

DOI:10.19651/j.cnki.emt.2620949

基于门控 Transformer 的边缘计算卸载与 无线充电器部署联合优化算法^{*}

贾本源¹ 刘祖深¹ 铁 奎²

(1. 中北大学仪器与电子学院 太原 030051; 2. 中电科思仪科技(安徽)有限公司 蚌埠 233010)

摘 要: 无线供能移动边缘计算网络能够为物联网应用提供充足的算力与稳定的能量供应。然而,与云端相比,边缘侧任务到达和无线信道状态具有更强的时变性,固定部署方案难以及时适应任务负载的变化。为此,本文考虑由多个无线充电器组成的无线供能移动边缘计算网络,构建了一个以最大化无线设备计算完成率为目标的优化问题,联合优化无线充电器在线部署决策、无线设备任务卸载比率和资源分配。针对该混合整数非凸优化问题,本文将其分解为两个子问题,并提出一种基于门控 Transformer 的联合优化算法,以有效应对长期网络动态性和高维动作空间带来的挑战。本文中的无线充电器部署采用候选位置集合上的在线离散决策建模,通过在不同时间帧更新部署方案来适应网络状态变化。仿真结果表明,与已有算法相比,所提算法在常规负载下计算完成率平均提升约 48%,在高负载计算密集型场景下性能提升可达 60%,且具有良好的稳定性和收敛性。

关键词: 移动边缘计算;计算卸载;无线供能传输;Transformer

中图分类号: TN929 **文献标识码:** A **国家标准学科分类代码:** 510.5015

Gated transformer-based joint optimization algorithm for task offloading and wireless charger deployment for mobile edge computing

Jia Benyuan¹ Liu Zushen¹ Tie Kui²

(1. School of Instrument and Electronics, North University of China, Taiyuan 030051, China;

2. Ceyear Technologies (Anhui) Co., Ltd., Bengbu 233010, China)

Abstract: Wireless powered mobile edge computing networks provide a promising solution for supplying both computational capability and stable energy to Internet of Things applications. However, compared with cloud environments, task arrivals and wireless channel conditions at the network edge exhibit stronger temporal dynamics, making fixed deployment schemes insufficient for adapting to workload variations. To address this issue, this paper considers a multi-charger Wireless powered mobile edge computing network and formulates an optimization problem that maximizes the computation completion rate of wireless devices by jointly optimizing wireless charger online deployment decisions, task offloading ratios, and resource allocation. To solve this mixed-integer non-convex problem, the original problem is decomposed into a wireless charger deployment decision subproblem and a resource allocation subproblem. Then, a Gated Transformer-based joint optimization algorithm is proposed to effectively handle long-term network dynamics and high-dimensional action spaces. In this work, wireless charger deployment is modeled as an online discrete decision over a predefined set of candidate locations, where the deployment scheme is updated frame by frame to adapt to network dynamics. Simulation results show that, compared with baseline algorithms, the proposed method improves the average computation completion rate by about 48% under normal workloads and by up to 60% in high-load computation-intensive scenarios, while maintaining good stability and convergence.

Keywords: mobile edge computing; computational offloading; wireless power transfer; Transformer

0 引 言

随着物联网应用向智能化发展,无线设备面临算力不

足与能量受限的问题^[1]。结合移动边缘计算(mobile edge computing, MEC)与无线供能传输技术(wireless power transfer, WPT),无线供能移动边缘计算(WP-MEC)网络

收稿日期:2026-03-30

^{*} 基金项目:安徽省重点研究与开发计划(202304a05020011)项目资助

成为解决这一问题的关键方案^[2]。一个 WP-MEC 网络由一个或多个无线充电器、一个 MEC 服务器和多个具有能量采集功能的无线设备组成。无线充电器部署一个射频能量发射器,为无线设备提供 WPT 服务。MEC 服务器为无线设备提供计算服务。无线设备可对任务进行本地计算和边缘卸载任务到 MEC 服务器。无线充电器广播射频信号向无线设备供能,无线设备收集能量用于本地计算和边缘卸载。目前的各类研究主要集中在固定无线充电器位置的场景下的资源分配或卸载决策。文献[3]针对在信道衰落下传统方法难以实时求解 WP-MEC 网络中的卸载决策问题,提出了一种基于深度强化学习(deep reinforcement learning, DRL)的在线卸载算法,实现了毫秒级做出次优的卸载决策。针对具有多个无线充电器与多个无线设备的 WP-MEC 网络,文献[4]设计了基于扰动 Lyapunov 优化理论的算法,联合优化通信资源和计算资源以最大化 WP-MEC 网络的计算数据量。文献[5]研究了具有多个无线充电器的 WP-MEC 网络中的协作计算问题,提出了一种双层优化方法,联合配对决策与多维资源分配以降低 WP-MEC 网络能耗。文献[6]研究多个无线充电器 WP-MEC 网络的长期能量供应最小化问题,提出了一个两阶段的基于多智能体 DRL 的分布式计算卸载算法,其中高层智能体优化发射功率和 WPT 时长,低层智能体基于本地观测优化卸载决策、延迟和 CPU 频率。

研究[7-10]调查了无线充电器在无线供能通信网络(wireless-powered communication network, WPCN)网络中的部署。文献[7]提出基于 Grundy 的着色算法确定最小无线充电器数量,并提出黑洞优化算法求解无线充电器的最优位置,提高无线充电器的覆盖效率,即以最少的无线充电器去最大化供能覆盖区域。文献[8]考虑多天线无线充电器的部署,提出基于区域离散化和功率分配的子模集优化算法,提高了所有无线设备的充电效率。文献[9]研究配备有向天线的无线充电器部署,提出了一种基于区域离散化与覆盖支配集抽取的贪心算法,提升了 WPCN 网络的整体充电效率。文献[10]研究在恶劣场景中无线充电器的部署,提出了一种考虑容错的算法,即使有少量无线充电器故障也能最大化任务的充电效率。虽然上述工作涉及了无线充电器的部署优化,但大多针对静态网络,无法应对边缘侧任务的快速动态变化。原有固定位置的无线充电器很难维持较高的任务计算完成率。边缘计算还需要考虑动态的任务生成和任务卸载决策,对能量供应和任务计算进行联合优化。然而,由于信道的时变性及多无线设备之间的资源竞争,WP-MEC 网络中的能量传输与任务卸载存在复杂的耦合关系,难以在有限时间内求得无线充电器部署的最优解,并且 DRL 算法采用的深度神经网络(deep neural network, DNN)结构在捕捉多无线设备信道间的复杂依赖关系时存在局限。

针对上述挑战,本文提出一种基于 Gated Transformer-XL

的 DRL(GTXL-DRL)联合优化算法。本文将时变 WP-MEC 环境下的无线充电器部署决策建模为马尔可夫决策过程,在每个时间帧根据当前任务到达和信道状态,对候选位置集合上的部署方案进行在线更新。Gated Transformer-XL(GTXL)通过多头自注意力机制^[11],能够全局建模多无线设备状态之间的复杂依赖关系,并有效捕捉时间帧间的时序相关性,并行分析所有无线设备的信道增益和任务量变化,从而弥补传统 DNN 在长序列处理中局部感知能力有限的问题。与标准 Transformer-XL 主要依赖注意力机制和记忆单元建模长程时序依赖不同,本文进一步结合 WP-MEC 场景中任务负载具有多连续时间帧相关性、无线信道存在瞬时波动的特点,对网络结构进行了针对性改进:一方面,GTXL 在多头自注意力模块输出和前馈网络输出后分别引入门控单元,对不同于子层输出进行筛选,抑制由短时信道起伏引入的无效扰动;另一方面,在策略学习过程中,增加记忆单元来保留多个连续时间帧的历史状态信息,使模型在当前决策能够同时利用当前观测与近期历史信息,优化下一步输出的部署决策。记忆单元长度的设置既要能够覆盖任务负载变化的短中期相关性,又要避免过长历史带来的噪声累积与计算开销上升。基于上述改进,GTXL 能够更适应 WP-MEC 网络中任务动态性与信道随机性并存的决策场景,在提升部署决策时序建模能力的同时,增强策略学习的稳定性与鲁棒性。

本文的主要贡献如下:

1)构建了多无线充电器 WP-MEC 网络模型,建立了以最大化任务计算完成率为目标的混合整数非凸优化问题,联合优化无线充电器在线部署决策、无线设备任务卸载比率和资源分配变量。其中,无线充电器部署被建模为候选位置集合上的离散选址决策。

2)为求解上述问题,将原问题分解为无线充电器部署决策子问题与凸资源分配子问题。针对部署决策子问题,提出 GTXL-DRL 算法,通过多头自注意力机制从历史状态中提取任务到达的时间规律与空间分布特征,以提高时变环境下的部署决策质量;同时引入门控机制,以减弱无线信道波动对策略学习的影响。对于资源分配子问题,则采用拉格朗日对偶法高效求解。

3)仿真结果表明,GTXL-DRL 算法在常规负载下计算完成率平均提升约 48%。在中高负载和无线设备性能较低的场景下,其优势更加明显,最高提升可接近 60%。GTXL 中的门控机制使模型收敛后的波动率降低约 15%。仿真实验验证了 GTXL-DRL 在提升无线设备计算完成率方面的优越性,以及门控机制降低噪声干扰有效性。因此,GTXL-DRL 在复杂、动态的 WP-MEC 网络中更适合在线决策。

1 系统模型

1.1 网络模型

如图 1 所示,WP-MEC 网络包括 N 个配备单天线的无

线设备、 M 个配备单天线的无线充电器和一个 MEC 服务器。无线设备集合记为 $\mathcal{N} = \{1, 2, \dots, N\}$ ，无线充电器集合记为 $\mathcal{M} = \{1, 2, \dots, M\}$ 。 N 个无线设备固定分布在一个二维区域 Ω 中， M 个无线充电器可以部署在 Ω 中的候选位置，记为集合 $\mathcal{C} = \{C_1, C_2, \dots, C_M\}$ 。 Ω 中央部署了一个拥有高计算能力的 MEC 服务器，可以接收并计算无线设备卸载的任务^[12]。无线充电器广播射频信号向无线设备供能，每个无线设备均装载一个可以收集能量的可充电电池，为本地计算和边缘卸载提供能量。相邻的无线充电器可以覆盖共同的无线设备，这表示某个无线设备可以被多个无线充电器影响。在通信过程中，由于 WPT 与任务卸载沿着同一个信道传输，为避免相互干扰，每个无线设备均采用时分复用(time division duplexing, TDD)电路。定义 $h_{n,m}$ 表示第 n 个无线设备与第 m 个无线充电器之间的信道增益，定义 h_n 表示第 n 个无线设备与 MEC 服务器之间的信道增益。

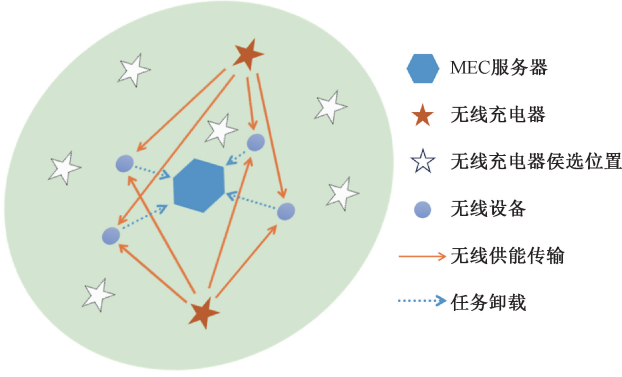


图 1 具有多个无线充电器的 WP-MEC 网络模型

Fig. 1 WP-MEC network model with multiple wireless chargers

WP-MEC 网络时间被划分为多个长度为 T 的连续时间帧。在每一时间帧开始时，无线充电器向设备供能，WPT 所需的时间记为 aT ，其中 $a \in [0, 1]$ 。第 n 个无线设备收集的能量为：

$$E_n = \sum_{m=1}^M \mu P h_{n,m} a T \quad (1)$$

其中， $\mu \in (0, 1)$ 表示能量收集效率， P 表示无线充电器的发射功率。在每一个时间帧内的 WPT 结束之后，每个无线设备需要利用收集到的能量完成一个计算任务。为了高效地处理可分割的计算任务，本文考虑部分卸载策略^[13]，该策略将一个完整的计算任务分割成两个部分，一部分在无线设备本地计算，另一部分卸载到 MEC 服务器完成计算任务。本文中，无线充电器的部署优化体现在候选位置集合上的离散选址决策，并随时间帧进行在线更新。系统根据任务负载和信道状态的变化，对各时间帧的部署方案进行自适应调整，以提升网络整体计算完成率。

1.2 计算模型

1) 本地计算模式

在 WPT 之后，无线设备对任务进行本地计算。第 n

个无线设备的本地计算任务数据量 $D_{L,n}$ 满足：

$$D_{L,n} = \frac{f_n t_n}{\varphi} \quad (2)$$

其中， f_n 表示第 n 个无线设备的 CPU 频率， $t_n \in [0, (1-a)T]$ 表示第 n 个无线设备的计算时间， φ 表示处理一个比特数据量所需的 CPU 频率。本地计算的能耗为：

$$E_{L,n} = k_n (f_n)^3 t_n \quad (3)$$

其中， k_n 表示计算能量效率系数^[14]。

2) 边缘计算模式

在 WPT 之后，无线设备卸载任务给 MEC 服务器，MEC 服务器计算任务并回传计算结果给无线设备。设 $\tau_n T$ 为第 n 个无线设备的卸载时间，其中 $\tau_n \in [0, 1-a]$ 。MEC 服务器计算速率远大于无线设备，且回传给无线设备的计算反馈数据量远小于卸载到 MEC 服务器的数据量，因此 MEC 服务器进行任务计算和回传计算结果的时间可忽略不计^[15]。每个时间帧内，WPT 和任务卸载的时间满足：

$$\sum_{n=1}^N \tau_n + a = 1 \quad (4)$$

第 n 个无线设备的卸载任务数据量 $D_{O,n}$ 满足：

$$D_{O,n} = B \tau_n T \log_2 \left(1 + \frac{P_n h_n}{\omega} \right) \quad (5)$$

其中， B 表示无线设备任务卸载的通信带宽， P_n 表示第 n 个无线设备的传输功率， ω 表示噪声功率。任务卸载的能耗为：

$$E_{O,n} = P_n \tau_n T \quad (6)$$

1.3 问题模型

在具有多个无线充电器的 WP-MEC 网络中， N 个无线设备的任务计算完成率 Q 定义为：

$$Q(\mathbf{c}, \mathbf{x}, \boldsymbol{\tau}, a) = \frac{\sum_{n=1}^N (\mathbf{x} D_{L,n} + (1-\mathbf{x}) D_{O,n})}{\sum_{n=1}^N D_n} \quad (7)$$

其中， $\mathbf{c} = \{c_m | m \in \mathcal{M}\}$ 表示无线充电器在候选位置集合上的部署决策， $\mathbf{x} = \{x_n | n \in \mathcal{N}\}$ 表示任务卸载比率， $\boldsymbol{\tau} = \{\tau_n | n \in \mathcal{N}\}$ 表示卸载时间分配向量， $D_n = D_{L,n} + D_{O,n}$ 表示第 n 个无线设备需要计算的任务数据量。本文研究最大化 WP-MEC 网络的任务计算完成率，联合优化了无线充电器部署决策、无线设备任务卸载比率和资源分配。该优化问题可以定义为：

$$\begin{aligned} & \text{P1: } \max_{\mathbf{c}, \mathbf{x}, \boldsymbol{\tau}, a} Q \\ & \text{s. t. } \begin{cases} \text{C1: } \sum_{n=1}^N \tau_n + a = 1 \\ \text{C2: } a \geq 0 \\ \text{C3: } \tau_n \geq 0 \\ \text{C4: } x_n \in [0, 1] \\ \text{C5: } E_{L,n} \leq E_n \\ \text{C6: } E_{O,n} \leq E_n \end{cases} \end{aligned} \quad (8)$$

由于式(8)中的目标函数是非凸的且需要优化离散变量 $\mathbf{c}$, 问题 P1 是一个难以求解的混合整数非凸问题。然而, 一旦给定 $\mathbf{c}$, P1 被转化为以下凸问题:

$$\begin{aligned} \text{P2: } & \max_{\mathbf{x}, \boldsymbol{\tau}, \mathbf{a}} Q \\ \text{s. t. } & \text{C1} \sim \text{C6} \end{aligned} \quad (9)$$

因此, 本文将问题 P1 转化为两个子问题, 即无线充电器部署决策子问题和资源分配子问题 P2。由于资源分配子问题 P2 是凸问题, 采用拉格朗日对偶方法可求出最优解 $\{\mathbf{x}^*, \boldsymbol{\tau}^*, \mathbf{a}^*\}$ 。

解决问题 P1 的主要挑战在于无线充电器部署决策子问题。传统优化算法通常需要在候选位置集合上反复迭代搜索部署方案, 计算复杂度较高, 在信道增益和任务数据量随机变化的 WP-MEC 网络中难以满足在线决策需求。为此, 本文提出 GTXL-DRL 算法, 通过与环境持续交互, 自适应学习时变网络状态下的近似最优部署策略, 从而避免传统方法在每次状态变化后重新求解所带来的高计算开销。进一步地, 引入 GTXL 增强 DRL 智能体对时序依赖关系的建模能力, 使其能够从历史信道增益和任务量变化中提取有效信息, 从而实现适用于动态 WP-MEC 网络的在线部署决策。

1.4 无线充电器在线部署物理实现的补充分析

本文中的无线充电器在线部署主要表示部署状态的在线重构, 其物理实现并不是指充电器本体在每一时间帧都发生大范围位移, 而是对应于实际系统中的两类场景。一类是无人机搭载式无线充电器, 该场景下无线充电器安装于低空无人机平台, 候选位置集合可理解为无人机在服务区域内预先规划的悬停点或航路点, 系统根据任务负载与信道状态变化选择下一时段的悬停位置。第二类是固定站点上的可开关天线阵列式无线充电器。该场景下无线充电器物理位置保持不变, 通过切换阵列单元、波束方向或射频前端端口, 使能量覆盖重心在不同候选区域之间切换。因此, 本文的候选位置集合既可对应实际空间中的候选部署点, 也可等效表示为不同波束覆盖状态的在线切换, 具有一定的工程实现基础。

在实际系统中, 在线部署还伴随一定的部署代价。对于无人机搭载方式, 代价主要体现为位置迁移带来的飞行能耗与调整时延; 对于可开关天线阵列方式, 代价主要体现为阵列重构、波束切换及附加电路控制开销。本文重点研究动态环境下部署决策与任务卸载的联合优化, 为保证模型简洁性, 未将上述代价纳入考虑, 而是将其作为在线部署可实现性的补充分析。

2 GTXL-DRL 算法

在 WP-MEC 网络中, 环境状态具有时间动态性与空间耦合性。任务到达量、无线设备剩余可用能量以及信道增益会随时间帧不断变化, 呈现出时序相关特征; 同一时间帧内不同无线设备之间、不同信道之间以及候选部署位置之

间又存在复杂的耦合关系。对于这种高维状态空间, 若仅采用传统深度神经网络或仅依据当前时刻观测进行决策, 难以同时描述多个连续时间帧的动态信息与多状态分量之间的全局关联, 从而导致策略学习不稳定、收敛速度缓慢。为此, 本文提出了基于门控 Transformer 的边缘计算卸载与无线充电器部署联合优化算法 GTXL-DRL。算法选择 GTXL 模块而不是 LSTM 或 GRU, 主要基于问题 P1 结构本身。LSTM 和 GRU 主要通过递归方式对历史信息进行压缩, 适合处理以时间递推关系为主的单一时间序列建模问题, 并不适合解决 P1 这类高维度的决策问题。GTXL 中的多头自注意力机制能够并行处理高维状态空间, 动态捕捉不同无线设备信道条件之间的复杂关联, 因此相较于 LSTM 和 GRU, 能利用全局特征的长期依赖关系以优化部署决策。同时, 为了智能体有效学习长期依赖关系, GTXL 采用门控机制与记忆单元进行优化, 可以提升学习效率与策略的收敛速度, 实现 WP-MEC 网络整体任务计算完成率的最大化。

2.1 MDP 建模

将系统模型建模成马尔可夫决策过程 (markov decision process, MDP) 模型。问题 P1 可以通过一个包含 3 个元素的元组 (S, A, R) 来定义, S 表示状态空间, A 表示动作空间, R 表示奖励函数。

1) 状态空间设计

考虑到信道增益和任务数据量随机变化, 本文利用 GTXL 提取多个时间帧的状态信息, 以提升 DRL 在第 t 个时间帧的决策性能。在第 t 个时间帧开始时观察到的系统状态是 $S_t = \mathbf{h}_t \otimes \mathbf{D}_t$, $t \in [1, \dots, T]$, 其中 $\mathbf{h}_t = \{h_{t,n,m}, h_{t,n} | n \in \mathcal{N}, m \in \mathcal{M}\}$ 表示在第 t 个时间帧的无线设备到无线充电器之间和无线设备到 MEC 服务器之间的信道增益, $\mathbf{D}_t = \{D_{t,n} | n \in \mathcal{N}\}$ 表示无线设备的计算任务数据量, $\otimes$ 表示连接状态向量。

2) 动作空间设计

定义 $A_t = \mathbf{c}_t$, 其中 $\mathbf{c}_t = \{c_{t,m} | m \in \mathcal{M}\}$ 表示第 t 个时间帧内无线充电器在候选位置集合上的部署选择。该动作描述的是面向时变网络状态的在线部署决策。

3) 奖励函数设计

在每个时间帧, DRL 智能体会基于当前状态和执行动作作出响应以得到一个奖励。为了实现 WP-MEC 网络计算完成率最大化, 奖励函数应与 GOT 式(8)中的目标函数正相关。在第 t 个时间帧的奖励函数为

$$R_t = \frac{\sum_{n=1}^N (x D_{L,n} + (1-x) D_{O,n})}{\sum_{n=1}^N D_n} \quad (10)$$

2.2 GTXL-DRL 算法总览

用于求解 P1 的 GTXL-DRL 算法流程图如图 2 所示。在第 t 个时间帧, GTXL 以信道增益 $\mathbf{h}_t = \{h_{t,n,m}, h_{t,n} | n \in \mathcal{N}$,

$m \in \mathcal{M}$ 和任务数据量 $\mathbf{D}_t = \{D_{t,n} | n \in \mathcal{N}\}$ 作为输入,并根据部署策略 π_{θ_t} 输出一组松弛的部署动作 $\hat{\mathbf{c}}_t$,然后将松弛动作 $\hat{\mathbf{c}}_t$ 转换为 $K \times N$ 个候选的部署动作。其中 K 是一个决定每个松弛动作候选集数量的参数。对于每组候选的

部署动作,基于 P2 的最优资源分配 $\{\mathbf{x}_t^*, \boldsymbol{\tau}_t^*, \mathbf{a}_t^*\}$ 计算出一个最优动作 $\mathbf{c}_t^*$,并获得相应的,且均满足式(9)中的约束条件。基于 P1 优化的部署动作 $\mathbf{c}_t^*$,得到奖励 R_t ,并将获得的 $(\mathbf{h}_t, \mathbf{D}_t, \mathbf{c}_t^*)$ 添加到记忆池中用于更新 GTXL 模型。

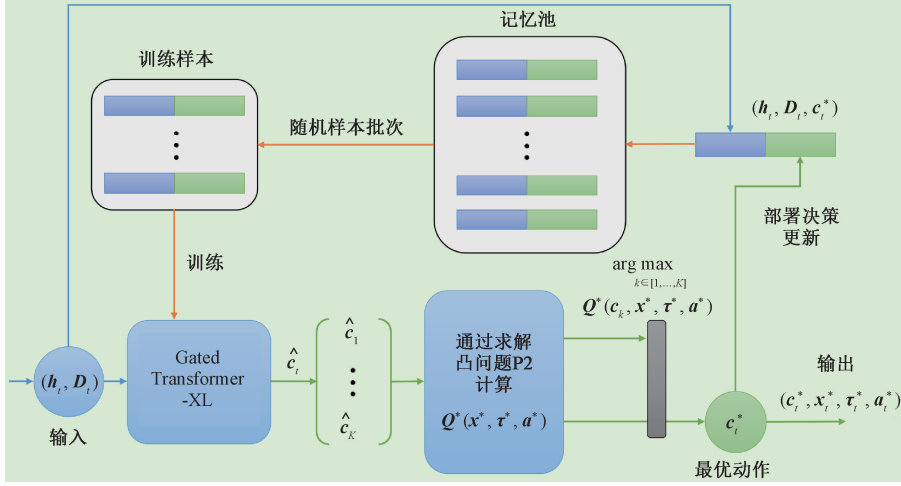


图 2 GTXL-DRL 算法流程图

Fig. 2 Flowchart of the GTXL-DRL algorithm

1) 模型更新

在第 t 个时间帧,一个新的经验 $(\mathbf{h}_t, \mathbf{D}_t, \mathbf{c}_t^*)$ 被添加到记忆池中。每隔 δ 个时间帧,从记忆池中随机选择一批训练样本来训练 GTXL。通过 Adam 算法来更新 GTXL 模型,损失函数为:

$$L(\theta) = \frac{1}{|\Gamma|} \sum_{\tau \in \Gamma} (\pi_{\theta}(\mathbf{h}_{\tau}, \mathbf{D}_{\tau}) - \mathbf{c}_{\tau})^2 \quad (11)$$

其中, Γ 表示所选训练样本的集合, $|\Gamma|$ 表示所选训练样本的集合的数量。

2) 动作生成

GTXL 模型的输入为 $\mathbf{h}_t$, 输出为一个松弛的部署动作 $\hat{\mathbf{c}}_t$, 由一个松弛方法表示:

$$\hat{\mathbf{c}}_t = f_{\theta_t}(\mathbf{h}_t) \quad (12)$$

其中, $\hat{\mathbf{c}}_t = \{\hat{c}_{t,i} | i = 1, \dots, N\}$, $\hat{c}_{t,i}$ 表示 $\hat{\mathbf{c}}_t$ 的第 i 个候选部署动作。本文在隐藏层中使用 ReLU 激活函数,在输出层使用 sigmoid 激活函数,使松弛的部署动作满足 $\hat{c}_{t,i} \in (0, 1)$ 。然后,将 $\hat{c}_{t,i}$ 转换为 K 个部署动作。较大的 K 会输出更多数量的解,但模型的计算复杂度也会升高。本文采用分支动作生成规则以转换 $\hat{\mathbf{c}}_t$, 当 K 为奇数时:

$$\hat{\mathbf{c}}_t = \left\{ 2^{\frac{K-1}{2}} \hat{\mathbf{c}}_t \mid X \in \left\{ -\frac{K-1}{2}, -\frac{K-1}{2} + 1, \dots, \frac{K-1}{2} \right\} \right\} \quad (13)$$

当 K 为偶数时

$$\hat{\mathbf{c}}_t = \left\{ 2^{\frac{K}{2}} \hat{\mathbf{c}}_t \mid Y \in \left\{ -\frac{K}{2}, -\frac{K}{2} + 1, \dots, \frac{K}{2} \right\} \right\} \quad (14)$$

例如,候选动作的数量 $K = 3$ 表示将一个松弛的 $\hat{\mathbf{c}}_t$ 转

换为 3 个部署动作 $0.5\hat{\mathbf{c}}_t$, $\hat{\mathbf{c}}_t$ 和 $2\hat{\mathbf{c}}_t$ 。使得模型可以高效地获取 K 倍的训练样本以提升训练效果。基于式(13)和(14),可获得第一个时间帧的卸载部署决策 $\hat{\mathbf{c}}_{1,i}$ 。对候选位置集正则化处理,可将 $\hat{\mathbf{c}}_{1,i}$ 对应到某个候选位置上。通过迭代执行模型更新 $(\mathbf{h}_t, \mathbf{D}_t)$ 和动作生成,改进 GTXL 的部署策略 π_{θ_t} , 以根据获得一个近似最优的无线充电器部署位置 $\mathbf{c}_t^*$ 。GTXL-DRL 算法的伪代码如算法 1 所示。

算法 1: GTXL-DRL 算法

- 1: 初始化状态 S_t , 记忆池 O , GTXL 中的参数 θ_t ;
- 2: 设置迭代次数 E 和训练时间间隔 δ ;
- 3: **for** $e = 1, 2, \dots, E$ **do**
- 4: 生成松弛的部署动作 $\hat{\mathbf{c}}_t$;
- 5: 根据规则式(13)和(14),将当前动作 $\hat{\mathbf{c}}_t$ 转换为 K 个扩展动作;
- 6: 将所有新生成的 K 个动作输入凸问题 P2 中计算剩余决策 $\{\mathbf{x}_t^*, \boldsymbol{\tau}_t^*, \mathbf{a}_t^*\}$;
- 7: 执行具有最大奖励 R_t^* 的动作 $\mathbf{c}_t^*$;
- 8: 将状态转移 $(S_t, \mathbf{c}_t^*)$ 存放到记忆池 O 中;
- 9: **If** $t \bmod \delta = 0$ **then**
- 10: 从记忆池 $\{(\mathbf{h}_{\tau}, \mathbf{D}_{\tau}, \mathbf{c}_{\tau}^*) \mid \tau \in \Gamma_t\}$ 均匀采样一批数据集;
- 11: 使用采样的数据集训练 GTXL, 并通过 Adam 算法来更新 θ_t ;
- 12: **end**
- 13: **end**

根据算法 1, GTXL-DRL 的计算开销主要包含两个阶段:一是基于 GTXL 的状态特征提取与动作生成;二是对候选部署动作对应的资源分配子问题 P2 进行求解。在状态特征提取阶段, GTXL 的自注意力机制的时间复杂度为 $O(L^2d)$, 其中 L 为序列长度, d 为隐藏层维度。在动作生成阶段, GTXL 首先输出一个松弛动作向量, 再构造 K 个离散候选部署动作, 复杂度为 $O(K)$ 。对于每一个候选部署动作, 都要求解一次凸资源分配子问题 P2。用拉格朗日对偶法求解 P2 的复杂度为 $O(N)$ 。一共需要对 K 个候选动作分别求解并比较奖励值, 该部分总复杂度为 $O(KN)$ 。其中 L 为有限时间窗口长度, d 为固定参数, K 通常取固定较小的常数, 所以在线执行阶段的主要计算开销集中在求解 P2。因此 GTXL-DRL 在线执行阶段的总体时间复杂度为 $O(N)$ 级别, 显著优于传统强化学习在庞大动作空间中的穷举搜索 $O(2^N)$, 满足边缘计算的毫秒级时延要求。

2.3 GTXL

在 WP-MEC 网络中, 环境 $(\mathbf{h}_t, \mathbf{D}_t)$ 的动态性和策略 π_{θ_t} 的高维度会导致模型难以学习和收敛速度缓慢, 可以使用 GTXL 来缓解这种现象^[16]。GTXL 既可以基于时隙变化捕获动态环境下的依赖关系, 也擅长处理高维度的决策问题, 适合用于 WP-MEC 网络的复杂决策。因此, 本文提出了 GTXL-DRL 算法。采用 GTXL 来从每个时间帧的信道增益和任务数据量中提取特征, 使用门控机制来帮助无线充电器查询历史记忆并只关注特定记忆。GTXL 的多头注意力模块将注意力机制应用于输入的不同线性映射, 并将其拼接在一起, 使模型能够关注全局, 从而有助于捕捉输入中不同特征之间的关系。多层感知器模块对输入序列中每个位置的特征执行独立的非线性变换^[17], 增强模型捕捉输入序列中不同位置的复杂模式和关系的能力。门控层用于在特征经过不同模块时对其进行加权。该模型可以通过学习适当的权重来控制输入信息流, 使其更适合特定任务和数据分布。

在门控机制设计上, 本文分别在多头注意力输出和前馈网络输出后引入门控, 以增强模型对有效信息的筛选能力。对于任意输入特征 u , 门控模块定义为:

$$\text{Gate}(u) = u \odot \sigma(W_g u + b_g) \quad (15)$$

其中, $\sigma(\cdot)$ 为 Sigmoid 激活函数, W_g 和 b_g 分别表示线性变换的权重矩阵和偏置项, $\odot$ 表示逐元素乘法。该门控方式通过对特征的不同维度赋予自适应权重, 实现对无效扰动信息的抑制。模型更新过程可表示为:

$$\begin{cases} x' = \text{LN}_1(x) \\ h^{\text{att}} = \text{MHA}(x', x', x') \\ x = x + \text{Gate}_1(h^{\text{att}}) \\ x'' = \text{LN}_2(x) \\ h^{\text{mlp}} = \text{MLP}(x'') \\ x = x + \text{Gate}_2(h^{\text{mlp}}) \end{cases} \quad (16)$$

其中, $\text{LN}(\cdot)$ 、 $\text{MHA}(\cdot)$ 和 $\text{MLP}(\cdot)$ 分别表示层归一化、多头自注意力和前馈神经网络, 可以看出, 门控机制嵌入在子层输出与残差连接之前, 分别对注意力输出和前馈输出进行自适应调节, 从而减弱短时波动和噪声对策略学习的影响。经过 GTXL 编码后, 对所有 token 特征进行层归一化和平均池化, 得到全局表示向量, 再通过输出头映射为候选部署位置上的动作权重。输出头由两层全连接网络构成, 其输出经过 Sigmoid 函数压缩到 $[0, 1]$ 区间, 用于生成后续部署决策。

3 仿真结果

本节在具有多个无线充电器的 WP-MEC 网络中, 评估 GTXL-DRL 算法的性能。采用的无线充电器为 Powercast 公司生产的 TX91501 型射频能量发射器, 输出功率 $P = 3 \text{ W}$ 。采用的无线设备为 Powercast 公司生产的 P2110 Powerharvester。WP-MEC 网络是半径 50 m 的圆形区域。其中 MEC 服务器在中心, 无线充电器服从平均分布有 25 个候选位置集, 无线设备随机分布。平均信道增益 $h_{n,m}$ 遵循自由空间路径损耗模型^[9]:

$$h_{n,m} = A_d \left(\frac{3 \times 10^8}{4\pi f_c} \right)^2 d_{n,m}^{-d_e} \quad (17)$$

其中, $A_d = 4.11$ 表示天线增益, $f_c = 915 \text{ MHz}$ 表示载波频率, $d_{n,m}$ 表示第 n 个无线设备到第 m 个无线充电器之间的距离, $d_e = 2.8$ 表示路径损耗指数。对于无线设备, 设置能量收集效率 $\mu = 0.51$, 计算能量效率系数 $k_n = 10^{-26}$, 任务卸载的通信带宽 $B = 2 \text{ MHz}$, 噪声功率 $\omega = 10^{-13} \text{ W}$ 。

为了更准确地反映 WP-MEC 网络在实际复杂环境中的通信状况, 本文在自由空间路径损耗模型的基础上, 引入了 Rayleigh 衰落信道模型。考虑到无线设备与无线充电器、MEC 服务器之间存在多径效应和非视距传输, 第 i 个无线设备与目标节点之间的实际信道增益 h_i 可以表示为: $h_i = h_{\text{loss}} \cdot |\alpha|^2$ 。其中, h_{loss} 表示大尺度衰落, $\alpha \in N(0, 1)$ 表示服从小尺度 Rayleigh 衰落的复高斯随机变量。计算任务数据量 D_n 服从正态分布, 均值 $\eta = 1.0 \text{ Mbit}$, 标准差 $\sigma = 0.8 \text{ Mbit}$ 。本文采用 3 层 TransformerBlock 堆叠而成的 GTXL 编码器, 每层包含 4 头多头自注意力模块、前馈神经网络模块以及门控单元。模型特征维度设置为 128, 隐藏层维度为 512。为验证所提算法的性能, 本文选取的所有对比算法均采用相同的状态空间、动作空间和奖励函数, 并在相同训练轮数和仿真环境下进行训练与测试。最终实验结果均在多个随机种子下重复运行取平均值。

图 3 显示了 GTXL-DRL 算法的训练收敛性能。图 3(a) 通过比较 5 种不同学习率下的正则化奖励, 发现 GTXL-DRL 算法在过小的学习率下收敛缓慢, 而在过大的学习率下收敛性能较差。在图 3(b) 中, 发现 Transformer 层数在

等于 3 时正则化奖励达到峰值。当层数增加到 4 层时,性能反而下降。这是因为在 WP-MEC 场景下,状态空间的马尔可夫性较强,过深的网络会引入了过拟合,导致智能

体对历史噪声过于敏感,而忽略了当前帧的主导信息。因此,在实验中 Transformer 层数设为 3 层是捕捉长程依赖与保持模型泛化能力之间的最佳平衡点。

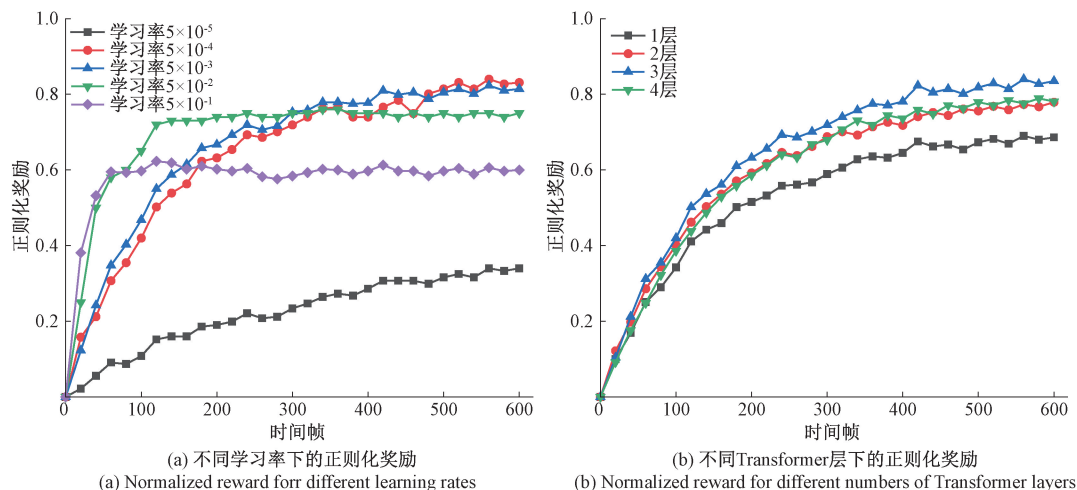


图 3 不同算法参数下 GTXL-DRL 的收敛性能

Fig. 3 Convergence performance of the GTXL-DRL algorithm under different algorithm parameter settings

为了更好地评估 WP-MEC 网络环境下 GTXL-DRL 框架的性能,本节将 GTXL-DRL 算法与 5 个典型强化学习算法进行比较,如图 4 所示。Reinforce 算法是一种基本的策略梯度算法,需要在一个完整的时间序列结束后,根据累积回报来更新策略。它实现简单、对模型假设要求较少,在无线任务到达较为平稳、网络规模较小时具有一定优势。Policy Gradient 算法是一类直接对策略进行优化的 DRL 方法,通过不断调整策略参数,使期望回报最大化。该方法结构相对简单,易于实现,在 WP-MEC 环境中可以较快得到一个可行的卸载与部署策略。深度 Q 网络(deep Q-network, DQN)算法利用深度神经网络来近似 Q 值函数,适合处理高维状态空间中的决策问题^[18]。在 WP-MEC 环境中,DQN 能够在一定程度上平衡能量供给和任务卸载决策,并通过经验回放等机制提高样本利用率。近端策略优化(proximal policy optimization, PPO)算法是一种典型的在线策略优化方法,其核心思想是在策略更新过程中引入裁剪机制,对新旧策略之间的变化幅度进行限制,从而避免因单次更新过大而导致训练不稳定^[19]。在 WP-MEC 环境中,PPO 的优势主要体现在更新过程较为平稳,能够在复杂约束条件下维持较稳定的策略改进过程,从而为无线充电器部署与任务卸载联合决策提供可靠的基线。柔性动作-评价(soft actor-critic, SAC)算法是一种基于最大熵强化学习思想的离策略 Actor-Critic 方法,该方法在最大化长期累计回报的同时,引入策略熵项以增强探索能力,从而提升算法在复杂连续动作空间中的搜索效率^[20]。SAC 的优势主要表现为能够利用其最大熵机制维持更充分的策略探索,并借助离策略学习加快早期收敛

过程,因此在动态信道变化和任务负载波动条件下,更容易较快获得性能较优的部署与卸载策略。

图 4(a)和(b)表明,在不同计算任务数据量和不同 CPU 频率的仿真场景下,GTXL-DRL 的任务计算完成率均优于 5 种对比框架。当计算任务数据量较大或无线设备 CPU 频率较低时,对比框架的计算完成率下降明显,而 GTXL-DRL 仍能维持较高水平。SAC 通过最大熵机制保证了较好的策略探索空间,但在处理高负载下的长序列依赖问题时,缺乏对历史信道和能量状态的记忆,导致在高数据量时难以维持较高的计算完成率。PPO 对动态复杂环境的适应性较弱,在任务量较小、系统资源宽裕时能维持不错的完成率,但随着任务量增加,PPO 保守的策略更新机制使其难以及时跳出次优策略,导致任务难以计算完成。DQN 框架虽然依靠经验回放机制维持了中等水平的性能,但缺乏时序记忆导致其在高负载下表现乏力。Policy Gradient 与 Reinforce 框架由于更新方差大、对复杂动态环境适应性差,在高负载场景下性能严重退化,计算完成率出现了大幅下降。上述的对比实验结果表明,GTXL-DRL 框架在同样的 WP-MEC 环境中,通过引入 GTXL,更好地利用了多时间帧的信道和任务信息,获得更高的任务计算完成率和更稳定的决策性能,验证了 GTXL 在捕捉长期依赖与优化高维动作的优势,并且 GTXL 擅长提取全局特征,倾向于算出接近最优的解。

为了进一步研究 GTXL 中门控单元的作用机理,本文将仿真环境划分为低信噪比和高信噪比两种典型场景。如图 5(a)所示,在 $SNR = 5$ dB 的恶劣信道环境下,由于缺乏门控机制的过滤,模型对瞬时噪声过拟合,导致收敛曲

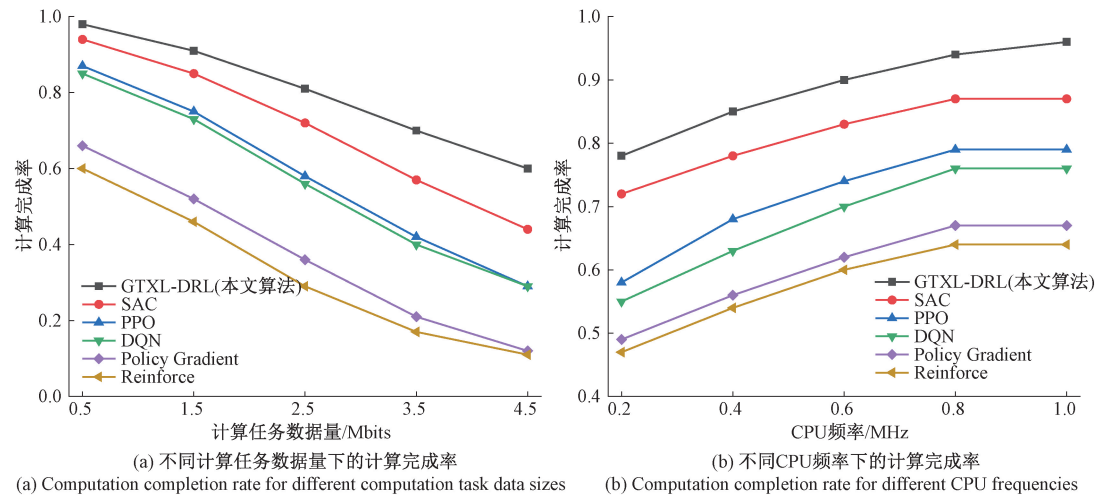


图 4 不同算法的性能对比

Fig. 4 Performance comparison of different algorithms

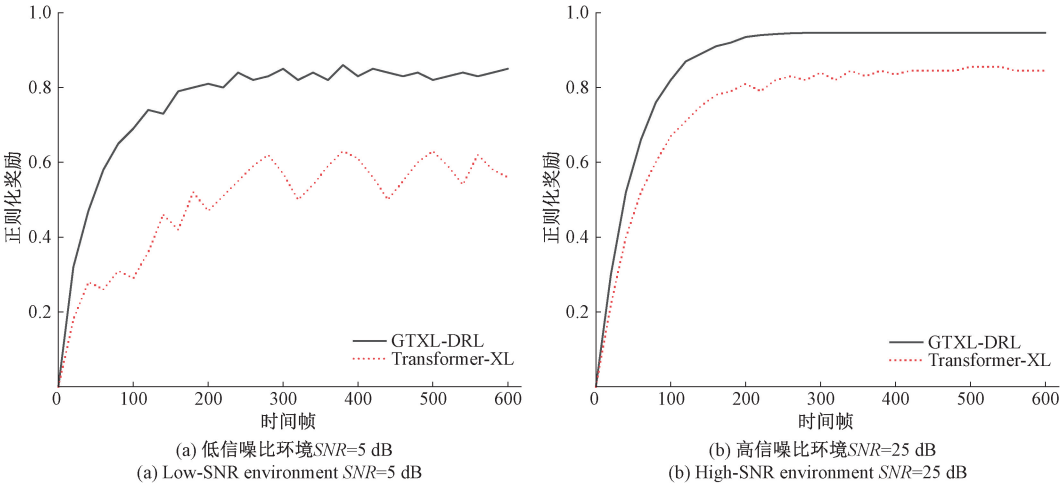


图 5 门控机制在不同信噪比环境下的性能对比

Fig. 5 Performance comparison of the gating mechanism under different signal-to-noise ratio environments

线剧烈震荡,且最终奖励值较低。相比之下,GTXL-DRL 通过门控单元自适应地降低了高噪声的影响,展现出极强的抗噪鲁棒性。而在图 5(b)所示的 $SNR = 25$ dB 理想环境下,虽然两者差距缩小,但 GTXL-DRL 仍能凭借对微小波动的平滑处理,实现更快的收敛速度与更高的稳定性。这一对比实验证明了门控机制是解决边缘侧高动态、高噪声问题的关键。

图 6 研究了设备数量对 GTXL-DRL 算法计算完成率的影响。从图 6(a)中可以看出,随着无线设备数量的增加,所有算法的平均计算完成率均呈现下降趋势。这是由于在无线频谱带宽与 MEC 服务器算力总量固定的前提

下,设备数量的增加导致了更激烈的资源竞争。值得注意的是,GTXL-DRL 的下降斜率明显缓于其余 3 种算法。这得益于其 Transformer 结构能够有效对多设备间的干扰关系进行全局建模,通过更精细的时隙调度来缓解资源冲突,维持较高的系统吞吐量。从图 6(b)中可以看出,随着无线充电器数量的增加,模型的计算完成率越高,这是因为更多的无线充电器使无线设备收集到了更多的能量。另外,从图 6 的时间维度来看,无论设备或充电器数量如何变化,GTXL-DRL 算法均收敛至稳定状态。这进一步验证了所提算法在动态 WP-MEC 环境下,能够快速自适应调整策略,具有良好的收敛稳定性。

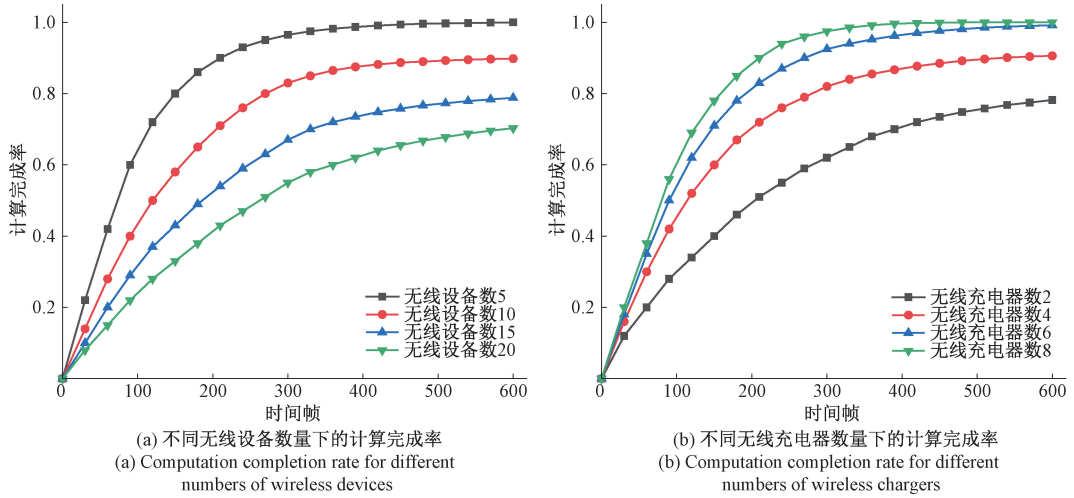


图 6 无线设备与无线充电器数量对计算完成率的影响

Fig. 6 Effect of the numbers of wireless devices and wireless chargers on the computation completion rate

4 结 论

本文针对边缘侧任务到达的动态性,固定位置的无线充电器很难维持高的任务计算完成率,考虑了 WP-MEC 网络中多个无线充电器部署位置的优化。本文的目标是通过联合优化无线充电器部署决策、无线设备任务卸载比率和资源分配以最大化无线设备计算完成率之和。为了解决高维度的混合整数非凸问题,本文把原问题分成了两个子问题。为了解决传统算法无法有效捕捉长期依赖关系且易受噪声干扰的缺陷,提出了一种 Transformer 驱动的移动边缘计算任务卸载和无线充电器部署算法。仿真实验验证了 GTXL-DRL 在提升无线设备计算完成率方面的优越性,以及门控机制降低噪声干扰有效性。未来工作将进一步结合无人机搭载和可开关天线阵列等更贴近实际场景的在线部署机制,并将位置迁移、波束切换等部署代价及其计算开销纳入统一优化模型。

参考文献

[1] 陈晓,仇洪冰,李燕龙. 边缘辅助的自适应稀疏联邦学习优化算法[J]. 电子与信息学报, 2025, 47(3): 645-656.
CHEN X, QIU H B, LI Y L. Adaptively sparse federated learning optimization algorithm based on edge-assisted server[J]. Journal of Electronics & Information Technology, 2025, 47(3): 645-656.

[2] 许方耀,池凯凯,毛科技,等. 无线供能下基于节点间合作的计算卸载方案[J]. 传感技术学报, 2025, 38(5): 886-893.
XU F Y, CHI K K, MAO K J, et al. Computation offloading scheme based on inter-node cooperation under wireless power transfer[J]. Chinese Journal of Sensors and Actuators, 2025, 38(5): 886-893.

[3] HUANG L, BI S ZH, ZHANG Y J A. Deep reinforcement learning for online computation offloading in wireless powered mobile-edge computing networks [J]. IEEE Transactions on Mobile Computing, 2019, 19(11): 2581-2593.

[4] DENG X M, LI J, SHI L, et al. Wireless powered mobile edge computing: Dynamic resource allocation and throughput maximization[J]. IEEE Transactions on Mobile Computing, 2020, 21(6): 2271-2288.

[5] ZHANG L J, SONG Q Y, WU M R, et al. Joint terminal pairing and multi-dimensional resource allocation for cooperative computation in a WP-MEC system [J]. IEEE Transactions on Green Communications and Networking, 2023, 7(3): 1447-1456.

[6] LIU X Y, CHEN A P, ZHENG K CH, et al. Distributed computation offloading for energy provision minimization in WP-MEC networks with multiple HAPs[J]. IEEE Transactions on Mobile Computing, 2024, 24(4): 2673-2689.

[7] 戴海鹏,陈贵海,徐力杰,等. 一种高效有向无线充电器的布置算法[J]. 软件学报, 2015, 26(7): 1711-1729.
DAI H P, CHEN G H, XU L J, et al. Effective algorithm for placement of directional wireless chargers[J]. Journal of Software, 2015, 26(7): 1711-1729.

[8] DAI H P, ZHANG Y K, WANG W J, et al. Placing wireless chargers with multiple antennas[J]. IEEE Transactions on Mobile Computing, 2023, 23(6): 7517-7536.

[9] NEELAGANDAN P, BALAJI S, PAVITHRA R. Advanced charger placement strategies in sensor

- networks using graph theory and evolutionary algorithms[J]. IEEE Sensors Journal, 2025, 25(18): 35609-35621.
- [10] YU N, DAI H P, CHEN L, et al. Fault-tolerant wireless charger placement[J]. IEEE Transactions on Mobile Computing, 2024, 24(5): 3903-3917.
- [11] 熊漩,严佩敏. 融合多头自注意力机制的中文分类方法[J]. 电子测量技术, 2020, 43(10): 125-130.
- XIONG X, YAN P M. Chinese classification method based on multi-head self-attention [J]. Electronic Measurement Technology, 2020, 43(10): 125-130.
- [12] MACH P, BECVAR Z. Mobile edge computing: A survey on architecture and computation offloading[J]. IEEE Communications Surveys & Tutorials, 2017, 19(3): 1628-1656.
- [13] TANG Q, LYU H M, HAN G J, et al. Partial offloading strategy for mobile edge computing considering mixed overhead of time and energy[J]. Neural Computing and Applications, 2020, 32(19): 15383-15397.
- [14] GUO S T, XIAO B, YANG Y Y, et al. Energy-efficient dynamic offloading and resource scheduling in mobile cloud computing[C]. IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications. IEEE, 2016: 1-9.
- [15] 郭澳归,叶成荫,申佳欣. 能耗约束下量子遗传算法时延最小化卸载策略[J/OL]. 电子测量技术, 1-7[2026-03-30]. <https://link.cnki.net/urlid/11.2175.tn.20251231.1321.003>.
- GUO AO G, YE CH Y, SHEN J X. Energy consumption constraint-based quantum genetic algorithm delay minimization offloading strategy [J/OL]. Electronic Measurement Technology, 1-7[2026-03-30]. <https://link.cnki.net/urlid/11.2175.tn.20251231.1321.003>.
- [16] WANG ZH Y, GOUDARZI M, BUYYA R. TF-DDRL: A transformer-enhanced distributed DRL technique for scheduling IoT applications in edge and cloud computing environments[J]. IEEE Transactions on Services Computing, 2025, 18(2): 1039-1053.
- [17] WANG H, LIU CH H, YANG H M, et al. Ensuring threshold AoI for UAV-assisted mobile crowdsensing by multi-agent deep reinforcement learning with transformer [J]. IEEE/ACM Transactions on Networking, 2023, 32(1): 566-581.
- [18] 黄鑫,张志佳,孙平,等. 基于深度强化学习的路径规划算法综述[J]. 机器人, 2026, 48(01): 196-216.
- HUANG X, ZHANG ZH J, SUN P, et al. A survey of path planning algorithms based on deep reinforcement learning[J]. Robot, 2026, 48(01): 196-216.
- [19] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimization algorithms [J]. ArXiv preprint arXiv: 1707. 06347, 2017.
- [20] HAARNOJA T, ZHOU A, ABBEEL P, et al. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor [C]. International conference on machine learning. PMLR, 2018: 1861-1870.

作者简介

贾本源, 硕士研究生, 主要研究方向为无线通信与人工智能。

E-mail: 13994270272@163.com

刘祖深(通信作者), 博士, 研究员, 主要研究方向为微波通信测量技术与仪器的研究与开发。

E-mail: liuzushen@ceyear.com

铁奎, 研究员, 主要研究方向为移动通信测试技术。

E-mail: tiekui@ceyear.com