

引用格式: 李新民, 张徐浩, 刘易阳, 等. 基于强化学习的无人机通信感知一体化能耗优化 [J]. 电子科技大学学报, 2025, 54(1): 23-28.
LI X M, ZHANG X H, LIU Y Y, et al. Energy optimization algorithm for ISAC-enabled unmanned aerial vehicles system via reinforcement learning[J].
Journal of University of Electronic Science and Technology of China, 2025, 54(1): 23-28.

基于强化学习的无人机通信感知一体化 能耗优化



李新民^{1,2}, 张徐浩², 刘易阳², 曹 辉^{2*}, 张晓强²

(1. 成都大学 计算机学院, 成都 610100; 2. 西南科技大学 信息工程学院, 绵阳 621000)

摘要: 通信感知一体化技术与灵活可控无人机通信技术结合将成为 6G 无线网络“万物智联”的潜在技术。针对无人机通信感知一体化系统中无人机能量受限问题, 提出了一种基于强化学习的无人机功率分配和轨迹设计的联合优化算法。该算法在用户通信速率和目标感知波束增益约束下, 通过构建与无人机能耗、发射波束和通信速率相关的线性加权奖励函数, 以实现智能化的无人机功率分配和轨迹设计, 从而最小化无人机通信感知一体化系统的能耗。仿真结果表明, 该方案相较于基准方案降低了 12.36%~21.08% 的无人机能耗, 并拥有更优的收敛性能。

关键词: 通信感知一体化; 无人机; 强化学习; 能耗优化

中图分类号: TN92

文献标志码: A

DOI: 10.12178/1001-0548.2023331

Energy optimization algorithm for ISAC-enabled unmanned aerial vehicles system via reinforcement learning

LI Xinmin^{1,2}, ZHANG Xuhao², LIU Yiyang², CAO Hui^{2*}, and ZHANG Xiaoqiang²

(1. College of Computer Science, Chengdu University, Chengdu 610100, China;

2. School of Information Engineering, Southwest University of Science and Technology, Mianyang 621000, China)

Abstract: Integrated sensing and communication (ISAC) technology combined with the cost-effective and flexibly controllable unmanned aerial vehicle (UAV) becomes the potential technology to enable a variety of applications for the future "Internet of Everything" in the sixth generation (6G) communication system. To reduce the energy consumption of ISAC-enabled UAV systems, a joint optimization algorithm based on reinforcement learning (RL) is proposed to design UAV's trajectory and allocate the transmit power. Under constraints of user communication rate and the target sensing beam pattern gain, this algorithm can achieve intelligent decision-making for UAV trajectory and power allocation by constructing a linearly weighted reward function related to UAV energy consumption, transmit beamforming pattern gain, and communication rate. The simulation results indicate that the proposed scheme can reduce energy consumption by 12.36% to 21.08% in comparison to the benchmark schemes. Furthermore, the proposed scheme demonstrates superior convergence.

Key words: integrated sensing and communication; unmanned aerial vehicle; reinforcement learning; energy optimization

未来第六代 (the sixth generation, 6G) 无线网络在支撑极致通信体验的同时, 能够实现高精度的信息感知、海量节点的万物智联以及空天地一体的高效无缝覆盖等功能^[1]。为了满足多样的业务形态和极致的通信感知性能需求, 无线通信与雷达感知技术的融合已经成为 6G 研究的重要

方向^[2]。通信感知一体化 (integrated sensing and communication, ISAC) 技术旨在提高频谱效率, 实现 6G 无线网络中通信和感知物理基础设备的共享, 作为一种集成无线通信和雷达感知能力, 实现两者功能融合互助的关键技术, 引起了工业界和学术界的广泛关注^[3]。

收稿日期: 2023-12-27

基金项目: 国家自然科学基金 (62201479); 四川省科技计划 (2023NSFSC1388)

作者简介: 李新民, 博士, 主要从事 6G 通信系统、无人机通信、智能算法等方面的研究。

*通信作者 E-mail: Caoh@swust.edu.cn

通信感知一体化技术的研究工作目前主要集中在传统的地面网络, 针对 ISAC 系统在基础理论、信号处理、网络架构、通信安全等方面进行研究^[4]。然而当前的移动通信网络面临着两个固有的感知局限性^[5]。首先是地面的障碍物可能会遮挡视距链路 (line of sight, LoS), 从而导致感知变得困难甚至不可行。其次, 从远处目标到地面接入点的回波信号的严重传播损耗, 使得感知性能下降。而无人机具有广域覆盖、灵活部署、视距通信等优点^[6], 其灵活可控的特性为飞行轨迹设计带来了方便, 也为通信感知一体化系统带来了全新的设计视角。无人机与通信感知一体化系统的结合, 为 6G 无线网络设计带来了更大的弹性, 提升了网络的内生感知能力, 对网络智能等研究工作具有重要推动意义。在无人机灵活部署和通信增强等优势推动下, 有望彻底改变现有 ISAC 网络系统架构, 实现更加灵活高效的联合设计^[7]。

无人机通常是能量受限的, ISAC 辅助的无人机通信系统往往需要联合优化轨迹设计和通信资源分配, 以此提高能量的使用效率, 提升系统的服务质量^[8]。在支持灾区紧急通信背景下, 文献 [9] 针对多无人机的按需部署和通信连接保持提出了多无人机协作轨迹优化问题, 通过马尔可夫博弈转化, 提出了多智能体强化学习的分布式轨迹优化算法。在多无人机协作目标感知场景中, 文献 [10] 为了平衡无人机的雷达感知和通信性能, 联合优化了无人机的发射功率、资源分配、飞行轨迹, 以最大化感知性能和资源分配的均衡性。文献 [11] 联合优化无人机的目标感知和地面用户的功率控制, 在概率视距信道下最大化速率。文献 [12] 研究了支持射频和光通信的多无人机实时协作轨迹设计问题, 以提升部署效率和最小化交互开销为目标, 将其表述为分布式协作优化问题进行求解。同时, 通信和感知性能对通信感知一体化系统而言也十分重要, 所以通信和感知性能的折中也是系统设计的关键, 文献 [13] 将无人机作为空中双功能接入点, 利用无人机的高机动性和强视距链路来实现高效的感知和通信。

此外, 无人机与通信感知一体化系统的融合也带来了更多的挑战, 对通信时延、感知精度、保密性能等提出了更高的要求。文献 [14] 提出了一种无人机辅助通信感知一体化车联网的多感知数据融合的节能计算卸载策略, 显著降低了无人机能耗和数据融合任务的延迟。针对无人机通信系统的用户

隐私要求, 文献 [15] 讨论了无人机到地面用户通信和感知信号的物理层安全问题。针对集成了通信和感知功能的无人机在功能需求上的不对称性, 文献 [16] 提出了一种自适应 ISAC 机制, 通过灵活配置无人机按需进行感知, 避免了频谱资源的浪费, 提高了系统性能和资源利用率。目前, 针对无人机通信感知一体化系统的研究工作, 主要涉及波束设计、通感性能优化、数据处理算法等。然而目前针对 ISAC 辅助的无人机设备能量受限问题的研究较少, 但该问题是在 6G 通信感知智能网络中广泛应用无人机场景的重大挑战, 所以在通信感知一体化系统设计中降低无人机能耗十分必要。

本文研究了无人机通信感知一体化系统中无人机能量受限的问题。在该系统中无人机服务地面用户并且对潜在的空中目标进行感知。首先建立以最小化无人机能耗为优化目标, 以通信速率和感知波束增益为约束的系统模型, 然后在保证无人机通信和感知性能的前提下, 将优化目标解构为无人机飞行轨迹和功率分配联合优化问题。然而, 在无人机动态环境中能耗优化问题通常难以求解, 本文将该问题转化为马尔可夫决策过程, 采用强化学习方法求解。具体而言, 将无人机建模为智能体, 形成与无人机能耗相关的加权奖励函数, 使无人机能够自主性的改进策略, 实现轨迹和功率分配的智能化决策, 从而实现最小化无人机能耗的目标。

1 系统模型和问题形成

如图 1 所示, 本文考虑一个 ISAC 辅助的单无人机系统。该系统由配备 M 根天线的均匀线性阵列单无人机基站、 K 个地面用户和 N 个固定高度为 H_0 的空中目标组成, 空中目标通常是需要监视的外来侵入无人机或其他执行空中作业任务的无人机。无人机的总任务时长 T 被划分为 L 个 ISAC 子帧, 每个子帧具有 T_f 个时隙, 包含通信时隙和感知时隙两部分, 每个时隙的时长为 $\delta_t = T/(LT_f)$, 且通信所占时隙多于感知时隙总时隙数 $N = T/\delta_t$ 。本文采取三维笛卡尔坐标系, 其中第 n 个空中目标和第 k 个地面目标的水平坐标分别被固定在 $\tilde{\mathbf{u}}_n = [\tilde{x}_n, \tilde{y}_n]^T$ 和 $\mathbf{u}_k = [x_k, y_k]^T$ 。无人机在时隙 t 个时隙的水平坐标可以表示为 $\mathbf{q}_t = [q_{x_t}, q_{y_t}]$, 其飞行高度为 H , 同时 $\mathbf{q}^I = [x^I, y^I]$ 和 $\mathbf{q}^F = [x^F, y^F]$ 分别表示无人机的任务起点和终点的水平坐标, 因此, 无人机的位置约束可以表示为:

$$q_0 = \bar{q}^1, q_T = q^F \quad (1)$$

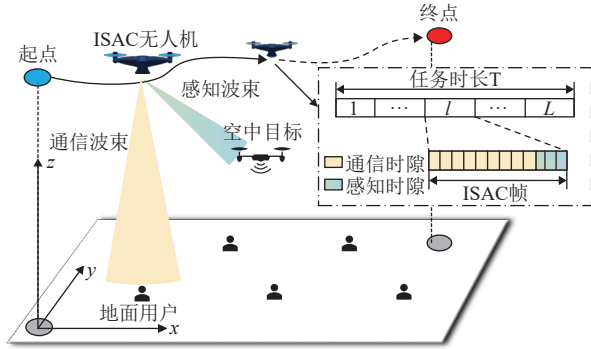


图1 无人机通信感知一体化系统图

无人机由于自身体积、载荷和功耗等约束, 通常在频谱、硬件和能量资源上是受限的。用户调度和资源分配会占用无人机大量的计算资源和系统能量, 本文假定无人机以轮询方式公平有序的为用户分配时间和频谱资源^[17]。为了提高无人机的感知精度, 假定无人机感知空中目标时处于悬停状态。因为感知周期一般较小, 固定无人机位置可以避免运动引入的复杂多普勒频移, 提升感知精度。无人机和地面用户之间的通信链路由视距链路主导, 遵循自由空间路径损耗模型, 从无人机到地面用户的信道功率增益可以表示为^[18]:

$$\beta_k(\mathbf{q}_t, \mathbf{u}_k) = \beta_0 d(\mathbf{q}_t, \mathbf{u}_k)^{-2} = \frac{\beta_0}{H^2 + \|\mathbf{q}_t - \mathbf{u}_k\|^2} \quad (2)$$

式中, β_0 表示参考距离为1 m处的信道功率增益。

无人机对第 k 个用户位置处的发射阵列响应向量为:

$$\mathbf{a}^H(\mathbf{q}_t, \mathbf{u}_k) = \left[1, e^{-j2\pi \frac{d}{\lambda} \sin(\theta(\mathbf{q}_t, \mathbf{u}_k))}, \dots, e^{-j2\pi \frac{d}{\lambda} (M-1) \sin(\theta(\mathbf{q}_t, \mathbf{u}_k))} \right] \quad (3)$$

式中, d 是半波长天线间距; λ 是载波波长; θ 是无人机和用户地理路径的仰角, $\sin(\theta(\mathbf{q}_t, \mathbf{u}_k)) = \frac{H}{\sqrt{\|\mathbf{q}_t - \mathbf{u}_k\|^2 + H^2}}$ 。无人机和第 k 个用户之间的基带等效信道可以表示为:

$$\mathbf{h}_k^H(\mathbf{q}_t, \mathbf{u}_k) = \sqrt{\beta_k(\mathbf{q}_t, \mathbf{u}_k)} e^{-j \frac{2\pi d(\mathbf{q}_t, \mathbf{u}_k)}{\lambda}} \mathbf{a}^H(\mathbf{q}_t, \mathbf{u}_k) \quad (4)$$

无人机在时隙 t 的基带发射信号为:

$$\mathbf{x}[t] = \mathbf{w}_c[t] c_k[t] \quad 1 \leq k \leq \mathcal{K} \quad (5)$$

式中, $c_k[t]$ 是无人机向第 k 个地面用户发射的信号, $c_k \sim CN(0, 1)$; $\mathbf{w}_c[t] \in \mathbb{C}^{M \times 1}$ 是波束成形向量。

因此, 第 k 个地面用户处的接收信号 $y_k(t)$ 可以表示为:

$$y_k(t) = \mathbf{h}_k^H(\mathbf{q}_t, \mathbf{u}_k) \mathbf{x}[t] + n_k[t] \quad (6)$$

式中, $n_k[t] \sim CN(0, \sigma_k^2)$ 表示第 k 个用户处的加性高斯白噪声。

采用最大比发送预编码的传输波束成形向量的表达式为 $\mathbf{w}_c[t] = \frac{\sqrt{p_t} \mathbf{h}_k^H(\mathbf{q}_t, \mathbf{u}_k)}{\|\mathbf{h}_k^H(\mathbf{q}_t, \mathbf{u}_k)\|}$, 无人机在时隙 t 的发射功率可以表示为:

$$P_t[t] = \|\mathbf{w}_c[t]\|^2 \quad (7)$$

在时隙 t 第 k 个地面用户的信噪比可表示为:

$$\gamma_k[t] = \frac{|\mathbf{h}_k^H(\mathbf{q}_t, \mathbf{u}_k) \mathbf{w}_c[t]|^2}{\sigma_k^2} \quad (8)$$

因此在时隙 t 第 k 个用户的通信速率可以表示为:

$$R_k[t] = \log_2(1 + \gamma_k[t]) \quad (9)$$

在设计ISAC系统中, 由目标反射的通信信号也可用于空中目标参数估计, 信号 $c_k[t]$ 可以被进一步用于执行感知任务^[19]。从无人机到空中目标 $\tilde{\mathbf{u}}_n$ 的发射波束增益可以表示为:

$$\Gamma(\mathbf{q}_t, \tilde{\mathbf{u}}_n) = E \left[\left| \mathbf{a}^H(\mathbf{q}_t, \tilde{\mathbf{u}}_n) \mathbf{x}[t] \right|^2 \right] = \mathbf{a}^H(\mathbf{q}_t, \tilde{\mathbf{u}}_n) (\mathbf{w}_c[t] \mathbf{w}_c^H[t]) \mathbf{a}(\mathbf{q}_t, \tilde{\mathbf{u}}_n) \quad (10)$$

旋翼无人机的功率直接取决于无人机的工作模式^[20]。无人机在时隙 t 的功耗可以分为悬停功耗和飞行功耗两种, 悬停时执行感知任务, 飞行时执行通信任务。飞行功耗是其飞行速度 $\mathbf{v}[t] \in \mathcal{R}^{2 \times 1}$ 的函数, 不同工作模式下的无人机功耗可以表示为:

$$P_u[t] = \begin{cases} P_h & \text{感知任务} \\ P_f & \text{通信任务} \end{cases} \quad (11)$$

式中, 无人机的悬停功率 $P_h = P_o + P_i$, P_o 和 P_i 是取决于无人机重量、空气密度、旋翼角速度和面积等参数的气动功率, $P_o = \frac{\delta}{8} \rho s A \Omega^3 R^3$, $P_i = (1 + \kappa) \frac{W^{3/2}}{\sqrt{2\rho A}}$; P_f 是无人机的飞行功率, $P_f = P_o \left(1 + \frac{3\|\mathbf{v}[t]\|^2}{Q^2 r^2} \right) + P_i \left(\sqrt{1 + \frac{\|\mathbf{v}[t]\|^4}{4v_0^4}} - \frac{\|\mathbf{v}[t]\|^2}{2v_0^2} \right) + \frac{1}{2} r_0 \rho s A_r \|\mathbf{v}[t]\|^3$ 。

无人机飞行功耗模型的参数详见文献[20]。无人机飞行到时刻 t 的能耗为 $E_t = \sum_{i=1}^t (p_t[i] + p_u[i])$, 因此无人机在整个任务期间的能耗可以表示为:

$$E = \sum_{t=1}^{N_T} E_t \quad (12)$$

因此, 优化问题可以表述为:

$$\begin{aligned} \text{(P1): } & \min_{\{\mathbf{w}_c[t], \mathbf{q}_t\}} E \\ \text{s.t. C1: } & \mathbf{q}_0 = \tilde{\mathbf{q}}^I, \mathbf{q}_T = \mathbf{q}^F \\ & \text{C2: } \Gamma(\mathbf{q}_t, \tilde{\mathbf{u}}_n) \geq d(\mathbf{q}_t, \tilde{\mathbf{u}}_n)^2 \tilde{\Gamma}, 1 \leq n \leq N, 1 \leq t \leq N_T \\ & \text{C3: } \|\mathbf{w}_c[t]\|^2 \leq P_{\max}, 1 \leq t \leq N_T \\ & \text{C4: } \log_2(1 + \gamma_k[t]) \geq R_{\min}^k, 1 \leq k \leq \mathcal{K} \end{aligned} \quad (13)$$

式中, C1 是无人机的初始和最终位置约束。C2 确保目标方向所需的波束图增益和发射信号的实际波束图增益之间的差异不超过预定阈值, 其中 $\tilde{\Gamma}$ 表示波束图增益的阈值, $d(\mathbf{q}_t, \tilde{\mathbf{u}}_n)^2$ 用于表征对应的路径损耗。C3 是无人机的发射功率约束, 其中 P_{\max} 是无人机的最大发射功率。C4 保证地面用户的通信速率最低不小于预定阈值 R_{\min}^k 。

2 算法表述

无人机能耗优化问题 (P1) 是非凸的, 通常难以获得最优解。因此, 为了求解上述问题, 实现无人机轨迹和功率分配决策的智能化设计, 本文采用强化学习 Q-Learning 算法将问题建模为马尔可夫决策过程。Q-Learning 算法中智能体通过与环境的交互不断学习积累经验, 从而提升在马尔可夫决策过程中的决策能力, 该算法可用于解决通信网络中的路由控制、功率分配、部署优化、动态频谱访问等问题^[21]。马尔可夫决策过程由五元组 $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$ 构成, 其中状态空间 \mathcal{S} 是所有状态的集合, 动作空间 \mathcal{A} 是所有动作的集合, 状态转移概率 \mathcal{P} 表示从当前状态转移到下一状态的概率, 奖励函数 \mathcal{R} 用于衡量当前状态下所采取动作的价值, 折扣因子 γ 反映了所处状态下未来奖励的重要性。智能体在时隙 t 的状态可以表示为 $\mathbf{s}_t \in \mathcal{S}$, 然后在当前状态下选择一个动作 $\mathbf{a}_t \in \mathcal{A}$, 智能体根据所选动作与环境进行交互并得到反馈, 智能体会收到一个数值化的即时奖励 $r_{t+1} \in \mathcal{R}$, 同时进入一个新的状态 \mathbf{s}_{t+1} 。这一系列的状态和动作序列构成了智能体的策略 $\pi: \mathcal{S} \rightarrow \mathcal{A}$, 通过输入当前的状态从而输出最优动作, 实现从状态到动作的映射。强化学习的目标不是最大化即时奖励, 而是最大化长期回报

$g_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}$, 为了最大化 g_t , 定义动作价值函数 $Q_{\pi}(\mathbf{s}_t, \mathbf{a}_t) = \mathbb{E}_{\pi}[g_t | \mathbf{S} = \mathbf{s}_t, \mathbf{A} = \mathbf{a}_t]$, 表示智能体根据策

略 π , 从状态 \mathbf{s}_t 开始执行动作 \mathbf{a}_t 后所有可能决策序列的回报期望值。解决一个强化学习问题意味着找到一个最优策略, 使得长期回报值最大, 其对应的最优动作价值函数可以表示为:

$$Q_* = \max_{\pi} Q_{\pi}(\mathbf{s}_t, \mathbf{a}_t) \quad \forall \mathbf{s}_t \in \mathcal{S}, \forall \mathbf{a}_t \in \mathcal{A} \quad (14)$$

本文将 (P1) 问题映射为马尔可夫决策过程, 并将无人机作为 Q-learning 算法的智能体。定义学习率 α 和贪婪系数 $\epsilon \in (0, 1)$, 分别用于决定新旧经验值间的利用程度和平衡算法在探索与利用之间的权衡, 无人机基于 ϵ -greedy 策略进行动作选择。

1) 状态 \mathbf{s}_t : 无人机执行决策的过程中需要感知环境状态, 状态信息由无人机的位置坐标构成。无人机在时隙 t 的状态可以表示为 $\mathbf{s}_t = [q_x, q_y]$ 。

2) 动作 \mathbf{a}_t : 无人机执行动作空间中的动作与环境交互, 从而实现状态的转换。无人机在时隙 t 的决策动作包含飞行方向和发射功率两部分。飞行方向动作空间可以表示为 $A_{\theta} = \left[\frac{2\pi}{l_{\theta}}, 2\frac{2\pi}{l_{\theta}}, \dots, 2\pi \right]$, 其中 l_{θ} 表示无人机飞行方向空间的长度。发射功率动作空间可以表示为 $A_p = \left[\frac{P_{\max}}{l_p}, \frac{2P_{\max}}{l_p}, \dots, P_{\max} \right]$, 其中 l_p 代表发射功率动作空间的长度。因此, 智能体的动作可以表示为: $\mathbf{a}_t = [a_{\theta}, a_p]$, $a_{\theta} \in A_{\theta}$, $a_p \in A_p$ 。

3) 奖励 r_t : 为了使无人机能够自主求解最佳飞行轨迹和发射功率, 将采集的通信、感知和自身能耗信息经过运算后作为奖励反馈给无人机, 通过反复迭代训练, 无人机学得最佳动作策略, 能够选取长期收益最大的动作。无人机在时隙 t 的奖励为:

$$r_t = c + \lambda_1(\Phi + \lambda_2 E_r) \quad (15)$$

式中, $c \in \mathcal{R}$ 是常数系数; 参数 $\lambda_1 \in [0, 1]$ 为系统的能耗惩罚系数, 无人机总能量为 E_{\max} , 则无人机在时隙 t 的剩余能量为 $E_{r,t} = E_{\max} - E_t$, 若剩余能量 $E_{r,t} \leq E_l$ 则 $\lambda_1 = 0$, 反之 $\lambda_1 = 1$, E_l 是保证无人机安全工作所需要的最低能量值; 参数 $\lambda_2 \in [0, 1]$ 为服务惩罚系数, 若无人机服务完所有用户则 $\lambda_2 = 1$, 反之 $\lambda_2 = 0$, Φ 是无人机已服务用户的计数器。

无人机在训练过程中需要对给定策略计算 Q 值函数并更新策略, 通过不断迭代更新 Q 值表, 从而使算法收敛到最优动作策略。Q 值函数更新规则可以表示为:

$$Q_{\pi}(s_t, a_t) = Q_{\pi}(s_t, a_t) + \alpha[r_t + \gamma \max_{\tilde{a}} Q_{\pi}(s_{t+1}, \tilde{a}) - Q_{\pi}(s_t, a_t)] \quad (16)$$

本文通过强化学习算法, 将无人机路径规划与功率分配问题转化为马尔可夫决策过程, 智能体通过持续学习积累经验, 逐渐逼近最优解。通过将 Q-learning 的奖励函数设计为采集的通信感知和自身能耗信息的线性组合, 在能耗和服务质量之间找到平衡, 提升了任务执行效率和能量利用率。基于 Q-learning 的无人机能耗最小化算法步骤如下。

算法 1 基于 Q-Learning 的无人机能耗最小化方法

初始化: Q-Learning 参数 $\{\alpha, \gamma, \epsilon\}$, 最大训练次数 M_{\max} , 无人机能量 E_{\max} 和 E_l , $Q(s_t, a_t)$ 值;

for episode $i = 1: M_{\max}$ do

 获得初始状态 s_0 ;

 while $E_{r,t} \geq E_l$ do

 if $\Phi \leq K$ then

 由 ϵ 贪婪策略选择当前状态 s_t 下的动作 a_t , 依据式 (15) 获取奖励 r_t , 由式 (12) 计算无人机剩余能量 $E_{r,t}$, 由式 (16) 更新 Q 值;

 else

 由式 (12) 计算系统能耗 E ;

 end if

$s_t \leftarrow s_{t+1}$, $a_t \leftarrow a_{t+1}$, $t \leftarrow t+1$;

 end while

end for

上述算法的时间复杂度为 $O(L|S||A|M_{\max})$ ^[22], 其中 L 是无人机从 E_{\max} 到 E_l 的最大决策步数, $|S|$ 和 $|A|$ 是状态列表和动作列表的列表长度, M_{\max} 是最大训练次数。

3 仿真结果与分析

本文将最大功率方案与贪婪方案作为基准方案, 通过不同参数设置与本文方案进行对比, 以验证所提方案的有效性和稳定性。仿真平台为 Python3.9, CPU 为 Intel i7-8750H@2.20GHz, 设定无人机在 $350 \text{ m} \times 350 \text{ m}$ 的规定区域内飞行, 无人机飞行速度为 20 m/s , 飞行高度固定为 50 m , 无人机的起点水平位置坐标和终点水平位置坐标分别为 $[50, 50] \text{ m}$ 和 $[300, 300] \text{ m}$ 。在飞行区域内随机分布着 4 个需要服务的地面通信用户和 2 个潜在的空中感知目标。训练阶段算法的迭代次数为 400 次, 仿真参数详细设置如表 1 所示。

表 1 仿真参数

符号	参数描述	值
M	天线数	6
H_0	空中目标飞行高度/m	35
α	学习率	0.1
ϵ	贪婪系数	0.9
γ	折扣系数	0.9
E_{\max}	无人机最大能量/J	3.5×10^6
P_{\max}	无人机最大发射功率/dBm	30
σ^2	噪声功率/dBm	-80
\bar{r}	波束图增益阈值 ^[19]	$1 \times e^{-6}$
R_{\min}^k	最小通信速率/bps/Hz ⁻¹ ^[19]	1
δ_t	时隙长度/s	0.2
l_p	功率动作空间长度	10
l_{θ}	飞行动作空间长度	8
T	任务周期时长/s	500
L	ISAC子帧数目	500

在单个无人机执行通信感知一体化任务的情景中, 无人机基于 Q-learning 算法训练多次后的飞行轨迹如图 2 所示。无人机从起点位置开始飞行, 需要在能量有限的条件下服务每个地面用户, 并且感知潜在的空中目标, 完成任务或者能量到达安全飞行的最低限度时自动飞回终点位置。无人机执行一次动作后主动收集相关信息, 同时根据奖励机制进行反馈奖励, 从而实现无人机自主飞行轨迹决策。图 2 中也描述了基于最大功率方案和贪婪方案的无人机轨迹。为了满足设定的通信和感知阈值需求, 同时提高其通信和感知质量, 本文方案中无人机尽可能地为用户靠近以获得更好的链路质量, 但为了减少系统能耗, 无人机并不飞达用户正上方, 相比于贪婪方案而言无人机更加智能化决策飞行轨迹。

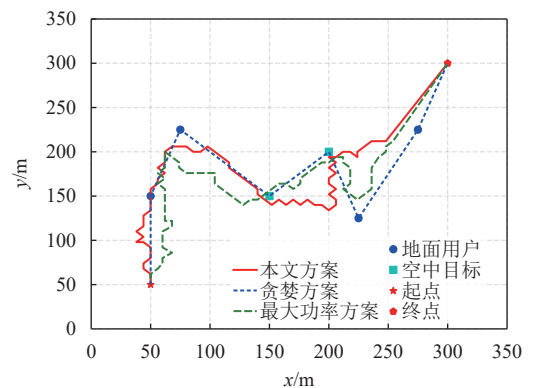


图 2 无人机的飞行轨迹

图 3 展示了 4 种方案下无人机能耗和训练次数的关系。从仿真结果可知, 随着训练次数的增加所提方案的能耗值迅速降低, 表明本文方案能够降低无人机的能耗, 实现无人机自主飞行轨迹决策。算

法收敛时, 本文方案的能耗为 4 576.78 J、而最大功率方案和贪婪方案的能耗分别为 5 222.57 J 和 5 799.49 J。相比于最大功率方案, 所提方案收敛性更加优越, 降低了无人机系统 12.36% 的系统能耗, 相比于贪婪方案, 能降低 21.08% 的系统能耗。而 DQN 方案呈现较差的收敛特性, 这可能是由于 DQN 算法在训练过程中出现了过拟合现象。观察图中曲线发现 DQN 算法并未快速收敛且存在较大的波动性, 在有限的训练次数中 DQN 算法性能不够好。具体而言 DQN 算法在 300 次训练后能耗有效降低, 但在 300~400 次训练后持续出现波动。相较于本文算法, 能耗值很高且不稳定。

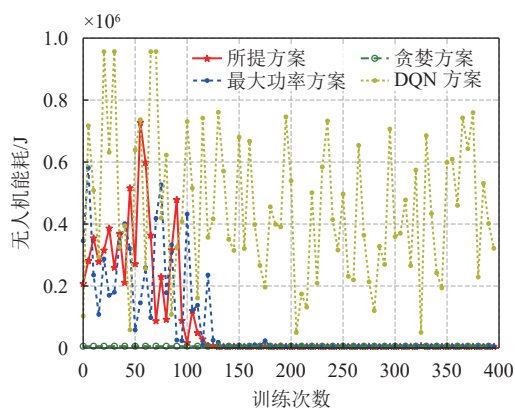


图 3 无人机能耗和训练次数的关系

图 4 为仿真条件都相同的情况下, 使用本文所提强化学习 Q-learning 方案在不同学习率条件下的能耗收敛性能对比。从图中可以明显的看出, 基于 Q-learning 算法的无人机能耗的收敛速度和稳定性受到学习率选择的显著影响。具体而言, 在本文方案中, 较小的学习率会导致更好的收敛性能。当学习率 $\alpha = 0.1$ 时算法大约在训练到 130 次时开始收敛, 而当 $\alpha = 0.5$ 时算法大约需要训练 170 次才能实现收敛, 而当学习率更大时, 如 $\alpha = 0.7$ 时算法需要训练大约 260 次才能基本收敛。因此, 在本文提出的 Q-learning 方案中, 采取学习率为 0.1 是合适的。

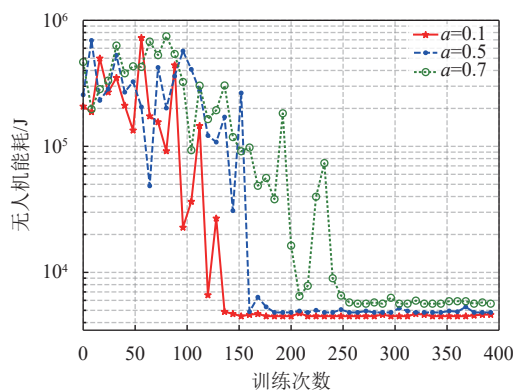


图 4 不同学习率下无人机能耗的收敛性能

4 结束语

本文研究 ISAC 辅助的单无人机系统的能耗优化问题。其中无人机充当双功能基站, 在与地面用户通信的同时感知空中目标的位置。通过联合优化无人机的飞行轨迹和发射功率来平衡感知和通信的性能需求, 在满足发射波束增益和通信速率约束的基础上, 最小化无人机的能耗。为求解这一复杂问题, 提出了基于强化学习 Q-learning 算法的联合无人机轨迹和发射功率的优化方法, 构建与无人机能耗相关的奖励函数, 使无人机能够自主性的改进策略。仿真结果表明, 本文方案与基准方案相比能有效的降低无人机能耗, 同时具有更优的收敛性。本文工作将推动未来 6G 通信感知一体化架构中多无人机通信感知一体化系统的能耗优化研究。

参考文献

- [1] NGUYEN D C, DING M, PATHIRANA P N, et al. 6G internet of things: A comprehensive survey[J]. *IEEE Internet of Things Journal*, 2021, 9(1): 359-383.
- [2] 尉志青, 冯志勇, 李怡恒, 等. 太赫兹通信感知一体化波形: 现状与展望[J]. *通信学报*, 2022, 43(1): 3-10.
WEI Z Q, FENG Z Y, LI Y H, et al. Terahertz joint communication and sensing waveform: Status and prospect[J]. *Journal on Communication*, 2022, 43(1): 3-10.
- [3] ZHANG J A, LIU F, MASOUIROS C, et al. An overview of signal processing techniques for joint communication and radar sensing[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2021, 15(6): 1295-1315.
- [4] 孙长印, 梁有为, 江帆, 等. 场景化毫米波特征选择和波束预测算法[J]. *电子科技大学学报*, 2023, 52(5): 689-698.
SUN C Y, LIANG Y W, JIANG F, et al. Scenario-based mmWave feature selection and beam prediction algorithm[J]. *Journal of University of Electronic Science and Technology of China*, 2023, 52(5): 689-698.
- [5] LIU F, MASOUIROS C, LI A, et al. MU-MIMO communications with MIMO radar: From co-existence to joint transmission[J]. *IEEE Transactions on Wireless Communications*, 2018, 17(4): 2755-2770.
- [6] ZHANG J A, RAHMAN M L, WU K, et al. Enabling joint communication and radar sensing in mobile networks—A survey[J]. *IEEE Communications Surveys & Tutorials*, 2021, 24(1): 306-345.
- [7] CHEN X, FENG Z, WEI Z, et al. Performance of joint sensing-communication cooperative sensing UAV network[J]. *IEEE Transactions on Vehicular Technology*, 2020, 69(12): 15545-15556.
- [8] MENG K, WU Q, MA S, et al. Throughput maximization for UAV-enabled integrated periodic sensing and communication[J]. *IEEE Transactions on Wireless Communications*, 2022, 22(1): 671-687.

(下转第 51 页)