

doi:10.3969/j.issn.1003-3114.2025.05.007

引用格式:朱佳琳,张鹏浩,李南希,等.基于强化学习的无人机辅助高效能数据收集方法[J].无线电通信技术,2025,51(5):940-950.[ZHU Jialin, ZHANG Penghao, LI Nanxi, et al. High-efficiency UAV-assisted Data Collection Method Leveraging Reinforcement Learning[J]. Radio Communications Technology, 2025, 51(5): 940-950.]

## 基于强化学习的无人机辅助高效能数据收集方法

朱佳琳<sup>1</sup>,张鹏浩<sup>2</sup>,李南希<sup>1</sup>,蒋 峥<sup>1</sup>,朱剑驰<sup>1</sup>

(1. 中国电信股份有限公司研究院,北京 102209;

2. 北京邮电大学 信息与通信工程学院,北京 100876)

**摘要:**物联网(Internet of Things, IoT)作为6G发展的核心领域之一,在驱动网络架构变革以及支撑核心应用场景中扮演着关键角色,然而,IoT系统存在能量不均衡以及网络生命周期短暂等问题,严重制约了数据收集效率的提升。随着无人机(Unmanned Aerial Vehicle, UAV)技术的兴起,其高度机动性可以有效构建视距(Line of Sight, LOS)通信链路,进而提升通信速率,这在IoT系统的数据收集方面具有很好的应用价值,可以解决IoT网络因生命周期短暂导致的数据收集效率低下的问题。为此,利用UAV对地面IoT设备进行数据收集,构建空地协同的数据采集传输链路,提出了一种基于深度强化学习(Deep Reinforcement Learning, DRL)的智能数据收集方法,设计了一种预测神经网络,通过在基站(Base Station, BS)侧预测网络数据进一步提高数据收集效率,从而实现降低IoT设备能耗、延长网络寿命的目的。仿真结果表明,所提数据收集算法在设备所需能耗、能量均衡性等方面具有较好的性能优势,优于常见的数据收集算法。同时,所提数据收集网络架构在预测数据占比12.5%时可以延长1.2倍的网络寿命。仿真证明了设计的预测神经网络在均方误差(Mean Squared Error, MSE)以及平均绝对误差(Mean Absolute Error, MAE)指标均优于其他对比网络。

**关键词:**6G;物联网;数据收集;无人机;强化学习

中图分类号:TP18;V279

文献标志码:A

开放科学(资源服务)标识码(OSID):

文章编号:1003-3114(2025)05-0940-11



## High-efficiency UAV-assisted Data Collection Method Leveraging Reinforcement Learning

ZHU Jialin<sup>1</sup>, ZHANG Penghao<sup>2</sup>, LI Nanxi<sup>1</sup>, JIANG Zheng<sup>1</sup>, ZHU Jianchi<sup>1</sup>

(1. China Telecom Research Institute, Beijing 102209, China;

2. School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China)

**Abstract:** The Internet of Things (IoT), as one core area of 6G development, plays a crucial role in driving network architecture changes and supporting core application scenarios. However, IoT systems suffer from energy imbalances and short network lifecycles, which severely restrict the improvement of data collection efficiency. With the rise of Unmanned Aerial Vehicle (UAV) technology, its high maneuverability can effectively construct Line of Sight (LOS) communication links, thereby improving communication speed. This has great application value in data collection of IoT systems and can solve the problem of low data collection efficiency caused by the short lifecycle of IoT networks. To this end, UAVs are used to collect data from ground IoT devices and build a data collection and transmission link for air-to-ground collaboration. An intelligent data collection method based on Deep Reinforcement Learning (DRL) is proposed. In addition, a predictive neural network is designed to further improve data collection efficiency by predicting network data at

收稿日期:2025-05-27

基金项目:国家重点研发计划(2024YFE0200102)

Foundation Item: National Key R&amp;D Program of China (2024YFE0200102)

the Base Station (BS) side, thereby achieving the goal of reducing IoT device energy consumption and extending network lifespan. Simulation results show that the proposed data collection algorithm has good performance advantages in terms of device energy consumption and energy balance, and is superior to traditional data collection algorithms. At the same time, the proposed data collection network architecture can extend the network lifespan by 1.2 times when the predicted data accounts for 12.5%. In addition, simulations have shown that the designed predictive neural network outperforms other compared networks in terms of Mean Squared Error (MSE) and Mean Absolute Error (MAE) metrics.

**Keywords:** 6G; IoT; data collection; UAV; reinforcement learning

## 0 引言

近年来, IoT 因其在智能医疗、智能家居、交通运输和智慧城市等领域的广泛应用而备受关注<sup>[1]</sup>。IoT 设备在环境监测、工业监控等方面发挥着数据采集和数据处理的基础作用,是 IoT 实现各种智能化应用的重要支撑。但由于 IoT 设备储能能力有限且许多设备通常需要在不充电的情况下持续采集和上传数据,当向 BS 传输数据时会产生大量的能量损耗,致使整个网络的寿命降低<sup>[2]</sup>。此外,由于 IoT 设备向 BS 传输数据时存在较少的直射路径,导致传输损耗较大进而使得传输失败概率增加<sup>[3]</sup>。

为解决上述问题,学术界已开展了一系列研究工作,文献[4-5]提出基于聚类的路由方法进行数据收集,通过有效的路由策略降低 IoT 设备向 BS 传输的能量损耗,从而提高 IoT 网络的寿命。此外,由于 UAV 具备高机动性和灵活部署等特性,通过与 IoT 设备建立 LOS 通信链路<sup>[6]</sup>,可有效降低 IoT 设备的能量损耗、提高通信效率,现已被广泛应用在数据收集场景中。为了实现 UAV 的高效能数据收集,文献[7]运用深度确定性策略梯度算法设计 UAV 飞行策略,提高了数据收集效率。文献[8]提出基于 Q 学习的强化学习算法来优化 UAV 的飞行轨迹,使得系统吞吐量最大化。

文献[9]创新设计多层聚类与 k-means 算法结合的框架,借助太阳能 UAV 实现了 IoT 设备数据采集量最大化。文献[10]围绕最小化 UAV 数据收集完成时间展开研究,优化数据采集流程。文献[11]利用迁移学习技术联合优化 UAV 的飞行速度和能量补给策略以提升数据采集效率。文献[12]则面向 UAV 群提出一种路径规划算法,实现了地面网络节点间的能量均衡。

尽管上述研究在提升数据收集效率方面取得了一定的进展,但大多数现有的研究工作主要集中在

IoT 设备端或数据传输阶段,相比之下,对于 BS 侧进行数据聚合和数据处理的关注相对较少。面向未来 6G AI 技术的发展需求,有望在 BS 侧对新业务数据进行智能化处理。为此,本文借助 UAV 的特性提出一种智能路径规划算法以实现 IoT 设备高效能的数据收集,此外,通过在 BS 侧搭载预测神经网络,对历史数据和实时信息的分析,可以预测未来数据的趋势,从而实现更高效的资源分配。借助本文构建的新型数据收集网络架构,UAV 辅助的数据收集流程与 BS 侧的智能处理环节实现了有机融合。与传统数据收集网络相比,它不仅能够大幅降低网络能耗,还能提高数据处理效率,从而构建更智能、更高效的 IoT 生态系统。

## 1 系统模型

### 1.1 数据收集网络模型

本文设计的数据收集网络模型如图 1 所示,包含数据收集和预测 2 个部分。利用 UAV 对地面 IoT 设备进行数据收集,通过在 BS 侧搭载预测神经网络模型基于 UAV 收集到的数据进行预测,从而快速得到 IoT 网络的全部数据集。假设 IoT 设备在三维空间服从均匀分布,且初始化 IoT 设备的剩余能量是相同的,当其能量耗尽时,则为无效 IoT 设备。当 BS 需要 UAV 进行数据收集时,会为各个 IoT 设备设定数据收集率  $\alpha$ , 其中  $\alpha \in [0, 1]$ 。 $\alpha = 0$  表示 IoT 设备不向 UAV 传输信息, $\alpha = 1$  表示 IoT 设备向 UAV 传输收集到的全部信息, $\alpha \in (0, 1)$  表示 IoT 设备向 UAV 传输收集到的比例为  $\alpha$  的部分信息。在数据收集时,假设 UAV 在飞行过程中并不开展数据收集工作,待 UAV 飞行一段时间  $\vartheta_i$  后,会进入静止状态开始时间为  $\varepsilon_i$  的数据收集,完成该阶段收集任务后,再次飞行一段时间,随后又静止下来继续进行数据收集,如此循环往复,直至所有 IoT 设备满足预设的  $\alpha$  标准,此时 UAV 飞往 BS,将所收集的数据进行卸载。

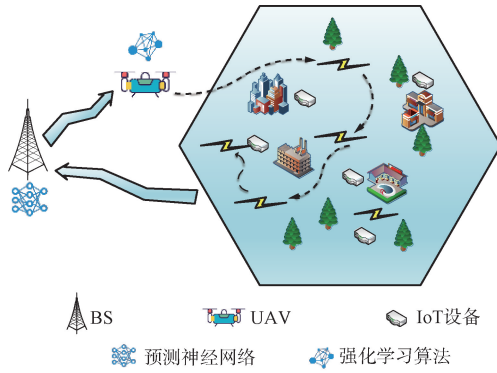


图 1 数据收集网络模型

Fig. 1 Data collection network model

### 1.2 信道模型

假设 UAV  $X$  与 IoT 设备  $S$  之间通信的平均数据传输速率用  $R(S, X)$  表示<sup>[13]</sup>:

$$R(S, X) = B \ln \left( 1 + \frac{P_T}{PL(S, X) N_0} \right), \quad (1)$$

式中:  $PL(S, X)$  为平均信道路径损耗,  $B$  为传输带宽,  $P_T$  为 IoT 设备的发射功率,  $N_0$  为噪声功率。考虑概率性空地路径损耗模型, UAV  $X$  与 IoT 设备  $S$  之间的平均路径损耗可表示为<sup>[14]</sup>:

$$PL(S, X) = p_{\text{LOS}}(S, X) PL_{\text{LOS}}(S, X) + [1 - p_{\text{LOS}}(S, X)] PL_{\text{NLOS}}(S, X), \quad (2)$$

式中:  $p_{\text{LOS}}(S, X)$  表示  $X$  与  $S$  之间 LOS 的概率。因其受环境因素以及仰角的影响,具体表达式为:

$$p_{\text{LOS}}(S, X) = \frac{1}{1 + \delta \exp(-\gamma[\theta - \delta])}, \quad (3)$$

式中:  $\theta$  为  $X$  相对于设备  $S$  的仰角,  $\delta$  和  $\gamma$  为环境参数<sup>[15]</sup>。式(2)中的  $PL_{\text{LOS}}(S, X)$  和  $PL_{\text{NLOS}}(S, X)$  分别为 LOS 和非视距 (Non Line of Sight, NLOS) 环境下的平均路径损耗,可表示为<sup>[13]</sup>:

$$\begin{cases} PL_{\text{LOS}}(S, X) = 10\vartheta \lg \left( \frac{4\pi f_c}{c} \|S - X\|_2 \right) + \mu_{\text{LOS}} \\ PL_{\text{NLOS}}(S, X) = 10\vartheta \lg \left( \frac{4\pi f_c}{c} \|S - X\|_2 \right) + \mu_{\text{NLOS}} \end{cases}, \quad (4)$$

式中:  $10\vartheta \lg \left( \frac{4\pi f_c}{c} \|S - X\|_2 \right)$  为自由空间路径损耗,  $\vartheta$  为路径损耗指数,  $f_c$  为载波频率,  $c$  为光速,  $\|\cdot\|$  为模 2 范数,  $\mu_{\text{LOS}}$ 、 $\mu_{\text{NLOS}}$  分别为 LOS 与 NLOS 环境下的路径损耗的均值。

### 1.3 数据收集策略衡量指标

衡量数据收集策略优劣的常用指标之一是网络生存时间<sup>[1,16]</sup>,其定义为网络中各个设备能量耗尽时的总运行时间。网络生存时间越长,说明数据收

集量越大,使用的数据收集策略越好<sup>[17-19]</sup>。与传统方法不同,本文中每一轮数据的收集率  $\alpha$  不一定为 1,因此在相同的轮次下,数据收集量将小于或等于传统方法,该数据收集策略下的网络生存时长也将大于传统方法。值得注意的是,除了数据收集,本文提出在 BS 侧进行数据预测,将数据收集侧采集到的数据量表示为  $C$ ,将 BS 侧预测得到的数据量表示为  $P$ ,将传统方法收集到的数据量表示为  $G$ ,且  $G=C+P$ 。为了将本文方法与传统方法进行比较,设计了代价函数,其表达式为:

$$C_f = \rho L_t - (1 - \rho) \left[ \sum_{i=1}^{G-C} (d_{p_i} - d_{r_i})^2 \right], \quad (5)$$

式中:  $\rho$  为取值为  $[0, 1]$  的权重因子,用以权衡网络生存时间和数据准确率之间的重要性关系;  $L_t$  表示网络的生存时间;  $d_{p_i}$  为预测网络预测出的第  $i$  个数据;  $d_{r_i}$  为 IoT 设备采集到的真实的第  $i$  个数据。  $C_f$  越大表明设计的数据收集策略越高效。

## 2 基于 SAC 的数据收集策略

为了在 UAV 完成数据收集过程中最小化 IoT 设备的总体能耗以最大化网络的生存时间,本文将设计相应的数据收集策略。考虑到 UAV 每次飞行的坐标位置、数据收集时的无线信道状况以及各个 IoT 设备的剩余电量等都会随着数据收集工作的推进而持续改变,这是一个不间断的动态变化过程。面对这类高维动作空间的处理问题, SAC (Soft Actor-Critic) 算法<sup>[20]</sup> 展现出更强的鲁棒性和更广泛的适应性。SAC 作为一种基于离线策略的最大熵强化学习算法,在决策过程中同步对策略即动作选择方式与价值函数展开学习。通过在奖励机制中融入策略的熵,促使算法对更多可行策略加以探索,从而提升选择的多样性与灵活性。相较于其他基于策略的 DRL 算法,如深度确定性策略梯度 (Deep Deterministic Policy Gradient, DDPG)<sup>[21]</sup> 以及 A3C (Asynchronous Advantage Actor-Critic)<sup>[22]</sup>, SAC 展现出更为卓越的效率。为此,本文提出一种基于 SAC 算法的强化学习方法,并将问题抽象为马尔可夫决策过程,具体给出了数据收集策略算法。

所提出的基于 SAC 的算法结构如图 2 所示,主要包含 2 个部分: Actor 和 Critic。Actor 包含 Policy Network,通过观察系统状态给出数据收集决策,同时更新策略。Critic 包含 Main Q Network 和 Target Q

Network,负责策略评估。此外,使用回放缓冲区来存储数据收集经验,用于训练 Actor 和 Critic 中的网络。

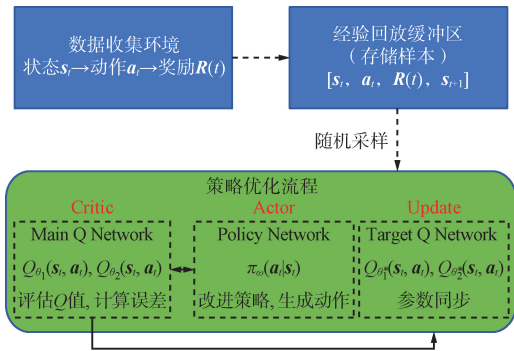


图2 基于SAC的数据收集算法结构

Fig.2 Structure of the data collection algorithm based on SAC

在 UAV 执行每个动作前, UAV 上的强化学习智能体收集得到以下环境参数:

$X_u(t)$  表示 UAV 在  $t$  时刻的三维坐标, 其  $x$ 、 $y$ 、 $z$  轴坐标分别表示为  $X_u^x(t)$ 、 $X_u^y(t)$ 、 $X_u^z(t)$ 。

$D(t)$  表示每个 IoT 设备在  $t$  时刻与 UAV 的距离, 第  $i$  个 IoT 设备与 UAV 的距离为  $D_i(t)$ 。

$O(t)$  表示在  $t$  时刻所有 IoT 设备所消耗的总能耗, 第  $i$  个 IoT 设备在  $t$  时刻的能耗为  $O_i(t)$ 。

$C(t)$  表示在  $t$  时刻所有 IoT 设备被 UAV 收集的数据量, 第  $i$  个 IoT 设备在  $t$  时刻被 UAV 收集的数据量为  $C_i(t)$ 。

$T$  表示整个数据收集过程的总时间, 假设 UAV 共做出  $q$  个动作最终完成数据收集, 则  $T = \sum_{i=1}^q t_i$ 。

令  $s$  表示状态空间, 在  $t$  时刻的状态向量表示为:

$$s_t = [X_u(t), D(t), O(t), C(t), T]。 \quad (6)$$

通过观察  $t$  时刻的状态, 智能体执行一个动作以指示 UAV 下一个时刻的位置坐标、飞行时长、飞行方向、飞行速度和需要进行数据收集的时长。将智能体在  $t$  时刻执行的动作向量表示为:

$$a_t = [\theta_t^{xy}, \theta_t^z, v_t^x, v_t^y, v_t^z, \varepsilon_t]， \quad (7)$$

式中:  $\theta_t^{xy}$  表示在  $t$  时刻 UAV 在  $xy$  平面的飞行方向夹角,  $\theta_t^z$  表示在  $t$  时刻 UAV 在  $z$  平面的飞行方向夹角,  $v_t^x$ 、 $v_t^y$ 、 $v_t^z$  分别表示  $t$  时刻 UAV 沿  $x$ 、 $y$ 、 $z$  轴上的飞行速度。

本文算法的伪代码如算法 1 所示。

算法 1 基于 SAC 的数据收集算法

输入: 用权重  $\theta_1$  和  $\theta_2$  初始化 Main Q Network  $Q_{\theta_1}(s_t, a_t)$  和  $Q_{\theta_2}(s_t, a_t)$ ; 用权重  $\theta_1^*$  和  $\theta_2^*$  初始化 Target Q Network  $Q_{\theta_1^*}(s_t, a_t)$  和  $Q_{\theta_2^*}(s_t, a_t)$ ; 用权重  $\omega$  初始化 Policy Network  $\pi_{\omega}(a_t | s_t)$ ; 初始化经验回放缓冲区  $\Gamma$   
输出:  $\theta_i$ 、 $\omega$ 、 $\theta_i^*$

1. for each epoch do
2. 收集初始观测状态  $s_0$ ,  $done = 0$
3. for  $done \neq 1$  do
4. 智能体接收来自 IoT 设备的数据采集请求, 并收集环境状态信息  $s_t$
5. 智能体根据状态信息和策略生成动作  $a_t$
6. 智能体引导 UAV 的飞行轨迹和数据采集时间并计算即时奖励  $R(t)$  并估计下一个状态  $s_{t+1}$
7. 将样本  $(s_t, a_t, R(t), s_{t+1})$  存储在  $\Gamma$  中
8. 通过计算式(9)中定义的  $J_Q(\theta)$  的梯度来更新  $\theta_i$ , 即  $\theta_i = \theta_i - \delta_Q \nabla \theta_i \frac{1}{B} \sum_{i=1}^B J_Q(\theta_i), i = 1, 2$
9. 通过计算式(10)中定义的  $J_{\pi}(\omega)$  的梯度来更新策略参数  $\omega$ , 即  $\omega = \omega - \delta_{\pi} \nabla \omega \frac{1}{B} \sum_{i=1}^B J_{\pi}(\omega)$
10. 通过  $\theta_i^* = \omega \theta_i + (1 - \omega) \theta_i^*, i = 1, 2$ , 更新 Target Q Network 的参数  $\theta_i^*$ .
11. 若数据收集任务完成, 则  $done = 1$ , 否则  $done = 0$
12. end for
13. end for

在飞行过程中, UAV 将目前的状态  $s_t$  输入其搭载的训练好的智能体, 智能体输出动作  $a_t$ , 得到一个即时奖励, 表示为:

$$R(t) = -O(t)。 \quad (8)$$

当 IoT 设备消耗的总能量越少, 其奖励值越大, 说明数据收集策略越好, 反之说明策略越差。

基于 SAC 的数据收集算法中有 3 个主要步骤, 具体如下。

步骤 1: 给定动作  $a_t$  后, 智能体指导 UAV 从  $t$  时刻开始飞行, 包括 UAV 与  $xy$ 、 $yz$ 、 $xz$  平面的飞行方向夹角、UAV 沿  $x$ 、 $y$ 、 $z$  轴上的飞行速度、时间。当飞行结束后智能体决策出数据收集时间  $\varepsilon_t$ , 同时对所有 IoT 设备进行  $\varepsilon_t$  时间的数据收集, 智能体根据本次的数据收集时间和数据收集量计算  $R(t)$ 。

步骤 2: 智能体得到下一个时刻的状态  $s_{t+1}$ , 将状态向量存储在经验回放  $\Gamma$ , 直到 UAV 对所有 IoT

设备完成数据收集且回到 BS,该轮次结束,结束标识符  $done = 1$ , 否则  $done = 0$ 。

步骤 3:智能体随机从缓冲区  $\Gamma$  中抽取一组数量为  $B$  的状态向量。通过计算式(9)和式(10)来更新网络参数<sup>[19]</sup>:

$$J_Q(\theta) = E_{(s_t, a_t) \sim \Gamma} \left[ \frac{1}{2} (Q_\theta(s_t, a_t) - Q_\theta^*(s_t, a_t))^2 \right], \quad (9)$$

$$J_\pi(\omega) = E_{s_t \sim \Gamma} [\alpha \ln(\pi_\omega(a_t | s_t)) - Q_\theta(s_t, a_t)], \quad (10)$$

式中:  $J_Q(\theta)$  为网络参数  $\theta$  的损失函数;  $E_{(s_t, a_t) \sim \Gamma}$  表示对从经验回放缓冲区  $\Gamma$  中采样得到的  $s_t$ 、 $a_t$  求期望;  $Q_\theta(s_t, a_t)$  为当前参数化的网络对状态  $s_t$  下采取动作  $a_t$  的估值;  $Q_\theta^*(s_t, a_t)$  为该状态下采取最优动作的值,通过求解最小化 MSE 使  $Q_\theta(s_t, a_t)$  不断逼近  $Q_\theta^*(s_t, a_t)$ ;  $J_\pi(\omega)$  为 Policy Network 的目标函数,通过计算  $J_\pi(\omega)$  的梯度,更新策略参数  $\omega$ ;  $E_{a_t \sim \pi_\omega}$  表示在给定策略  $\pi_\omega(a_t | s_t)$  下,对从该策略中采样得到的动作  $a_t$  求期望;  $\ln(\pi_\omega(a_t | s_t))$  用来衡量策略生成动作  $a_t$  的概率情况;  $Q_\theta(s_t, a_t)$  为通过网络参数对状态  $s_t$  下采取动作  $a_t$  的估值,通过最大化式(10)来更新 Policy Network 参数  $\omega$ ,使得高

价值动作  $a_t$  对应的策略概率增加。

### 3 预测神经网络

由于 UAV 采集到的数据是多维时间序列,即目标是在给定历史时间序列  $x_{1:T_p} \in I^{T_p \times D_{\text{dim}}}$  的情况下预测未来时间序列  $x_{T_p+1:T_p+\tau} \in I^{\tau \times D_{\text{dim}}}$ ,其中  $\tau$ 、 $T_p$  分别为未来和过去的时间步数,  $D_{\text{dim}}$  为维度数。为了提高预测准确性,本文将对具有跨维度依赖关系的 Crossformer<sup>[23]</sup> 进行改进,所设计的新网络称为 DepCrossformer,其主要分为 Encoder 和 Decoder 两大部分。其中,Encoder 通过对数据多尺度处理和依赖关系的捕捉,将原始的输入数据转化为不同层次、不同尺度的特征表示,这些特征表示包含了数据在不同粒度下的重要信息,为后续 Decoder 预测提供了丰富且有层次的信息基础。Decoder 基于 Encoder 输出的不同尺度的特征表示,在各个尺度上对数据进行预测,同时将在各个尺度上的预测结果相加并经过线性变换后得到最终的预测数据,这种方式能够综合考虑不同尺度下的信息,避免只关注单一尺度而忽略其他重要信息的问题。DepCrossformer 网络结构如图 3 所示。

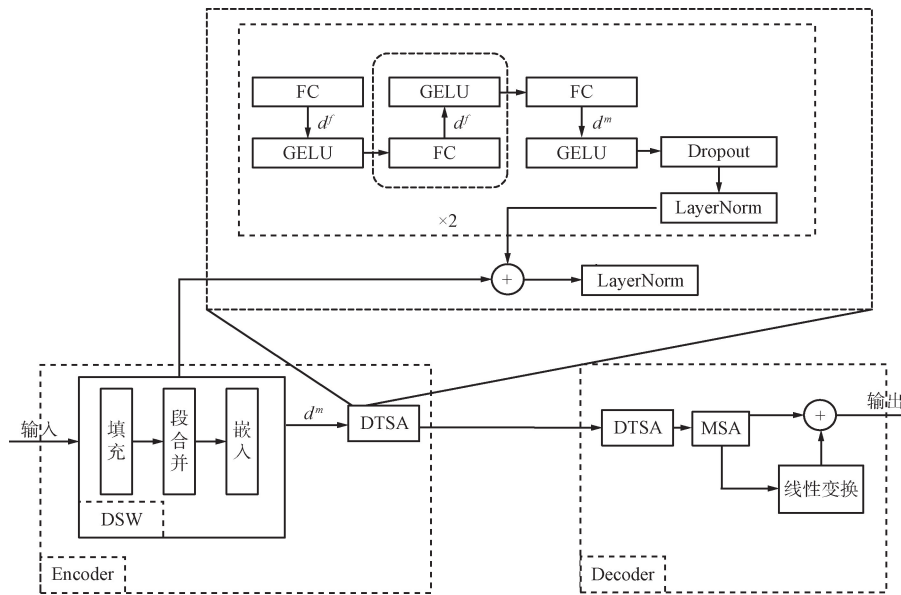


图 3 DepCrossformer 网络结构

Fig. 3 Network structure of DepCrossformer

考虑到 IoT 设备收集的数据具有跨维度的相关性,与维度分段式(Dimension-Segment-Wise, DSW)的思想相契合,因此复用 DSW 模块对不同尺度的信息进行嵌入处理。利用深度两阶段注意力层(Deep Two-Stage Attention, DTSA)分析不同时间点之间的关

联,捕捉时间序列数据中局部和全局的依赖关系。DTSA 是本文在 Crossformer 中两阶段注意力层(Two-Stage Attention, TSA)基础上进行改进的,通过在 TSA 模块之前引入一个多层的前馈神经网络(Multilayer Feedforward Neural Network, MFFN),能够将每个时间

步的嵌入表示映射到更高维的特征空间,通过复杂的非线性变换丰富了输入编码,从而提升了后续注意力机制的效果。具体地,MFFN 由全连接层和激活函数组成,其中激活函数采用 GELU 函数,其平滑地非线性转换能够更好地捕捉语义关系,帮助模型更好地处理语义信息。参数  $d^m$  和  $d^f$  分别代表 DSW 的输出维度和全连接(Fully Connected, FC)层的神经元数量。数据经过 MFFN 后,将其与原始数据相加后进行残差连接操作,通过残差连接,这些重要的原始特征不会在多层变换中丢失。此外,当 MFFN 学习到的变换接近恒等映射时,模型可以将信息从 Encoder 几乎无损地传递到 Decoder。在 Decoder 部分,基于 Encoder 输出的不同尺度的特征表示,在各个尺度上通过 DTSA 和交叉多头自注意力机制(Multi-Head Self-Attention, MSA)进行预测。

具体地,给定一个二维数组  $\mathbf{Z} \in \mathbb{R}^{L \times D \times d_{\text{mod}}}$  作为 DTSA 的层输入,其中  $L$  和  $D$  分别为片段和维度的数量, $\mathbf{Z}$  为 DSW 嵌入或更低层 DTSA 的输出。为了方便,用  $\mathbf{Z}_{:,d}$  表示在维度  $d$  上所有时间步数的向量。对每个维度应用 MSA,如式(11)所示:

$$\begin{cases} \mathbf{Z}^M = \mathbf{Z} + \text{MFFN}(\mathbf{Z}) \\ \bar{\mathbf{Z}}_{:,d}^M = \text{LayerNorm}(\mathbf{Z}_{:,d}^M + \text{MSA}^{\text{time}}(\mathbf{Z}_{:,d}^M, \mathbf{Z}_{:,d}^M, \mathbf{Z}_{:,d}^M)), \\ \mathbf{Z}^L = \text{LayerNorm}(\bar{\mathbf{Z}}^M + \text{MLP}(\bar{\mathbf{Z}}^M)) \end{cases} \quad (11)$$

式中:  $1 < d \leq D$ ,  $\text{LayerNorm}$  表示层归一化,  $\text{MLP}$  表示一个多层前馈网络,  $\text{MSA}^{\text{time}}(\mathbf{Q}, \mathbf{K}, \mathbf{V})$  表示多头自注意力层,所有维度共享相同的  $\text{MSA}^{\text{time}}$  层,  $\mathbf{Z}^M$  表示 MFFN 的输出,  $\bar{\mathbf{Z}}^M$  和  $\mathbf{Z}^L$  表示  $\text{MSA}^{\text{time}}$  和  $\text{MLP}$  的输出。跨时间阶段的计算复杂度为  $O(DL^2)$ ,在此阶段后,同一维度中时间片段之间的依赖关系在  $\mathbf{Z}^L$  中被捕获,并将  $\mathbf{Z}^L$  作为跨维度阶段的输入,以捕获跨维度依赖关系。

## 4 实验结果与分析

### 4.1 实验设置

本文实验中选用公开数据集 ETTh1<sup>[24]</sup> 作为 IoT 设备需采集的数据,其中包含了电力相关的 7 组不同时间序列数据集,共  $7 \times 14\ 400$  条数据。在 BS 侧,利用此数据集训练本文所提出的预测模型 DepCrossformer,以预测未来一段时间内的电力数据,按照 3 : 1 : 1 将其划分为训练集、测试集和验证集。

本文提出的 SAC 算法和 DepCrossformer 模型实验设置参数如表 1 所示。对于 DepCrossformer 模型

总训练轮数为 20,假设验证损失在 3 个轮次内没有下降,训练过程将提前终止。对于对比的基线模型,均使用与本文 DepCrossformer 模型相同的实验设置。本文对于 DepCrossformer 和基线模型等均使用 PyTorch 实现,并在配备 48 GB 内存的单张 NVIDIA A6000 GPU 上进行训练。

表 1 SAC 算法和 DepCrossformer 相关参数

Tab. 1 Key parameters of SAC algorithm and DepCrossformer

参数	参数值
SAC 算法优化器	Adam
SAC 算法学习率	$1 \times 10^{-5}$
SAC 算法策略学习率	$3 \times 10^{-7}$
SAC 算法的 discount 参数 $\xi$	0.99
SAC 算法的缓冲区 $\Gamma$	$3 \times 10^5$
SAC 算法中网络的隐藏层数目	2
SAC 算法中每层隐藏状态输入维度	128
SAC 算法中每层隐藏状态输出维度	64
SAC 算法中批量大小	256
SAC 算法的激活函数	ReLU
SAC 算法中目标平滑系数 $\eta$	0.005
DepCrossformer 优化器	Adam
DepCrossformer 损失函数	MSE
DepCrossformer 批量大小	32
DepCrossformer 初始学习率	$1 \times 10^{-4}$
DepCrossformer 中 DSW 的输出维度 $d^m$	256
DepCrossformer 中 FC 层的神经元数量	512
DepCrossformer 中多头注意力头数量	4

### 4.2 对比方法

本文基于 SAC 算法对 UAV 的数据收集路径进行规划,通过最大化策略的熵来提升策略探索性,从而实现更高效的轨迹决策。为了评估本文算法的性能,将其与 DDPG 算法进行对比,该算法使用确定性策略梯度结合 Actor-Critic 架构进行连续动作空间的路径优化。此外,还引入了 2 种非学习型的启发式路径规划算法:① 中心点(Central Point, CP)路径算法,其将目标设为所有未完成数据收集的 IoT 设备的质心坐标,模拟集中式路径决策;② 随机(Random, RD)路径算法,模拟 UAV 以随机速度和方向飞行,代表无规则的探索式策略。所有算法的仿真环境设置一致,相关环境参数如表 2 所示。

表 2 环境参数

Tab. 2 Environmental parameters

参数	参数值
IoT 设备数目	35
IoT 设备分布	$x, y$ 轴上服从 $[0, 5000]$ 的均匀分布, $z$ 轴上服从 $[0, 200]$ 的均匀分布
路径损耗指数	2
IoT 发射功率/mW	2
UAV 天线增益/dB	1
每个 IoT 设备分配到的带宽/Hz	$1 \times 10^6 / 35$
噪声功率谱密度/(dBm/Hz)	$1 \times 10^{-20.4}$
自由空间损耗/dB	2
由环境决定的常数项系数 $\delta, \gamma$	10, 0.03
路径损耗指数	3
载波频率 $f_c$ /Hz	$2 \times 10^9$
LOS 环境下路径损耗的均值/dB	1
NLOS 环境下路径损耗的均值/dB	20
数据收集最小时间/s	5
IoT 设备的能量/J	30 000

对于 DDPG 算法, Actor 网络的学习率设置为  $1 \times 10^{-4}$ , Critic 网络的学习率设置为  $1 \times 10^{-4}$ , discount 参数  $\xi$  设置为 0.99, Actor、Critic 网络的隐藏层数目为 2, 每层隐藏状态输入维度设置为 256, 输出维度设置为 128, 批量大小设置为 256, 激活函数为 ReLU 函数, 目标平滑系数  $\eta$  设置为 0.005。

对于 CP 算法, 在仿真环境相同的情况下, UAV 每次动作的飞行目的地为所有未完成数据收集的 IoT 设备的坐标均值点, 飞行速度为 5 m/s, 飞行方向为当前 UAV 坐标相对于目的地的方向, 飞行时间为 60 s, 数据收集时间为 5 s。

对于 RD 算法, 在仿真环境相同的情况下, UAV 每次的飞行速度、方向、时间以及数据收集时间均为随机数。

### 4.3 实验结果

为了将本文所提数据收集网络架构与传统非预测数据网络的效率进行比较, 利用式(5)给出的代价函数进行评判, 根据算法 1 对强化学习模型展开训练可得到 UAV 的路径策略  $P_p$ , 如图 4 所示, 图中“SAC+DepCrossformer- $x$ ”表示 UAV 侧依据  $P_p$  策略执行数据收集任务, 同时 BS 侧运用 DepCrossformer 模型进行数据预测, 假设预测数据比例占总数据量的  $x$ , 本文模型的输入长度统一设为 168, 当模型预测输出长度为 24、48、72、96 时,  $x$  对应取值分别为 0.14、0.29、0.43、0.57。曲线“SAC+traditional”表示

UAV 侧采用  $P_p$  策略完成 100% 的数据收集工作, BS 侧无需进行数据预测。可以看出, 当  $\rho$  的取值超过一定阈值时, 本文所提数据收集网络架构下的效率指标值明显高于传统的数据收集网络。当  $\rho=1$  时, 即忽略预测网络引入的数据误差的情况下, 预测数据占比 12.5% 时本文所提数据收集网络架构带来的效率指标是传统方法的 1.18 倍; 当预测数据占比 57% 时, 效率可提升至原来的 2.38 倍。可见在数据收集过程中, 引入预测神经网络对于提高网络生命周期非常重要。

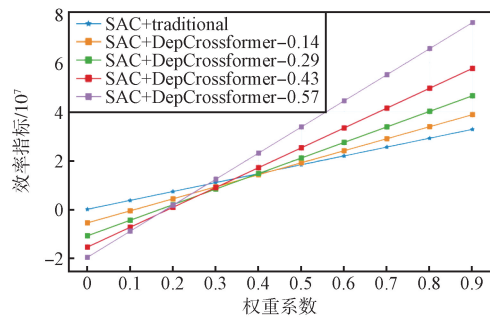


图 4 代价函数

Fig. 4 Cost function graph

图 5 为本文所提的 SAC 算法和其对比的 DDPG 算法的学习训练效果图, 可以看出 2 种算法随着训练回合的增加, 奖励值逐渐趋近于收敛。本文分别取 2 种强化学习算法训练效果最好时对应的数据收集策略进行仿真分析。

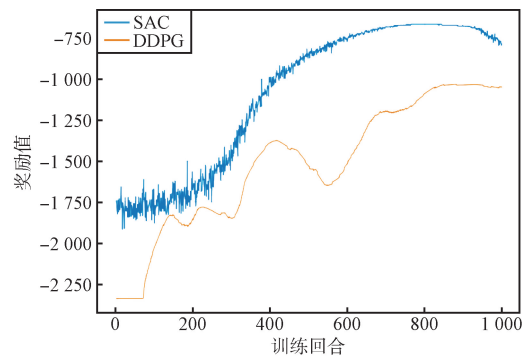


图 5 训练效果对比

Fig. 5 Training performance comparison graph

在给定的数据收集策略下, 每轮数据收集任务第  $i$  个 IoT 设备消耗的能量表示为  $C_{p_i}$ , 单个 IoT 设备的总能量表示为  $s_p$ , 整个数据收集网络的生存轮次为:

$$S_i = \min(s_p / C_{p_i}), i \in [1, n], \quad (12)$$

式中:  $n$  表示 IoT 设备的数量。

为了验证不同数据收集策略对网络生存状态的影响,数据收集网络的生存轮次如图 6 所示。其中图例中带有“trad”后缀的曲线代表不与预测神经网络相结合的传统数据收集方法,反之则为预测数据占比 12.5%时的仿真结果。

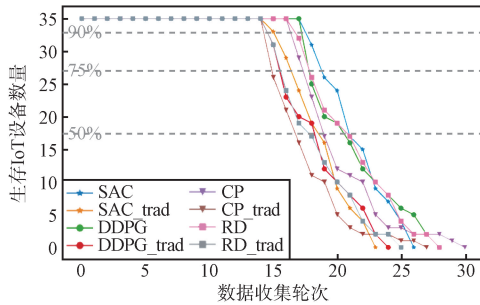


图 6 数据收集网络的生存轮次  
Fig. 6 Graph of network survival rounds

由图 6 可以看出,无论是否结合预测神经网络架构进行数据收集,本文提出的基于 SAC 的数据收集策略均具有更良好的网络生存状态。例如,在 SAC 与预测神经网络相结合的数据收集方法框架下,若要求所有 IoT 设备都存活,则基于 SAC 的策略能进行 18 轮完整的数据收集任务,而 CP 算法只能进行 17 轮,RD 算法只能进行 16 轮。此外,随着数据收集轮次的增加,生存 IoT 设备的数量会逐渐减少,但同一数据收集轮次下,基于 SAC 策略下的生存 IoT 数量明显高于其他算法。例如,在完成第 19 轮数据收集任务时,基于 SAC 的策略还有 89% 的设备存活,而 DDPG 算法下还有 71% 的设备存活,CP 算法下还有 66% 的设备存活,RD 算法下仅有 49% 的设备存活。当数据收集轮次增大到一定阈值后,考虑到有些算法设计时设备消耗的能量不均匀,生存的 IoT 数量反而更多,在这种情况下由于设备存活过少影响数据完整性,其生存状态性能优劣可以忽略不计。对比“SAC\_trad”曲线,在没有预测神经网络结合的数据收集策略下,IoT 设备仅能进行 15 轮次的数据收集,在等参数条件下本文所提数据收集架构可以带来 1.2 倍的网络寿命增益。由此可见,本文所提的基于 SAC 的数据收集策略具有更好的网络生存性能。

为了更直观清晰地说明各个数据收集算法的能耗性能,本文对每轮数据收集过程中 IoT 设备的能耗进行了分析,结果如表 3 和图 7 所示。表 3 给出了不同算法下数据收集过程中每轮 IoT 设备所消耗的总能量、所有 IoT 设备消耗能量的均值、标准差以

及最大最小差值。数据表明本文提出的基于 SAC 的算法总能耗最小,是 DDPG 算法能耗的 97.9%,是 RD 算法能耗的 89.4%,是 CP 算法能耗的 93.4%。此外,基于 SAC 算法的 IoT 设备能耗均值、标准差、最大最小差值均为所有方法中最小,说明本文所提算法在每轮数据收集任务中能量消耗最均衡,有利于保障网络的长期运行以及数据收集的完整性。

表 3 每轮数据收集 IoT 能量消耗  
Tab. 3 IoT energy consumption for each round of data collection

算法	总能耗	均值	标准差	最大最小差值
SAC	79 590.0	2 274.00	270.67	900.0
DDPG	81 286.8	2 322.48	327.54	995.8
RD	88 980.0	2 542.29	379.20	1 520.0
CP	85 240.0	2 435.43	326.97	1 270.0

单位:J

图 7 的箱型图呈现了在运用不同算法开展数据收集工作时各个 IoT 设备能耗的分布情况。可以看出,基于 SAC 的数据收集算法能够使各个 IoT 设备的能耗更加均衡,可以降低因 IoT 设备能耗不均衡带来的网络效率低下以及数据收集不完整等问题。

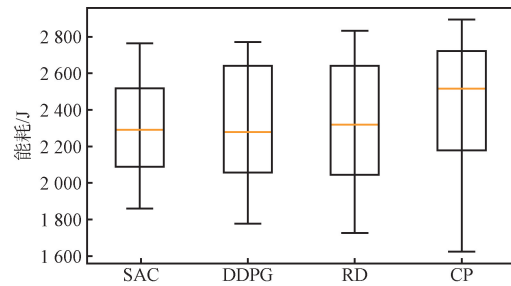


图 7 数据收集能量消耗分布  
Fig. 7 Distribution diagram of energy consumption for data collection

图 8 和图 9 展示了本文提出的 DepCrossformer 预测神经网络与文献[23]中 7 种其他网络的性能对比图,为了便于仿真对比,设置输入数据长度为 168,输出(预测)数据长度为 24 以及 48。

图 8 是在 ETTh1 数据集上测试各个网络输入数据长度为 168,输出(预测)数据长度为 24 时的性能,可以看出本文所给出的 DepCrossformer 网络在 MSE 和 MAE 指标上均优于其他网络。MSE 方面,

DepCrossformer 网络可以达到 LSTMa 网络的 46.62%,最差也能达到 Crossformer 网络的 99.34%。

MAE 方面,DepCrossformer 网络是 LSTMa 网络的 58.17%,是 Crossformer 网络的 98.91%。

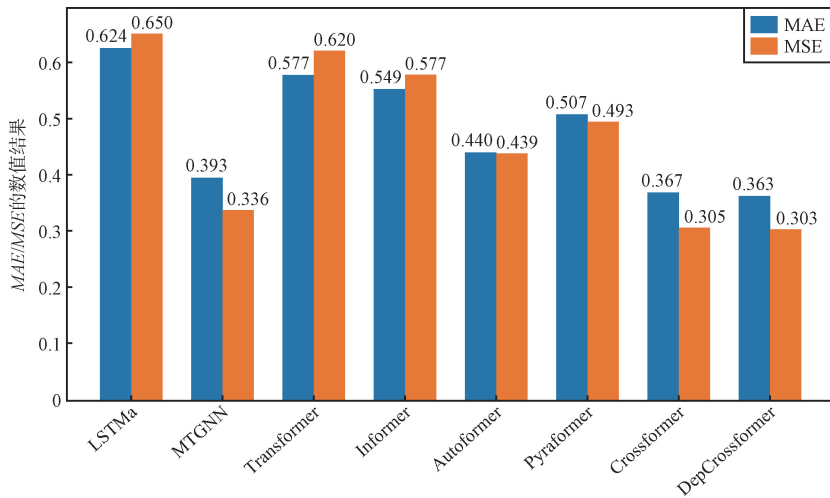


图 8 输入数据长度为 168,输出(预测)数据长度为 24 时的预测神经网络性能对比

Fig. 8 Performance comparison chart of prediction neural networks when the input data length is 168 and the output (predicted) data length is 24

图 9 是在 ETTh1 数据集上测试各个网络输入数据长度为 168,输出(预测)数据长度为 48 时的性能。MSE 方面,DepCrossformer 网络是 LSTMa 网络的 47.92%,是 Crossformer 网络的 98.01%。MAE 方面,DepCrossformer 网络是 LSTMa 网络的 57.33%,

是 Crossformer 网络的 98.22%。同样可以看出本文所给出的 DepCrossformer 网络在 MSE 和 MAE 指标上均优于其他方法。综上所述,可以看到在输出的预测数据长度为 24 和 48 时,本文所提预测神经网络在 MSE 和 MAE 指标上都优于其他对比网络。

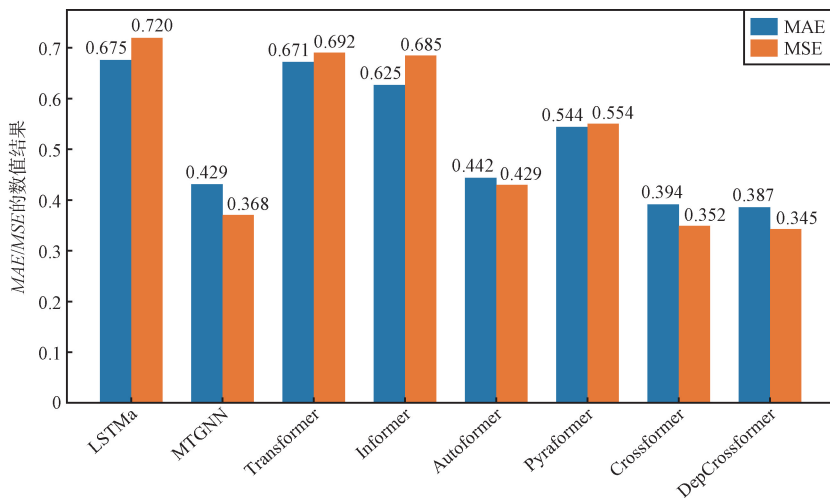


图 9 输入数据长度为 168,输出(预测)数据长度为 48 时的预测神经网络性能对比

Fig. 9 Performance comparison chart of prediction neural networks when the input data length is 168 and the output (predicted) data length is 48

## 5 结束语

本文通过构建 UAV 空地协同数据采集传输链

路,提出了一种基于 DRL 的智能数据收集方法,设计了一种预测神经网络,通过优化数据收集网络架构可以提升数据收集效率。仿真结果显示,该

算法在设备能耗、能量均衡性方面优于 DDPG、CP、RD 等对比算法。通过仿真统计在不同数据收集轮次下 IoT 设备的生存数量可以得到在预测数据占比 12.5% 时所设计的网络架构可以延长 1.2 倍的网络寿命。此外,本文设计的预测神经网络在 MSE、MAE 指标上优于其他 7 种对比网络。本文的研究成果为 IoT 数据收集提供了更高效的技术方案,同时可为 6G 时代网络架构优化提供。

#### 参考文献

- [1] KAUR G, CHANAK P, BHATTACHARYA M. Energy-efficient Intelligent Routing Scheme for IoT-enabled WSNs [J]. IEEE Internet of Things Journal, 2021, 8(14): 11440-11449.
- [2] WANG Z J, LIU R K, LIU Q R, et al. Energy-efficient Data Collection and Device Positioning in UAV-assisted IoT[J]. IEEE Internet of Things Journal, 2019, 7(2): 1122-1139.
- [3] ZHU B T, BEDEER E, NGUYEN H H, et al. UAV Trajectory Planning in Wireless Sensor Networks for Energy Consumption Minimization by Deep Reinforcement Learning[J]. IEEE Transactions on Vehicular Technology, 2021, 70(9): 9540-9554.
- [4] SHARMA N, AGARWAL U, SHAURYA S, et al. Energy-efficient and QoS-aware Data Routing in Node Fault Prediction Based IoT Networks [J]. IEEE Transactions on Network and Service Management, 2023, 20(4): 4585-4599.
- [5] PANDEY O J, YUVARAJ T, PAUL J K, et al. Improving Energy Efficiency and QoS of LPWANs for IoT Using Q-learning Based Data Routing [J]. IEEE Transactions on Cognitive Communications and Networking, 2021, 8(1): 365-379.
- [6] KHAWAJA W, GUVENC I, MATOLAK D W, et al. A Survey of Air-to-Ground Propagation Channel Modeling for Unmanned Aerial Vehicles [J]. IEEE Communications Surveys & Tutorials, 2019, 21(3): 2361-2391.
- [7] TALUKDAR N, RAGHAV A, HAZRA A, et al. A Deep Deterministic Policy Gradient Method for Optimizing Task Completion Time and Energy Efficiency in UAV-assisted IoT Networks [J]. IEEE Internet of Things Journal, 2025, 12(15): 31907-31917.
- [8] 周雯,花敏,许艺瀚,等. 能量受限无人机数据采集系统的飞行路线优化[J/OL]. 无线电通信技术(2025-05-13) [2025-05-23]. <https://link.cnki.net/urlid/13.1099.tn.20250527.1613.002>.
- [9] ISSA R, EZZEDDINE H, ABBAS N. Energy-efficient Solar-powered UAV-assisted Reliable Data Collection Considering IoT Elevation [C] // 2024 Conference on Renewable Energy Technologies and Modern Communications Systems: Future and Challenges. Shaqra: IEEE, 2024: 1-6.
- [10] LI M, LIU X, WANG H. Completion Time Minimization Considering GNs' Energy for UAV-assisted Data Collection [J]. IEEE Wireless Communications Letters, 2023, 12(12): 2128-2132.
- [11] CHU N H, HOANG D T, NGUYEN D N, et al. Joint Speed Control and Energy Replenishment Optimization for UAV-assisted IoT Data Collection with Deep Reinforcement Transfer Learning [J]. IEEE Internet of Things Journal, 2023, 10(7): 5778-5793.
- [12] WAN L T, ZHANG K, SUN L, et al. Energy-AOI-aware UAV-assisted Data Collection: A Multi-agent Deep Reinforcement Learning-based Trajectory Optimization [C] // 2023 IEEE 23rd International Conference on Communication Technology (ICCT). Wuxi: IEEE, 2023: 176-180.
- [13] GHORBEL M B, RODRIGUEZ-DUARTE D, GHAZZAI H, et al. Joint Position and Travel Path Optimization for Energy Efficient Wireless Data Gathering Using Unmanned Aerial Vehicles [J]. IEEE Transactions on Vehicular Technology, 2019, 68(3): 2165-2175.
- [14] MOZAFFARI M, SAAD W, BENNIS M, et al. Drone Small Cells in the Clouds: Design, Deployment and Performance Analysis [C] // 2015 IEEE Global Communications Conference (GLOBECOM). San Diego: IEEE, 2015: 1-6.
- [15] AL-HOURANI A, KANDEEPAN S, LARDNER S. Optimal LAP Altitude for Maximum Coverage [J]. IEEE Wireless Communications Letters, 2014, 3(6): 569-572.
- [16] BEHERA T M, MOHAPATRA S K, SAMAL U C, et al. Residual Energy-based Cluster-head Selection in WSNs for IoT Application [J]. IEEE Internet of Things Journal, 2019, 6(3): 5132-5139.
- [17] KIM T, QIAO D J. Energy-efficient Data Collection for IoT Networks via Cooperative Multi-hop UAV Networks [J]. IEEE Transactions on Vehicular Technology, 2020, 69(11): 13796-13811.
- [18] HU H M, XIONG K, QU G, et al. AoI-minimal Trajectory Planning and Data Collection in UAV-assisted Wireless Powered IoT Networks [J]. IEEE Internet of Things Journal, 2020, 8(2): 1211-1223.
- [19] LIU J, TONG P, WANG X J, et al. UAV-aided Data Collection for Information Freshness in Wireless Sensor Networks [J]. IEEE Transactions on Wireless Communications, 2020, 20(4): 2368-2382.

- [20] HAARNOJA T, ZHOU A, HARTIKAINEN K, et al. Soft Actor-critic Algorithms and Applications [ EB/OL ]. (2018-12-13) [ 2025-05-23 ]. <https://arxiv.org/abs/1812.05905>.
- [21] LILICRAP T P, HUNT J J, PRITZEL A, et al. Continuous Control with Deep Reinforcement Learning [ EB/OL ]. (2015-09-09) [ 2025-05-23 ]. <https://arxiv.org/abs/1509.02971>.
- [22] MNIEH V, BADIA A P, MIRZA M, et al. Asynchronous Methods for Deep Reinforcement Learning [ C ] // The 33rd International Conference on Machine Learning. New York: ICML, 2016: 1928-1937.
- [23] ZHANG Y H, YAN J C. Crossformer: Transformer Utilizing Cross-Dimension Dependency for Multivariate Time Series Forecasting [ C ] // The Eleventh International Conference on Learning Representations. Kigali: ICLR, 2023: 1-21.
- [24] ZHOU T, MA Z Q, WEN Q S, et al. Fedformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting [ C ] // International Conference on Machine Learning. Baltimore: PMLR, 2022: 27268-27286.

---

作者简介:

朱佳琳 女, (1996—), 硕士, 工程师。主要研究方向: 无人机通信、物联网技术、无线空口信令标准化等。

张鹏浩 男, (2000—), 硕士。主要研究方向: 语义通信、人工智能技术等。

李南希 男, (1990—), 博士, 高级工程师。主要研究方向: 大规模天线系统、5G 物理层技术、智能表面技术等。

蒋 峥 男, (1972—), 博士, 教授级高级工程师。主要研究方向: 通感一体化、无线空口信令和无线网络架构标准化等。

朱剑驰 男, (1981—), 硕士, 教授级高级工程师。主要研究方向: 无线通信技术研究 and 标准化、5G 标准化、6G 物理层技术等。