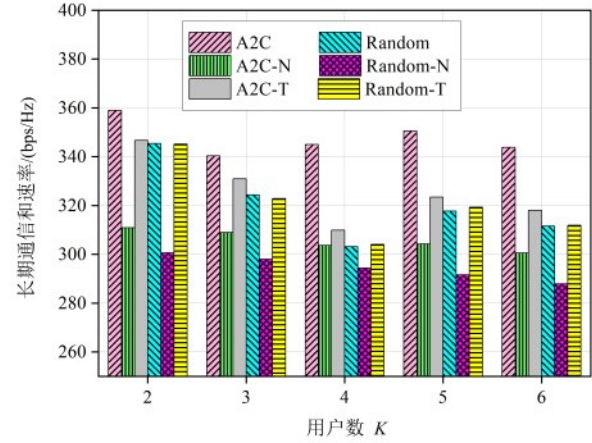


图7 不同模型策略下长期通信和速率与用户数间的关系

率随用户数量的变化情况。由于用户调度基于固定的贪婪策略, 用户数的增加仅提升网络的状态空间, 并不影响 A2C 的动作变量, 因此求解复杂度维持在相近水平, 反而因更丰富的状态信息获得了更多决策可能性, 使得其相较于对比策略, 性能增益的提升更为明显。同时, 各策略随着用户数的增加, 性能差距较小但呈现出波动现象。其原因在于, 一方面, 用户数的增加意味着无人机需遍历更多的用户, 在有限的飞行周期下, 其难以充分接近所有用户以建立高质量链路, 需要在各用户间折中, 从而造成系统通信性能的轻微下降。另一方面, 用户不断增加也意味着更多的接入可能性, 体现在多用户分集带来的信道质量改善, 因此系统性能也会出现提升的趋势。此外, Random 和 Random-T 在通信速率方面的性能相近, 这是因为接收流态天线并未直接影响用户可达速率, 而是参与感知 SINR 的计算, 因此两策略的性能差距体现在目标感知速率。在 6 用户情况下, Random-T 的感知速率均值为 60.47 bps/Hz, 目标间最低感知速率为 7.07 bps/Hz; 而 Random 策略由于具备接收流态天线, 感知速率均值上升至 86.00 bps/Hz, 目标间最低感知速率升至 15.51 bps/Hz。由此可见, 流态天线不仅可以提升通信速率, 而且在感知层面也具有良好的性能增益, 因此适用于 ISAC 网络架构。

图 8 则聚焦于算法层面, 比较了不同的算法的长期通信和速率随飞行周期变化的性能表现。所提 MW-A2C 算法在不同飞行周期下均优于其他算法。DDPG 算法则具备很强的不稳定性, 体现在性能呈现持续性波动, 在 $N=35$ 时与所提算法性能相当, 而当 $N=45$ 时性能甚至不如 Random。这是由于

DDPG 的 actor 网络通常利用 tanh 和 sigmoid 激活函数来构造动作输出, 同时在训练过程中采用 clip 函数来裁剪动作变量, 因此存在大量边界动作样本, 学习过程容易陷入局部最优, 面临边界动作输出问题。在 $N=45$ 时, DDPG 策略中多数的无人机轨迹为从起始位置到终止位置的连线, 无法根据用户与目标位置灵活调整无人机轨迹。TD3 算法与 DDPG 性能相似, 其同样存在因边界动作而造成的不稳定性, 但是 TD3 采用双 critic 网络, 能够缓解动作-值函数的过估计问题, 稳定性略强于 DDPG, 但仍不如所提算法。此外, DQN 和 QL 算法需要将动作空间离散化。例如, 将无人机飞行速度离散化为 3 个等级, 流态天线端口离散化为左右 2 个增量时, DQN 输出层所需要的神经元数为: 4 (可行方向) $\times$ 3 (可行速度) $\times$ 2³ (发射端流态天线可行增量集合) $\times$ 2³ (接收端流态天线可行增量集合) = 768。而 MW-A2C 的 actor 网络输出层神经元数量为: 1 (方向) + 1 (速度) + 3 (发射端流态天线数) + 3 (接收端流态天线数) = 8。因此, DQN 和 QL 需要更多的训练次数, 以完成完整状态-动作空间的学习。所以在相同的 1000 次训练周期下, 两者并未实现性能的收敛, 性能弱于所提 MW-A2C 算法。

图 9 展示了不同的算法策略, 其长期通信和速率随用户数量的变化情况。可以看出, 在用户数 $K=2$ 时, 由于 DQN 面临的状态-动作空间相对较小, 在高采样效率与确定性策略的加成下, 其性能优于所提 MW-A2C 算法。但是随着用户数的增加, 问题复杂度也随之提升, DQN 逐渐难以对应不断扩展的状态-动作空间, 而 MW-A2C 算法则具备良好的探索能力, 因此在用户数 $K \geq 3$ 时, 性能基本优

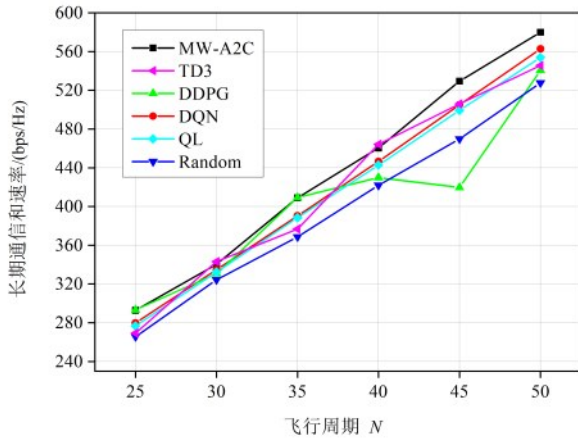


图8 不同算法策略下长期通信和速率与飞行周期间的关系

于其他基线策略。该结果说明, 所提 MW-A2C 策略在复杂场景下具有更好的可扩展性与鲁棒性。

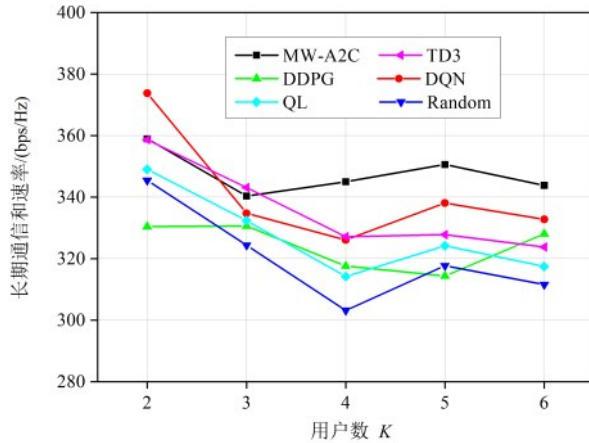


图9 不同算法策略下长期通信和速率与用户数间的关系

表4给出了不同的算法策略, 其总运行时间与用户数间的关系。所提 MW-A2C 算法与 DQN 算法的复杂度相当, 且显著低于 TD3 和 DDPG 算法。综合图8和图9中的性能曲线, 可以证明 MW-A2C 能够在低复杂度下获得更优的速率性能。此外, 随着用户数的增加, 所有算法的总运行时间均呈现增长态势, 由于用户数并不影响动作空间, 仅线性提升状态维度, 因此增长速率较为缓慢。该现象表明所提算法的复杂度对用户规模的敏感度较低, 验证了其在实际应用中的可扩展性。

5 结束语

本文围绕流态天线在无人机 ISAC 系统中的应用, 在收发两端均配置流态天线, 联合无人机轨迹设计、用户与目标调度、发射与接收天线位置选

表4 算法策略的总运行时间对比/s

	MW-A2C	TD3	DDPG	DQN	QL	Random
2用户	50.07	123.00	90.05	52.10	11.56	0.48
3用户	50.91	124.71	90.85	53.07	12.00	0.58
4用户	53.09	126.60	92.10	55.30	12.38	0.70
5用户	53.92	130.92	94.87	55.55	12.62	0.65
6用户	55.28	132.45	95.67	55.43	14.11	0.76

择, 提出了一种基于 MW-A2C 的多维资源协同优化策略。本文通过对子问题的拆分, 有效降低了无人机的计算负担, 在满足目标感知阈值的同时, 实现了系统总和通信性能的优化求解。仿真结果表明, 所提 MW-A2C 策略能够基于用户与目标位置, 灵活调整无人机飞行轨迹, 并且相比于 A2C-T、A2C-N、TD3、DDPG、DQN、QL 等基线策略, 在各种传输环境中, 均表现出更优的性能增益与鲁棒性, 同时验证了流态天线在 ISAC 系统中的优势。后续将进一步引入波束赋形设计问题, 基于多智能体强化学习, 探究面向多无人机协作网络的资源管理策略。

参考文献:

- [1] Liu F, Cui Y H, Masouros C, et al. Integrated sensing and communications: toward dual-functional wireless networks for 6G and beyond[J]. IEEE Journal on Selected Areas in Communications, 2022, 40(6): 1728-1767.
- [2] Lu S H, Liu F, Li Y X, et al. Integrated sensing and communications: Recent advances and ten open challenges[J]. IEEE Internet of Things Journal, 2024, 11(11): 19094-19120.
- [3] 冉鑫怡, 陈前斌, 徐勇军等. 基于深度学习的通感一体化系统综述[J]. 通信学报, 2025, 46(06): 233-250.
RAN X Y, CHEN Q B, XU Y J, et al. Survey on deep learning-based integrated sensing and communication systems[J]. Journal on Communications, 2025, 46(6): 233-250.
- [4] Meng K T, Wu Q Q, Xu J, et al. UAV-enabled integrated sensing and communication: Opportunities and challenges[J]. IEEE Wireless Communications, 2024, 31(2): 97-104.
- [5] Lyu Z, Zhu G, Xu J. Joint maneuver and beamforming design for UAV-enabled integrated sensing and communication[J]. IEEE Transactions on Wireless Communications, 2023, 22(4): 2424-2440.
- [6] Mu J S, Zhang R H, Cui Y H, et al. UAV meets integrated sensing and communication: Challenges and future directions[J]. IEEE Communications Magazine, 2023, 61(5): 62-67.
- [7] Wong K K, Shojacifard A, Tong K F, et al. Fluid antenna systems[J]. IEEE Transactions on Wireless Communications, 2021, 20(3): 1950-1962.

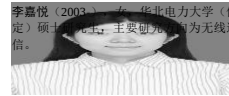
- [8] New W K, Wong K K, Xu H, et al. A tutorial on fluid antenna system for 6G networks: Encompassing communication theory, optimization methods and hardware designs[J]. IEEE Communications Surveys & Tutorials, 2025, 27(4): 2325-2377.
- [9] New W K, Wong K K, Wang C, et al. Fluid antenna systems: Redefining reconfigurable wireless communications[J]. IEEE Journal on Selected Areas in Communications, 2026, 44: 1013-1044.
- [10] Chen Y, Xu B, Li S, et al. Analysis and optimization for low-latency communications in slow fluid antenna multiple access systems[J]. IEEE Journal on Selected Areas in Communications, 2026, 44: 1290-1306.
- [11] Zhang C, Wang J, You L, et al. Fluid XL-MIMO: Multiuser antenna allocation exploiting the visibility region information[J]. IEEE Wireless Communications Letters, 2026, 15: 520-524.
- [12] Liang S Y, Zhu C, Yang Z H, et al. Rate maximization for fluid antenna system assisted semantic communication[J]. IEEE Transactions on Vehicular Technology, 2026, 75(7): 15329-15333.
- [13] Wu T, Yao J T, Zheng J C, et al. Unleashing more potential from FAS: A framework of FAS-CoNOMA systems[J]. IEEE Transactions on Communications, 2026, 74: 4820-4836.
- [14] Zou J Q, Xu H, Wang C, et al. Shifting the ISAC trade-off with fluid antenna systems[J]. IEEE Wireless Communications Letters, 2024, 13(12): 3479-3483.
- [15] Lou X, Xia W, Zhu Y, et al. Multi-target beamforming optimization for fluid antenna-enabled multi-static ISAC[J]. IEEE Transactions on Cognitive Communications and Networking, 2026, 12: 1382-1394.
- [16] Zhang Z, Wong K K, Dang J, et al. On fundamental limits for fluid antenna-assisted integrated sensing and communications for unsourced random access[J]. IEEE Journal on Selected Areas in Communications, 2026, 44: 136-149.
- [17] Tang B, Xu H, Wong K K, et al. Full-duplex FAS-assisted base station for ISAC[J]. IEEE Transactions on Wireless Communications, 2026, 25: 2922-2938.
- [18] Jiang H, Shi W, Chen Z, et al. Dynamic channel modeling of fluid antenna systems in UAV communications[J]. IEEE Wireless Communications Letters, 2025, 14(10): 3169-3173.
- [19] Abuzgaia N, Salem A, Elbarsha A. FAS enabled UAV for energy-efficient WPCNs[J]. IEEE Communications Letters, 2026, 30: 278-281.
- [20] Xu X R, Xu H, Wei D Y, et al. Transformer-based collaborative reinforcement learning for fluid antenna system (FAS)-enabled 3D UAV positioning[J]. IEEE Journal on Selected Areas in Communications, 2026, 44: 1128-1143.
- [21] Reda A, Mekawy T, Tsiftsis T A, et al. Toward joint optimization for UAV-integrated RIS-assisted fluid antenna systems[J]. IEEE Transactions on Vehicular Technology, 2026, 75(7): 13463-13473.
- [22] Zhang X H, Liu W C, Wang C J, et al. UAV-enabled fluid antenna systems for multi-target wireless sensing over LAWCNs[J]. IEEE Transactions on Consumer Electronics, 2026, 72(1): 1146-1155.
- [23] Fan J Y, Tang J, Song X K, et al. Fluid antenna-assisted ISAC design for emerging low-altitude scenarios[J]. IEEE Transactions on Cognitive Communications and Networking, 2026, 12: 8536-8548.
- [24] Wu M, Lai S, Lu W, et al. Rethinking fluid antennas for ISAC: A systematic review and a novel air-ground collaborative framework[J]. IEEE Transactions on Network Science and Engineering, 2026, 13: 3682-3700.
- [25] Liu W, Zhang X, Ren J, et al. UAV-enabled ISAC with fluid antennas for low-altitude wireless networks[J]. IEEE Transactions on Communications, 2026, 74: 9529-9546.
- [26] Meng K T, Wu Q Q, Ma S D, et al. Throughput maximization for UAV-enabled integrated periodic sensing and communication[J]. IEEE Transactions on Wireless Communications, 2023, 22(1): 671-687.
- [27] 3GPP. Technical specification group radio access network; Study on enhanced LTE support for aerial vehicles: 3GPP TR 36.777[R]. 2018.



邓丹昊(1996-), 女, 博士, 华北电力大学(保定)讲师、硕士生导师, 主要研究方向为无线通信、空天地一体化、通感算一体化、无线资源管理。



杨俊(2002-), 男, 华北电力大学(保定)硕士研究生, 主要研究方向为无线通信。



李嘉悦(2003-), 女, 华北电力大学(保定)硕士研究生, 主要研究方向为无线通信。

刘大维（2001-），男，华北电力大学（保定）硕士研究生，主要研究方向为无线体域网通信。

