

智能反射表面辅助的车载边缘任务卸载和资源分配策略

吴华明, 梁永辉

(天津大学应用数学中心, 天津 300072)

摘 要: 针对车载边缘计算场景对低时延、高能效与安全性的需求, 结合智能反射表面技术, 提出了一种基于深度强化学习的任务卸载与资源分配策略。首先, 不同于仅优化单一指标的方法, 构建了时延、能耗与任务完成率的加权综合指标作为系统优化目标。其次, 提出了一种具有任务完成感知能力的资源优化型双时延深度确定性策略梯度算法。该算法通过引入基于任务完成率的奖励函数, 并结合成功轨迹优先回放机制, 有效增强了稀疏环境下的学习稳定性, 从而提升了在通信暗区或链路不稳定环境下的卸载成功率和资源利用效率。实验结果表明, 与现有基线算法相比, 该算法可将系统平均成本降低约 25%, 平均时延降低 23.2%, 平均能耗降低 17.6%, 并使任务完成率提升 7%。

关键词: 车载边缘计算; 任务卸载; 资源分配; 深度强化学习; 智能反射表面

中图分类号: TN929.5

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2025171

Task offloading and resource allocation strategy for vehicular edge computing assisted by intelligent reflecting surfaces

WU Huaming, LIANG Yonghui

Center for Applied Mathematics, Tianjin University, Tianjin 300072, China

Abstract: To address the demands for low latency, high energy efficiency, and security in vehicular edge computing scenarios, a task offloading and resource allocation strategy based on deep reinforcement learning, combined with intelligent reflecting surface technology was proposed. Firstly, unlike approaches that optimized a single objective, a weighted comprehensive index of delay, energy consumption, and task completion rate was constructed as the system optimization objective. Furthermore, a resource-optimized twin-delayed deep deterministic policy gradient algorithm was designed, which was capable of perceiving task completion status. By introducing a task completion rate-based reward function and incorporating a successful trajectory prioritized replay mechanism, the learning stability of the algorithm in sparse environments was enhanced, thereby improving offloading success rate and resource utilization efficiency under poor signal coverage or unstable links. Experimental results demonstrated that, compared with state-of-the-art baseline algorithms, the proposed method reduces average system cost by about 25%, average latency by 23.2%, and average energy consumption by 17.6%, while increasing the task completion rate by 7%.

Keywords: vehicular edge computing, task offloading, resource allocation, deep reinforcement learning, intelligent reflecting surface

收稿日期: 2025-07-22; 修回日期: 2025-09-02

通信作者: 吴华明, whming@tju.edu.cn

基金项目: 国家自然科学基金资助项目(No.62071327);天津市科技计划基金资助项目(No.22ZYYYJC00020);天津大学学科交叉中心新兴方向培育专项基金资助项目

Foundation Items: The National Natural Science Foundation of China (No.62071327), Tianjin Science and Technology Planning Project (No.22ZYYYJC00020), Emerging Frontiers Cultivation Program of Tianjin University Interdisciplinary Center

0 引言

作为物联网的重要组成部分,车联网(IoV, Internet of vehicles)的核心目标是实现车辆之间、车辆与基础设施以及其他设备之间的高效互联互通,从而提升道路安全、交通效率和整体运输系统的智能化水平^[1]。随着IoV规模的扩大,大量具有实时性和敏感性特征的数据在车辆终端不断涌现^[2]。然而,车辆的本地计算资源有限,难以有效处理实时性强、计算密集型的任务,从而导致事故风险升高^[3]。车载边缘计算(VEC, vehicular edge computing)技术的兴起为解决上述挑战提供了一种有效方案,被广泛认为能够有效解决车联网环境中的计算瓶颈问题^[4]。VEC允许车辆将自身的计算任务卸载到道路附近部署的路侧单元(RSU, road-side unit)上进行处理,能够显著降低车辆端计算时延,提升IoV环境下任务处理效率^[5-6]。VEC技术不仅能够缓解车载计算压力,降低数据处理时延,同时有助于减少核心网络的数据负载,降低网络拥塞风险,同时也可提升数据安全性^[7-9]。

然而,车联网环境的高度动态性和多变性为任务卸载带来了诸多挑战^[10]。特别是在复杂城市环境或高速公路场景中,车与车、车与RSU之间的无线链路常因遮挡、多径效应或频繁变化的信道条件而面临严重的信号衰减与不稳定,导致任务卸载失败率显著上升,严重影响系统的可靠性与稳定性^[11]。为了应对这一问题,智能反射表面(IRS, intelligent reflecting surface)作为一种新兴的可重构环境感知技术,逐渐受到广泛关注^[12-13]。IRS通过大量低功耗的可编程反射单元,能对入射信号进行智能调控,从而使信号绕过遮挡区域,并增强接收信号的质量^[14-15]。这种特性使IRS在提升任务卸载成功率、优化无线资源利用方面具有显著潜力,为构建高效、可靠的车载边缘计算系统提供了新的解决思路^[16]。

目前,已有一些学者在IRS辅助的IoV方面展开了积极探索。例如,Qi等^[17]使用深度学习算法对车载用户功率分配进行优化,并结合块坐标下降算法求解IRS相位调整问题。Xie等^[18]提出了一种双IRS辅助的车载边缘计算架构,并通过深度强化学习和分数规划方法联合优化资源分配与相移控制,以提升任务卸载效率并实现实时决策。Wahid等^[19]聚焦于提升车联网计算卸载过程中的安全性与效率,引入零能耗智能反射表面辅助的深度强化学习卸载方案,

在保障通信机密性的同时,可以动态切换连接以优化任务卸载效率。Saleem等^[20]针对高频通信中信号损耗严重的问题,提出了一种面向时延感知的计算卸载框架,通过在车辆与RSU之间部署IRS,有效提升服务性能并显著降低时延。此外,一些学者也探索了将最新的深度强化学习算法和IRS相结合的研究。例如,在非正交多址接入的无线供能联邦学习网络中,Alishahi等^[21]引入IRS以辅助能量传输与数据通信,并提出了一种旨在最小化系统总能耗的联合资源分配方案。Sharma等^[22]将IRS应用于无人机通信,旨在利用IRS重构无线环境的能力来提升网络的安全性和覆盖范围,并基于此设计了一种结合联邦学习与长短期记忆网络的信道跟踪算法,利用前者进行安全和预估,并借助后者强大的序列数据处理能力来完成信道跟踪任务。

尽管如此,目前该领域仍面临两大核心挑战。首先,车载场景本身高度动态,涉及车辆的持续移动、信道状态的随机变化及任务特性的差异性,因此需要设计一种能够实时适应环境变化的联合优化策略,以协同决策任务卸载路径、通信资源分配、计算资源分配以及IRS相位矩阵配置^[23]。其次,现有研究多围绕以系统单一指标为核心的优化目标,难以在满足时延要求下同时兼顾整体能耗与任务完成率,因此迫切需要一种能够权衡多种性能指标、实现系统级综合优化的调度策略^[24]。

针对上述挑战,本文旨在深入研究IRS辅助下的车载边缘计算卸载与资源分配问题,重点关注在多目标约束下实现高可靠性与任务完成率、低时延与能耗的系统性能提升。本文首先构建了一个融合车辆任务卸载机制、通信与计算资源分配以及IRS反射相位调控机制的系统模型。在此基础上,提出一种基于双时延深度确定性策略梯度(TD3, twin delayed deep deterministic policy gradient)算法的感知任务完成情况的深度强化学习算法,通过深度强化学习框架实现对连续动作空间的高效探索,从而在动态环境中学习最优卸载与资源调度策略。最后,仿真实验验证了本文算法在多种复杂场景下的性能表现,表明本文算法相较于其他基线算法能显著提升任务卸载成功率,降低系统平均成本,并具备良好的鲁棒性和泛化能力,为车载边缘场景中的任务协同计算提供了有效解决方案。本文的主要工作如下。

1) 构建了一个 IRS 辅助的车载边缘计算场景, 建立了 VEC 系统的通信模型、时延模型、能耗模型和任务完成率模型, 全面刻画了计算任务在卸载过程中的关键性能指标。进一步, 构建了时延、能耗与任务完成率的综合性加权指标, 以作为系统的优化目标。

2) 提出了一个基于 TD3 算法的感知任务完成情况的深度强化学习算法, 该算法通过采用基于任务完成率的奖励函数, 并结合成功轨迹优先回放机制, 有效提升了通信暗区或链路不稳定环境下的卸载成功率和资源利用效率。

3) 通过仿真实验把本文算法和多种基线算法进行了性能对比, 结果表明本文算法可以显著降低系统时延和能耗, 同时显著提升任务完成率, 在动态车载边缘计算环境中表现出良好的稳定性与扩展性。

1 系统模型

1.1 IRS 辅助的车载边缘计算系统

考虑一个城市道路场景中智能反射表面辅助的车载边缘计算系统, 如图 1 所示, 场景中有车辆、配备边缘服务器的基站、智能反射表面以及一定范围的通信暗区。记车辆集为 $\mathcal{N} = \{1, 2, \dots, N\}$ 。本文采用一个离散时隙计算模型, 即将总时间划分为 M 个相等的时隙, 每个时隙 t 的持续时间记为 δ 。因此, 定义时间集合为 $\mathcal{T} = \{1, 2, \dots, M\}$ 。假设每辆车 n 在每个时隙 t 内生成一个计算任务, 表示为

$$\Theta_n = (D_n(t), C_n(t), T_n^{\max}(t)) \quad (1)$$

其中, $D_n(t)$ 表示任务数据大小, $C_n(t)$ 表示计算密度, $T_n^{\max}(t)$ 表示任务的最大可容忍时延。

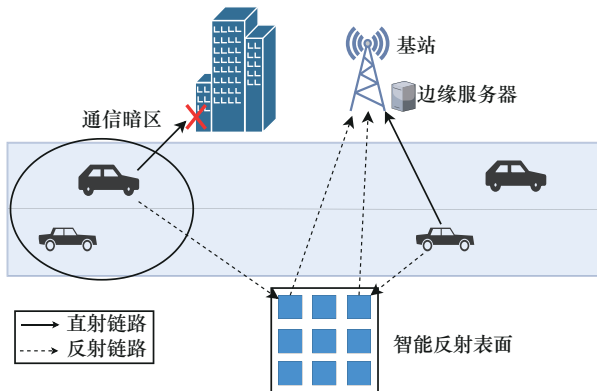


图1 智能反射表面辅助的车载边缘计算场景

由于许多实际任务具备可分性, 可通过设定卸载比例将任务划分为本地处理与远程处理部分。同时, 5G 移动边缘计算 (MEC) 架构与实际平台也已支持任务分片与动态卸载^[25]。因此本文采用部分卸载策略, 即车辆可以选择本地执行部分任务或卸载部分任务至边缘服务器, 设卸载比例为 $\alpha_n(t) \in [0, 1]$, 则本地执行任务大小为 $(1 - \alpha_n(t))D_n(t)$, 卸载任务大小为 $\alpha_n(t)D_n(t)$ 。系统中各实体间的交互为: 车辆通过直射链路或者通过智能反射表面建立折射链路, 可以把计算任务卸载至边缘服务器处, 边缘服务器处理完计算任务后, 再通过相同的链路把计算结果返回车辆。

1.2 通信模型

根据通信理论, 信号的复数表示为

$$x(t) = A(t) \cdot e^{j\omega(t)} \quad (2)$$

其中, $A(t)$ 表示振幅, $\omega(t)$ 表示相位。为简便起见, 此处把振幅设为 1, 即 $A(t)=1$ 。

设 IRS 反射相位矩阵为

$$\Phi = \text{diag}(e^{j\theta_1}, e^{j\theta_2}, \dots, e^{j\theta_N}) \quad (3)$$

从车辆 n 到边缘服务器的信号总增益可表示为

$$H_k = h_k s + h_{i,k} \Phi h_{i,b} s \quad (4)$$

其中, h_k 为直射路径上的信道增益, $h_{i,k}$ 是从车辆到 IRS 的信道增益, $h_{i,b}$ 是从 IRS 到基站边缘服务器的信道增益, s 是发送的信号。需要说明的是, 为突出重点关注的“卸载与资源联合决策”而不过度扩展物理层建模, 采用“块衰落/时标分离”假设: 在每个调度时隙内, 直射链路及经由 IRS 的级联链路的等效增益近似不变, 而相邻时隙之间则因小尺度衰落和多普勒效应而发生快变随机更新。本文涉及的“增益”均指“每时隙内恒定、跨时隙随机的有效增益”, 并非全局固定常数; 其作用相当于在更细的物理层模型 (如考虑小尺度衰落与多普勒效应的统计信道) 下对单个时隙进行等效平均, 从而便于聚焦系统层面的联合决策与算法设计。

因此, 信噪比 (SNR) 可表示为

$$\text{SNR} = \frac{P \cdot G}{\sigma^2} = \frac{P \cdot \left| (h_k + h_{i,k} \Phi h_{i,b}) \cdot \omega_k \right|^2}{\sigma^2} \quad (5)$$

其中, ω_k 表示基站处的波束成形矢量, P 表示发射功率, G 表示有效信道增益, σ 表示高斯噪声。

因此, 数据传输速率可表示为

$$R = B \cdot \lg(1 + \text{SNR}) = B \cdot \lg \left(1 + \frac{P \cdot \left| (h_k + h_{i,k} \Phi h_{i,b}) \cdot \omega_k \right|^2}{\sigma^2} \right) \quad (6)$$

为简化起见, 本文采用最大比传输方式, 将发射信号方向与信道增益方向对齐, 通过简单的设计, 增强目标用户的信号质量, 降低后续优化的复杂性。故波束成形矢量 ω_k 可以表示为

$$\omega_k = \frac{h_k + h_{i,k} \Phi h_{i,b}}{h_k + h_{i,k} \Phi h_{i,b}} \quad (7)$$

1.3 时延模型

设本地计算时延为 $T_n^l(t)$, 则有

$$T_n^l(t) = \frac{(1 - \alpha_n(t)) D_n(t) C_n(t)}{f_n^l(t)} \quad (8)$$

其中, $f_n^l(t)$ 代表车辆 n 在时隙 t 的计算能力。

由于计算结果的数据量通常很小, 为了简便起见, 这里忽略结果的回传时间。因此任务上传至边缘端处理这个过程的总时延由3个部分组成, 分别是上传时延 $T_n^u(t)$ 、边缘端排队时延 $T_n^q(t)$ 以及边缘端计算时延 $T_n^e(t)$ 。由于任务在边缘侧可能出现积压排队情况, 故设车辆 n 在边缘服务器处的队列长度为 $Q_n(t)$, 则其迭代式为

$$Q_n(t+1) = \max \{ Q_n(t) - f_n(t) \Delta t, 0 \} + \alpha_n(t) D_n(t) c_n(t) \quad (9)$$

其中, 第一项 $\{ Q_n(t) - f_n(t) \Delta t \}$ 表示在时隙 t 内处理完的任务量, 第二项 $\{ \alpha_n(t) D_n(t) c_n(t) \}$ 表示时隙 t 内新进入队列的任务量。

因此, $T_n^u(t)$ 、 $T_n^q(t)$ 、 $T_n^e(t)$ 可分别表示为

$$T_n^u(t) = \frac{\alpha_n(t) D_n(t)}{R_n(t)} \quad (10)$$

$$T_n^q(t) = \frac{Q_n(t)}{f_n^e(t)} \quad (11)$$

$$T_n^e(t) = \frac{\alpha_n(t) D_n(t) c_n(t)}{f_n^e(t)} \quad (12)$$

其中, $f_n^e(t)$ 表示边缘服务器分配给车辆 n 的计算资源。

因此, 车辆 n 在时隙 t 的总时延 $T_n(t)$ 为

$$T_n(t) = T_n^l(t) + T_n^u(t) + T_n^q(t) + T_n^e(t) \quad (13)$$

时隙 t 内所有车辆的总时延 $T(t)$ 为

$$T(t) = \sum_{n=1}^N T_n(t) \quad (14)$$

1.4 能耗模型

由于不考虑边缘服务器的能耗, 系统总能耗由本地计算能耗和传输能耗2个部分组成。设 $E_n^l(t)$ 代表本地车辆 n 在时隙 t 的本地计算能耗, 则有

$$E_n^l(t) = \kappa T_n^l(t) (f_n^l(t))^3 \quad (15)$$

其中, κ 是计算能耗参数。

由于忽略了计算结果回传时间, 此处也同样不考虑结果下传能耗, 故传输能耗仅由计算任务上传能耗组成。设任务上传能耗为 $E_n^u(t)$, 则有

$$E_n^u(t) = \frac{D_n(t)}{R_n(t)} P_n(t) \quad (16)$$

因此, 车辆 n 在时隙 t 的总能耗 $E_n(t)$ 为

$$E_n(t) = E_n^l(t) + E_n^u(t) \quad (17)$$

时隙 t 所有车辆的总能耗 $E(t)$ 为

$$E(t) = \sum_{n=1}^N E_n(t) \quad (18)$$

1.5 任务完成率模型

为综合评估系统性能, 在时延与能耗加权成本的基础上, 引入任务完成率作为系统处理能力的衡量指标^[26]。车辆 n 的任务若在最大容忍时延 $T_n^{\max}$ 内完成, 则视为“成功完成”。因此, 系统在时隙 t 的任务完成率 $\eta(t)$ 可定义为

$$\eta(t) = \frac{\sum_{n=1}^N \Gamma(t) D_n(t)}{\sum_{n=1}^N D_n(t)} \quad (19)$$

其中, $\Gamma(t)$ 为指示函数, 定义为

$$\Gamma(t) = \begin{cases} 1, & T_n(t) \leq T_n^{\max}(t) \\ 0, & T_n(t) > T_n^{\max}(t) \end{cases} \quad (20)$$

1.6 优化目标

为鼓励系统在保证能效与时延表现的同时尽可能地提高整体任务完成能力, 本文构建的改进型系统成本函数为

$$U(t) = \frac{\sum_{n=1}^N (\omega_1 T_n(t) + \omega_2 \beta E_n(t))}{\eta(t)} \quad (21)$$

其中, ω_1 与 ω_2 分别表示时延与能耗的权重系数, β 表示数值归一化系数。该目标函数能够在优化时延与能耗的同时, 提高系统整体计算服务的稳定性与可用性。设 x 、 f 、 p 、 θ 分别代表卸载比例、计算资源、

通信功率和智能发射表面反射相移, F_n^l 和 F_n^e 分别表示单个车辆可以使用的最大本地和边缘计算资源。

本文的优化目标为最小化系统成本, 即

$$\begin{aligned}
 & \min_{(x,f,P,\theta)} \frac{1}{M} \sum_{t=1}^M U(t) \\
 & \text{s.t. } C_1: x_n(t) \in [0, 1], \forall n \in \mathcal{N}, \forall t \in \mathcal{T} \\
 & C_2: 0 \leq f_n^e(t) \leq F_n^e, \forall n \in \mathcal{N}, \forall t \in \mathcal{T} \\
 & C_3: 0 \leq f_n^l(t) \leq F_n^l, \forall n \in \mathcal{N}, \forall t \in \mathcal{T} \\
 & C_4: 0 \leq P_n(t) \leq P_n^{\max}, \forall n \in \mathcal{N}, \forall t \in \mathcal{T} \\
 & C_5: 0 \leq \theta_n(t) \leq \pi, \forall n \in \mathcal{N}, \forall t \in \mathcal{T} \\
 & C_6: \omega_1 + \omega_2 = 1 \\
 & C_7: 0 < \eta(t) \leq 1
 \end{aligned} \quad (22)$$

其中, C_1 表示任务卸载比例介于 0 与 1 之间, 0 代表完全本地处理, 1 代表完全卸载至边缘端; C_2 和 C_3 分别表示车辆和边缘服务器的最大计算资源限制; C_4 表示车辆的最大发射功率限制; C_5 表示智能反射表面的相位角度限制; C_6 表示时延和能耗的权重加和为 1; C_7 表示时隙 t 内系统的总任务完成率的取值范围。

2 基于深度强化学习的感知型 TD3 算法

针对车载边缘场景中卸载决策、资源分配与 IRS 相位调控之间高度耦合、决策变量维度高、环境状态变化剧烈等挑战, 传统的确定性优化方法难以适应复杂动态环境^[27]。为此, 本文提出一种基于深度强化学习的感知型 TD3 (CR-TD3, completion-aware resource-optimized TD3) 算法, 用于解决车载边缘计算中的卸载与资源分配问题。

2.1 马尔可夫过程建模

本文给出马尔可夫过程的各个基本要素, 包括状态空间、动作空间和奖励函数。

设 $s_n(t)$ 为车辆 n 在时刻 t 的状态空间。状态空间 $s_n(t)$ 包含车辆的计算任务 $\phi_n(t)$ 、车辆速度 $v_n(t)$ 、车辆位置 $l_n(t)$ 、本地最大计算资源 F_n^l 、边缘端的最大计算资源 F_n^e 、最大通信功率 $P_n^{\max}$ 、边缘队列长度 $Q_n(t)$ 。因此, 系统在时刻 t 的整体状态空间可表示为

$$S(t) = (s_1(t), \dots, s_n(t), \dots, s_N(t)) \quad (23)$$

其中, $s_n(t) = [\phi_n(t), l_n(t), v_n(t), F_n^l, F_n^e, P_n^{\max}, Q_n(t)]$ 。

动作空间由以下几个部分组成: 任务卸载比例 $x_n(t)$ 、智能反射表面相移角度 $\theta_n(t)$ 、本地算力

$f_n^l(t)$ 、分配给车辆的边缘侧算力 $f_n^e(t)$ 、 $P_n(t)$ 。因此, 动作空间可表示为

$$A(t) = (a_1(t), \dots, a_n(t), \dots, a_N(t)) \quad (24)$$

其中, $a_n(t) = [x_n(t), \theta_n(t), f_n^l(t), f_n^e(t), F_n^e, P_n(t)]$ 。

本文所构建的奖励函数定义为目标函数 $U(t)$ 的相反数, 即

$$R(t) = -U(t) = -\frac{\sum_{n=1}^N (\omega_1 T_n(t) + \omega_2 \beta E_n(t))}{\eta(t)} \quad (25)$$

2.2 基于 TD3 的深度强化学习算法

Fujimoto 等^[28]提出的 TD3 算法显著提升了在连续控制任务中的性能与稳定性, 成为当前最常用的深度强化学习算法之一。TD3 算法基于演员-评论家 (Actor-Critic) 架构, 采用双 Q 网络结构以及时延策略更新。这样可以先让 Critic 网络收敛更准确的价值估计, 再指导 Actor 学习, 有助于策略稳定。TD3 算法引入目标策略扰动, 在生成的目标动作中加入噪声, 有效缓解策略对某些值函数“陡峭区域”的过度依赖。尽管 TD3 算法在标准连续控制任务中性能优秀, 直接应用于车载边缘计算仍存在很多不足, 例如缺乏对系统长期表现的度量, 无法优化全局任务完成能力; 缺少资源约束意识, 易输出不满足系统限制的决策等^[29]。

为解决上述问题, 本文提出 CR-TD3 算法, 该算法保持了 TD3 算法的稳定优势, 在原始 TD3 算法框架基础上, 结合车联网动态任务卸载特性进行关键性改进。

1) 引入任务完成率作为全局反馈信号, 更全面地反映系统性能, 避免策略陷入局部最优。

2) 联合建模高维动作空间, 覆盖卸载比例、通信功率、计算资源及智能反射表面相位角, 显著提升策略的综合能力。

3) 设计具有资源约束感知的奖励函数, 通过集成任务完成率的归一化系统成本, 引导策略向“高效+高完成率”的方向收敛。

4) 采用成功轨迹优先经验回放机制, 加快策略在稀疏奖励环境中的学习效率。

CR-TD3 算法通过强化策略网络的长期感知能力与系统约束感知能力, 通过双 Critic、时延更新与轨迹筛选机制, 不仅提升了训练稳定性, 也增强了算法对复杂动态环境的适应能力, 在车辆数量波

动、信道质量变化、任务需求多样化的场景下,依然能保持较高的任务完成率与系统效能。

Critic 网络目标值使用最小 Q 值策略构建。

$$y_i = r_i + \gamma \cdot \min_{j=1,2} Q_{\phi_j}(s_{i+1}, a'_{i+1}) \quad (26)$$

其中, $a'_{i+1} = \mu_{\theta}(s_{i+1}) + \varepsilon, \varepsilon \sim \mathcal{N}(0, \sigma^2)$ 。 (s_i, a_i, r_i, s_{i+1}) 样本来自成功轨迹经验池 $\mathcal{D}_{\text{success}}$, 即满足任务完成条件的历史轨迹集合, 用以强化策略对完成率的偏好。

对应的损失函数为

$$\mathcal{L}_Q(\phi_j) = \frac{1}{N} \sum_{i=1}^N \left(Q_{\phi_j}(s_i, a_i) - y_i \right)^2 \quad (27)$$

策略网络每 d 步进行一次更新, 其梯度定义为

$$\nabla_{\theta} J(\theta) = \frac{1}{N} \sum_{i=1}^N \nabla_a Q_{\phi_1}(s_i, a) \big|_{a=\mu_{\theta}(s_i)} \cdot \nabla_{\theta} \mu_{\theta}(s_i) \quad (28)$$

采用软更新策略更新目标网络参数。

$$\bar{\phi}_j^{\text{new}} \leftarrow \tau \phi_j + (1 - \tau) \bar{\phi}_j \quad (29)$$

$$\bar{\theta}_j^{\text{new}} \leftarrow \tau \theta + (1 - \tau) \bar{\theta} \quad (30)$$

CR-TD3 算法流程如图2所示, 伪代码如算法1所示。

算法1 CR-TD3

输入 $\phi_n(t), l_n(t), v_n(t), F_n^l, F_n^e, P_n^{\max}, Q_n(t)$

输出 卸载决策 x 以及资源分配策略 f, P, θ

1) 初始化评论器参数 ϕ_1, ϕ_2 , 策略网络参数 θ , 目标网络参数 $\bar{\phi}_1, \bar{\phi}_2, \bar{\theta}$; 成功轨迹经验池

$\mathcal{D}_{\text{success}}$, 软更新系数 τ , 探索噪声标准差 σ ;

2) for 每一轮训练 episode = 1 to G do

3) 初始化环境, 获取初始状态 s_0 ;

4) for 每个时间步 $t = 1$ to T do

5) 使用当前策略选择动作: $a_t = \mu_{\theta}(s_t) + N(0, \sigma)$;

6) 执行动作 a_t , 观察奖励 r_t 和新状态 s_{t+1} ;

7) 存储元组 (s_t, a_t, r_t, s_{t+1}) 到 $\mathcal{D}_{\text{success}}$;

8) 从 $\mathcal{D}_{\text{success}}$ 中随机采样 N 个样本 (s_i, a_i, r_i, s_{i+1}) ;

9) for 每个样本 i do

10) 计算目标动作: $a'_{i+1} = \mu_{\bar{\theta}}(s_{i+1}) + \varepsilon, \varepsilon \sim \mathcal{N}(0, \sigma^2)$;

11) 计算目标 Q 值: $y_i = r_i + \gamma \cdot \min_{j=1,2} Q_{\bar{\phi}_j}(s_{i+1}, a'_{i+1})$;

12) 更新评论器网络: $\mathcal{L}_Q(\phi_j) = \frac{1}{N} \cdot \sum_{i=1}^N \left(Q_{\phi_j}(s_i, a_i) - y_i \right)^2$;

13) end for

14) if 当前为策略时延更新步 then

15) 使用策略梯度更新 actor 网络:

$$\nabla_{\theta} J(\theta) = \frac{1}{N} \sum_{i=1}^N \nabla_a Q_{\phi_1}(s, a) \nabla_{\theta} \mu_{\theta}(s);$$

16) 执行软更新:

$$\bar{\phi}_j \leftarrow \tau \phi_j + (1 - \tau) \bar{\phi}_j$$

$$\bar{\theta}_j \leftarrow \tau \theta + (1 - \tau) \bar{\theta}$$

17) end if

18) end for

19) end for

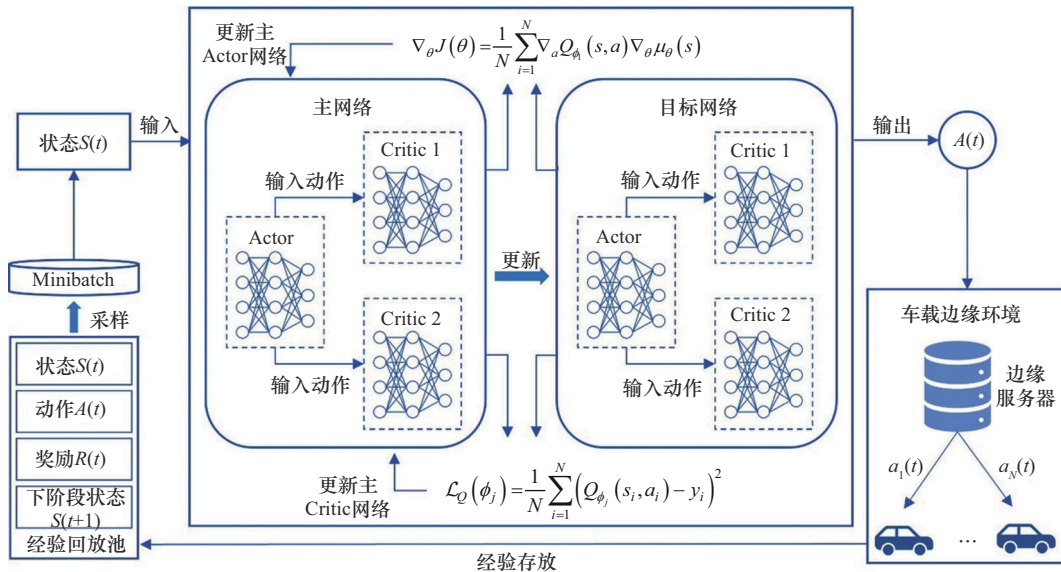


图2 CR-TD3 算法流程

需要强调的是,为确保 DRL 策略在训练与决策过程中始终满足式(22)定义的物理与资源约束,本文在 Actor 网络输出后加入了一个动作可行性映射层。该层针对不同类型的约束采用差异化机制:1) 对于卸载比例、发射功率等独立变量,通过线性变换将其映射到各自合法区间;2) 对于边缘总算力、带宽等存在总量或耦合约束的变量,采用基于 Softmax 的归一化分配方法,使分配总和自然满足预算上限;3) 对于极少数数值计算导致的边界波动,采用最终裁剪进行兜底。该机制不需要在线迭代即可生成可行性动作,有效避免了在无效区域的探索,从而显著提升了训练的稳定性与收敛速度。

3 仿真结果及分析

3.1 参数设置

所有仿真实验均在 Python 3.11 平台上进行,硬件环境包括 NVIDIA GeForce 2.1 GHz GPU、Intel Xeon Silver 4214R 2.40 GHz CPU 和 32 GB 内存。假设实验路段每 500 m 有一段 50 m 长的通信暗区。车辆以 60 km/h 的速度行驶,系统以每秒为一个时隙运行,每个小时隙产生一次大小为 0~2 Mbit 的计算任务,共设置 1 000 个时隙。设定信道带宽为 $W = 20$ MHz,能耗系数为 $\kappa = 10^{-26}$,路径损耗指数为 $l = 2.5$,高斯白噪声功率为 $N_0 = 10^{-13}$ 。仿真参数^[30-31]如表 1 所示。

表 1 仿真参数

参数	数值
车辆数量/辆	25
时隙时长/秒	1
任务数据大小/Mbit	0~2
基站覆盖范围/m	500
暗区覆盖范围/m	50
车辆速度/(km·h ⁻¹)	60
车辆计算资源上限/GHz	3
边缘服务器计算资源上限/GHz	6
最大传输功率/W	1
车辆处理每比特所需 CPU 周期数	3 000~4 500
能耗系数	10^{-26}
路径损耗因子	2.5
高斯白噪声功率/W	10^{-13}
信道带宽/MHz	20
归一化系数 β	15

3.2 数值归一化

时延和能耗的数值量级差距较大,这种不一致会使时延或能耗在实际决策过程中的作用被弱化,从而在任务卸载和资源分配策略中影响有限。为了解决这一问题,有必要对时延和能耗进行归一化处理。然而,由于计算能力和传输功率等变量在时延与能耗的计算中是随时隙而变动的,常规的归一化方法并不适用。鉴于此,本文采用了文献[32]提出的归一化方法,该方法可以将时延 T 与能耗 E 的量级调整至相同水平。具体做法是引入一个归一化系数 β ,使 $\frac{T}{\beta E} \approx 1$ 。本文在不同车辆数量条件下,对不同的 β 值进行了实验,不同 β 对 $\frac{T}{\beta E}$ 的影响如图 3 所示。结果表明,随着 β 的增大, $\frac{T}{\beta E}$ 的值逐渐减小。特别地,当 $\beta = 15$ 时,在各种车辆数量下都能观察到 $\frac{T}{\beta E} \approx 1$ 成立。因此,在后续仿真实验中,将归一化系数 β 设置为 10,以统一时延与能耗的尺度。

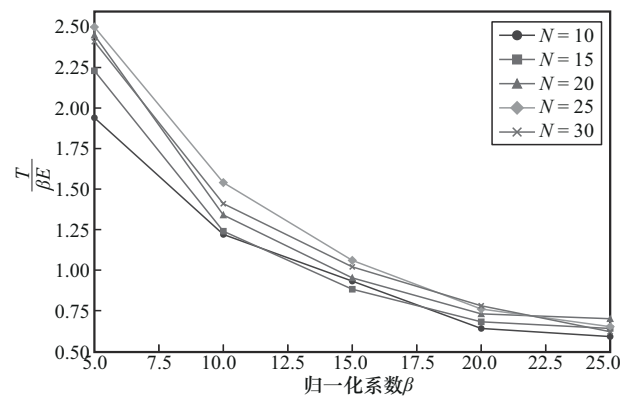


图 3 不同 β 对 $\frac{T}{\beta E}$ 的影响

3.3 对比算法

为体现 CR-TD3 算法在智能反射表面辅助的车载边缘计算场景中的整体性能与增强机制,本文选取以下 4 种具有代表性的算法作为对比对象。

1) 贪婪 (Greedy) 算法。每辆车独立选择当前时隙最利于自己的卸载比例与资源分配策略,不考虑全局协同或系统长期表现,仅优化当前决策效果。

2) 无 IRS 辅助的 CR-TD3 (CR-TD3_{NI}, non-IRS-aided CR-TD3) 算法。该算法为验证智能反射表面

的性能增益,将 IRS 带来的反射链路和通信增益去除,仅保留卸载决策与资源分配框架,用作非 IRS 辅助场景的对比基线。

3) 深度确定性策略梯度 (DDPG, deep deterministic policy gradient) 算法。经典的基于策略梯度的深度强化学习算法,用于对比 TD3 算法的性能表现。

4) 辅助车辆 (AVA, auxiliary vehicle) 算法^[30]。该算法结合联邦学习与多智能体强化学习,综合优化任务卸载、资源调度与任务完成率,代表现有文献中的先进协同卸载方案。

3.4 对比实验

图 4 为 CR-TD3 算法与 DDPG 在不同学习率下的平均奖励值变化趋势,其中阴影区域反映了不同学习率 (分别为 0.02、0.003 和 0.005) 带来的性能波动。由图 4 可知,随着训练回合数增加,2 种算法的平均奖励值均持续提升,体现出良好的学习性能和策略适应能力。其中,CR-TD3 算法在所有学习率下均表现出更高、更稳定的奖励值,尤其在训练后期性能波动明显较小,表明 CR-TD3 算法对学习率变化不敏感,具备更佳的鲁棒性,能够在多种超参数配置下稳定地收敛至更优策略。

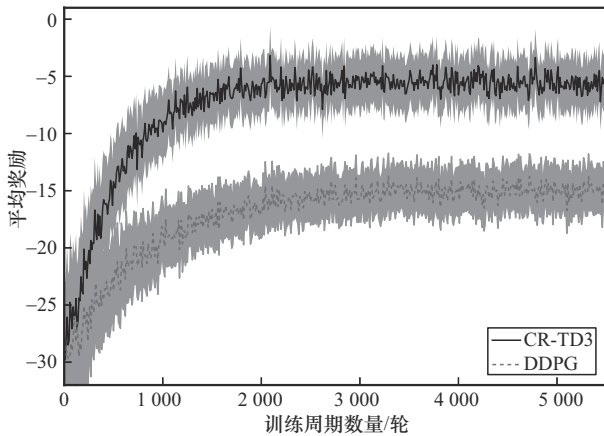


图4 CR-TD3算法与DDPG在不同学习率下的平均奖励值变化趋势

1) 系统平均成本分析

图 5 为不同车辆数量下各算法的系统平均成本。从图 5 可以看出,随着车辆数量的增加,所有算法的平均成本均有所上升,这主要是因为系统计算负载加重,通信资源竞争加剧,导致任务时延和能耗同步上升。Greedy 因采用局部贪婪策略,无法有效协调资源分配,导致在高负载情况下频繁出现

卸载失败和资源冲突,系统成本最高。DDPG 和 AVA 作为已有的深度强化学习基线方案,在小规模车辆场景下具有一定的优化能力,但随着车辆增多,其策略更新效率和泛化能力有限,表现逐渐趋于平庸。CR-TD3_{Ni} 虽具备深度强化学习优化框架,但缺乏 IRS 的反射链路的辅助和信道增强作用,在链路遮挡场景中传输失败率高,导致总体成本偏高。相比之下,CR-TD3 算法在所有车辆规模下均保持最低的系统平均成本,这得益于智能反射表面所提供的额外反射链路与信道增益,有效提升了卸载成功率,降低了传输重试与任务失败所带来的额外开销,体现出良好的全局调度与信道自适应能力。

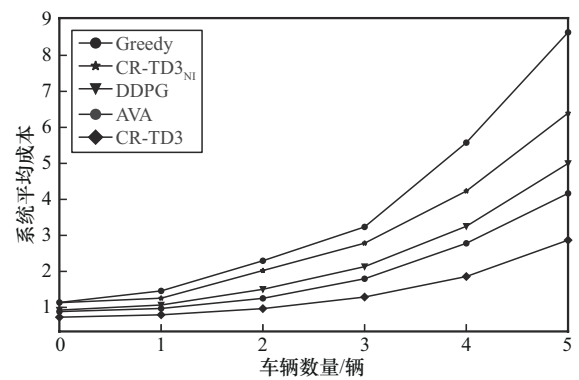


图5 不同车辆数量下各算法的系统平均成本

2) 系统平均时延分析

不同车辆数量下各算法的系统平均时延如图 6 所示。从图 6 可以看出,随着车辆数量的上升,系统平均时延普遍增加,但各算法表现存在显著差异。Greedy 只局限于短期成本,不能在资源紧张条件下灵活调整卸载策略,容易导致任务排队拥塞,从而时延始终居高不下。CR-TD3_{Ni} 虽然优化了调度框架,具备一定的资源调度能力,但由于缺乏 IRS 的辅助,其在复杂信道环境中传输稳定性较差,导致整体时延偏高。DDPG 和 AVA 在一定程度上通过策略学习和任务调度降低了排队等待时间,但在车辆数量激增时,优化效果仍有限。CR-TD3 算法借助 IRS 绕过信号遮挡区,实现高效稳定的通信链路,同时通过 TD3 策略动态感知环境状态,合理调控卸载与资源分配,有效缓解了排队拥堵和通信中断问题,表现出最低的系统平均时延,尤其在高负载情况下优势更加明显,充分验证了其在高动态、重负载环境下处理时延敏感型任务的能力。

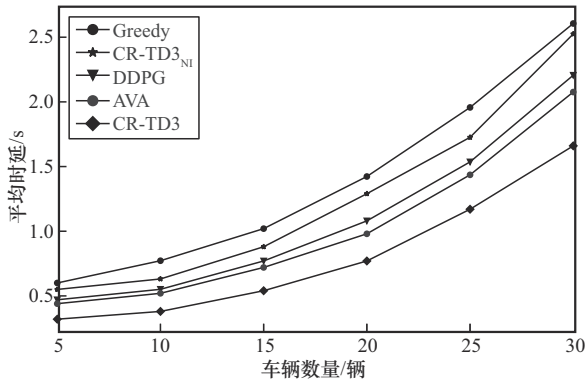


图6 不同车辆数量下各算法的系统平均时延

3) 系统平均能耗分析

能耗作为边缘计算中的关键指标，对系统能效具有重要影响。不同车辆数量下各算法的系统平均能耗如图7所示。从图7可以看出，Greedy因其短视，导致资源利用率较低，能耗水平最高。具体来看，当车辆数量达到30时，Greedy的能耗高达约0.15 J。CR-TD3_{NI}虽有优化框架，但缺乏IRS对信道质量的提升作用，在传输中需投入更大功率维持通信，导致其在高负载下能耗也达到了0.12 J。作为深度强化学习基线，DDPG与AVA通过资源调度在一定程度上降低了冗余能耗，其中性能较好的AVA在车辆数量为30时平均能耗为0.077 J。相比之下，CR-TD3算法在能耗方面表现最为优异。其通过IRS构建了低功耗、高增益的反射路径，显著减少了通信过程中的能量损耗。当车辆数量为30时，CR-TD3算法的平均能耗仅为0.06 J，相比Greedy降低了60%，相比AVA也降低了约22%。同时，与没有IRS辅助的CR-TD3_{NI}对比，能耗降低了50%，这充分证明了IRS技术在提升系统能效方面的巨大作用。最终，CR-TD3算法使系统在所有车辆数量下均保持最低能耗水平，显示出极佳的能耗控制能力与调度效率。

4) 系统平均任务完成率分析

任务完成率作为衡量系统服务质量的重要指标，其高低直接反映算法对资源的利用效率与系统稳定性。不同车辆数量下各算法的平均任务完成率如图8所示。从图8可以看出，随着车辆密度增加，负载加重，所有算法的任务完成率均呈下降趋势。Greedy的性能下降最为剧烈，在车辆数量达到30辆时，其任务完成率骤降至0.51，表明其在高负载场景中几乎无法保障服务质量。CR-

TD3_{NI}因缺少IRS构建的辅助链路，在通信暗区任务失败率较高，导致其在车辆数量为30辆时的完成率也仅有0.62。DDPG和AVA在中等负载下表现尚可，但在车辆数量超过20辆后，由于策略泛化能力不足，完成率也开始急剧下降，在车辆数量为30辆时分别为0.66和0.72。相较之下，CR-TD3算法在所有实验场景下均表现出最优异的性能和鲁棒性。在车辆数量从5辆增加到15的低到中负载场景下，CR-TD3算法始终保持着1.0的任务完成率，即便是在车辆数量为30辆的高负载极限场景下，依然能维持0.83的高成功率。这一性能不仅远超Greedy（高出0.32），也显著优于次优的AVA（高出0.11）。CR-TD3算法的优越性充分体现了其在IRS辅助的通信链路增强与基于DRL的资源自适应优化方面的双重优势，展现了卓越的系统稳定性和服务保障能力。

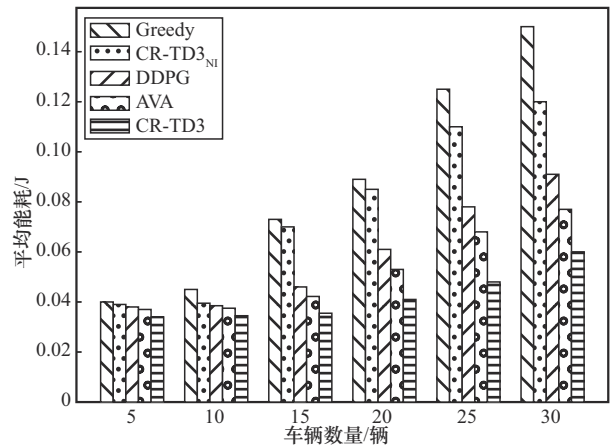


图7 不同车辆数量下各算法的系统平均能耗

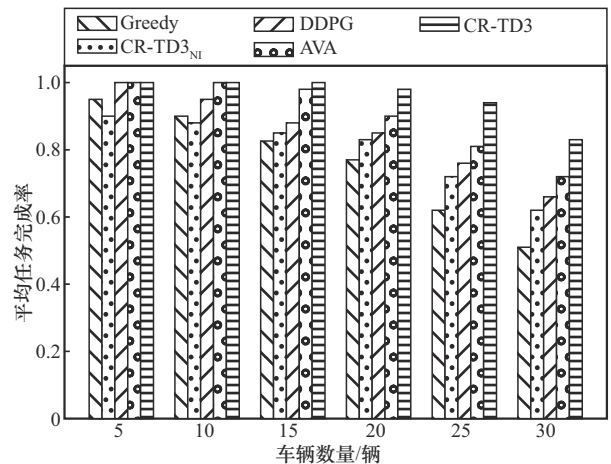


图8 不同车辆数量下各算法的平均任务完成率

3.5 敏感性分析

为进一步验证CR-TD3算法在极端场景下的鲁棒性,本节对关键环境参数进行敏感性分析。本文重点关注当计算任务负载发生突发性激增时系统的性能表现。本节试验保持车辆总数为20辆,将任务数据大小的生成范围分别设置为普通负载[0, 1.5] Mbit、高负载[1.5, 3] Mbit这2种情景,以模拟从平稳到任务激增的动态变化。

不同任务负载下各算法的平均任务完成率如图9所示。从图9可以看出,随着平均任务数据量从“普通负载”增加到“高负载”,所有算法的平均任务完成率均有所下降,这符合资源竞争加剧下的客观规律。然而,不同算法的性能下降幅度存在显著差异。Greedy由于缺乏对系统未来状态的预判,在任务负载激增时性能恶化最为严重,其平均任务完成率从普通负载下的0.77骤降至高负载下的0.58,在所有算法中表现最差。DDPG表现居中,平均任务完成率从0.86下降至0.74。作为一种深度强化学习算法,它能够学习并适应环境变化,因此表现优于前两者。然而,相比于TD3算法,DDPG在策略稳定性和价值估计准确性上存在不足,这使其在高负载的复杂决策场景中,性能下降较为明显。即使在“高负载”场景下,CR-TD3算法的平均任务完成率依然能保持在0.85。这一性能不仅显著优于次优的AVA(0.80),更是远超Greedy(0.58)。从性能下降的平稳性来看,CR-TD3算法的平均任务完成率仅从0.98下降至0.85,次优的AVA则从0.91下降至0.79,两者的稳定性相近,但CR-TD3算法在各种负载下均保持了最高的平均任务完成率。这充分表明CR-TD3算法具有相当的鲁棒性。

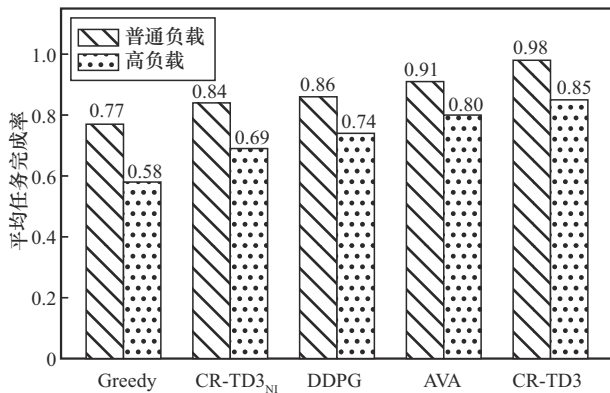


图9 不同任务负载下各算法的平均任务完成率

4 结束语

本文围绕车载边缘计算环境中信道遮挡和卸载失败率较高的问题,引入智能反射表面技术,构建了一个IRS辅助的车载边缘计算场景,并提出了一种基于深度强化学习的感知型TD3算法。首先,详细介绍了IRS辅助的VEC系统的通信模型、时延模型、能耗模型和任务完成率模型,全面刻画了任务在卸载过程中的各项关键性能指标。在此基础上,进一步抽象出以时延、能耗与任务完成率的综合性加权指标为优化目标的联合优化问题,并将其建模为马尔可夫决策过程,以适应后续的算法求解。随后,提出了CR-TD3算法,该算法结合了成功轨迹优先回放机制,可以对卸载决策、通信资源、计算资源以及IRS相位角度控制等多变量进行联合优化决策。最后,通过仿真实验把CR-TD3算法和多种基线算法进行了性能对比,从系统平均成本、时延、能耗以及任务完成率4个方面验证了CR-TD3算法的优越性。实验结果表明,CR-TD3算法充分发挥了智能反射表面和深度强化学习两者的优势,可以显著降低系统时延和能耗,同时提升任务完成率,在动态车载边缘计算环境中表现出良好的稳定性与扩展性。未来研究将探索集中式与分布式相结合的协同决策机制,通过引入联邦学习和多智能体强化学习方法,使车辆具备自主决策任务卸载的能力,并且有利于本地数据的隐私保护。

参考文献:

- [1] NOOR-A-RAHIM M, LIU Z L, LEE H, et al. 6G for vehicle-to-everything (V2X) communications: enabling technologies, challenges, and opportunities[J]. Proceedings of the IEEE, 2022, 110(6): 712-734.
- [2] LIU L, CHEN C, PEI Q Q, et al. Vehicular edge computing and networking: a survey[J]. Mobile Networks and Applications, 2021, 26(3): 1145-1168.
- [3] DANDALA T T, KRISHNAMURTHY V, ALWAN R. Internet of vehicles (IoV) for traffic management[C]//Proceedings of the 2017 International Conference on Computer, Communication and Signal Processing (ICCCSP). Piscataway: IEEE Press, 2017: 1-4.
- [4] LU S D, SHI W S. Vehicle as a mobile computing platform: opportunities and challenges[J]. IEEE Network, 2024, 38(6): 493-500.
- [5] MEHAR S, SENOUCI S M, KIES A, et al. An optimized roadside units (RSU) placement for delay-sensitive applications in vehicular networks[C]//Proceedings of the 2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC). Piscataway: IEEE Press, 2015: 121-127.
- [6] PLOTNIKOV P V, TAMBOVTSEV G I, VLADYKO A G. Performance evaluation of V2X model with a mobile road side units[C]//Proceedings

- of the 2023 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex (TIRVED). Piscataway: IEEE Press, 2023: 1-4.
- [7] SUN C, NI W, BU Z Y, et al. Energy minimization for intelligent reflecting surface-assisted mobile edge computing[J]. IEEE Transactions on Wireless Communications, 2022, 21(8): 6329-6344.
- [8] HAFEEZ T, XU L N, MCARDLE G. Edge intelligence for data handling and predictive maintenance in IIOT[J]. IEEE Access, 2021, 9: 49355-49371.
- [9] GHEISARI M, PHAM Q V, ALAZAB M, et al. ECA: an edge computing architecture for privacy-preserving in IoT-based smart city[J]. IEEE Access, 2019, 7: 155779-155786.
- [10] XIE J H, ZHOU X, CHENG L. Edge computing for real-time decision making in autonomous driving: review of challenges, solutions, and future trends[J]. International Journal of Advanced Computer Science and Applications, 2024, 15(7): 598-610.
- [11] WU Q Q, ZHANG R. Towards smart and reconfigurable environment: intelligent reflecting surface aided wireless network[J]. IEEE Communications Magazine, 2020, 58(1): 106-112. [
- [12] BASHARAT S, ALI HASSAN S, PERVAIZ H, et al. Reconfigurable intelligent surfaces: potentials, applications, and challenges for 6G wireless networks[J]. IEEE Wireless Communications, 2021, 28(6): 184-191.
- [13] 朱政宇, 王梓珏, 徐金雷, 等. 智能反射面辅助的未来无线通信: 现状与展望[J]. 航空学报, 2022, 43(2): 025014.
- ZHU Z Y, WANG Z X, XU J L, et al. Future wireless communication assisted by intelligent reflecting surface: State of art and prospects[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(2): 025014.
- [14] 朱政宇, 徐金雷, 孙钢灿, 等. 基于 IRS 辅助的 SWIPT 物联网系统安全波束成形设计[J]. 通信学报, 2021, 42(4): 185-193.
- ZHU Z Y, XU J L, SUN G C, et al. Secure beamforming design for IRS-assisted SWIPT Internet of Things system[J]. Journal on Communications, 2021, 42(4): 185-193.
- [15] YANG Y Z, GONG Y, WU Y C. Intelligent-reflecting-surface-aided mobile edge computing with binary offloading: energy minimization for IoT devices[J]. IEEE Internet of Things Journal, 2022, 9(15): 12973-12983.
- [16] BUZZI S, GROSSI E, LOPS M, et al. Foundations of MIMO radar detection aided by reconfigurable intelligent surfaces[J]. IEEE Transactions on Signal Processing, 2022, 70: 1749-1763.
- [17] QI K W, WU Q, FAN P Y, et al. Reconfigurable-intelligent-surface-aided vehicular edge computing: joint phase-shift optimization and multiuser power allocation[J]. IEEE Internet of Things Journal, 2025, 12(1): 764-777.
- [18] XIE Y B, SHI L, LI Z H, et al. Efficient task offloading in double road-side RIS-assisted vehicular edge computing networks using deep reinforcement learning[J]. IEEE Transactions on Vehicular Technology, 2025, 74(7): 11353-11365.
- [19] WAHID A, AYZED MIRZA M, AHMED M, et al. Toward secure and scalable vehicular edge computing with zero-energy RIS using DRL[J]. IEEE Access, 2024, 12: 129330-129346.
- [20] SALEEM O, BIN ASIF A, RIBOUH S, et al. RIS assisted cooperative computation offloading for autonomous vehicle in mobile edge computing[C]//Proceedings of the 2024 IEEE 100th Vehicular Technology Conference (VTC2024-Fall). Piscataway: IEEE Press, 2024: 1-6.
- [21] ALISHAHI M, FORTIER P, ZENG M, et al. Energy minimization for IRS-aided wireless powered federated learning networks with NOMA[J]. IEEE Internet of Things Journal, 2024, 11(9): 16339-16350.
- [22] SHARMA I, GUPTA S K. Channel tracking in IRS-based UAV communication systems using federated learning[J]. Journal of Electrical Engineering, 2023, 74(6): 521-531.
- [23] KONG L, HE J G, AI Y, et al. Channel modeling and analysis of reconfigurable intelligent surfaces assisted vehicular networks[C]//Proceedings of the 2021 IEEE International Conference on Communications Workshops (ICC Workshops). Piscataway: IEEE Press, 2021: 1-6.
- [24] LI A C, LIU Y, LI M, et al. Joint scheduling design in wireless powered MEC IoT networks aided by reconfigurable intelligent surface[C]//Proceedings of the 2021 IEEE/CIC International Conference on Communications in China (ICCC Workshops). Piscataway: IEEE Press, 2021: 159-164.
- [25] ZHENG G Y, XU C, LONG H, et al. MEC in NOMA-HetNets: a joint task offloading and resource allocation approach[C]//Proceedings of the 2021 IEEE Wireless Communications and Networking Conference (WCNC). Piscataway: IEEE Press, 2021: 1-6.
- [26] FARD MOSHIRI P, SIMSEK M, KANTARCI B. Joint optimization of completion ratio and latency of offloaded tasks with multiple priority levels in 5G edge[J]. IEEE Transactions on Network and Service Management, 2025, 22(2): 1357-1371.
- [27] ZHENG W, YAN L B. Latency minimization for IRS-assisted mobile edge computing networks[J]. Physical Communication, 2022, 53: 101768.
- [28] FUJIMOTO S, VAN HOOFF H, MEGER D. Addressing function approximation error in actor-critic methods[J]. arXiv Preprint, arXiv: 1802.09477, 2018.
- [29] WU Z P, JIA Z P, PANG X Y, et al. Deep reinforcement learning-based task offloading and load balancing for vehicular edge computing[J]. Electronics, 2024, 13(8): 1511.
- [30] WU H M, GU A Q, LIANG Y H. Federated reinforcement learning-empowered task offloading for large models in vehicular edge computing[J]. IEEE Transactions on Vehicular Technology, 2025, 74(2): 1979-1991.
- [31] LIANG Y H, TANG H J, WU H M, et al. Lyapunov-guided offloading optimization based on soft actor-critic for ISAC-aided Internet of vehicles[J]. IEEE Transactions on Mobile Computing, 2024, 23(12): 14708-14721.
- [32] ZHANG J, HU X P, NING Z L, et al. Energy-latency tradeoff for energy-aware offloading in mobile edge computing networks[J]. IEEE Internet of Things Journal, 2018, 5(4): 2633-2645.

[作者简介]



吴华明 (1986—), 男, 浙江余姚人, 博士, 天津大学教授、博士生导师, 主要研究方向为边缘计算、物联网、DNA 存储。



梁永辉 (1998—), 男, 山东菏泽人, 天津大学硕士生, 主要研究方向为边缘计算、深度学习。