

基于协作感知的通感算一体化车联网多维资源调度方案

杨新杰¹, 李睿哲¹, 梅杰¹, 马楠²

(1. 宁波大学信息科学与工程学院, 浙江 宁波 315211; 2. 北京邮电大学信息与通信工程学院, 北京 100876)

摘要: 为实现通感算一体化车联网 (integrated sensing, communication and computation based internet of vehicles, ISCC-IoV) 系统中通信、感知与计算性能的联合优化, 本文面向多路侧单元 (roadside unit, RSU) 覆盖场景下车辆感知请求驱动的端-边协作感知任务, 提出一种通、感、算联合资源分配方案。该方案定义了一种综合考虑感知数据量、任务时延与能耗的系统效用函数, 并联合优化了 RSU 选择及三类资源分配, 旨在最大化系统效用。形成的问题为非凸且 NP-hard 的混合整数非线性规划问题。为此, 提出了一种双 Critic 异构多智能体强化学习 (dual-Critic heterogeneous multi-agent reinforcement learning, DCHMARL) 算法。DCHMARL 将车辆和 RSU 作为异构智能体, 构建双价值 (Critic) 结构并联合设计分层策略 (Actor) 结构, 提高了训练的稳定性 and 效率。仿真结果表明, 所提资源分配方案在系统动态条件下能实现更优的通、感、算综合性能。此外, 与传统强化学习算法对比, 所提 DCHMARL 显著提高了系统效用, 降低了任务失败率, 且表现出更优的收敛性和稳定性。

关键词: 通感算一体化; 协作感知; 资源分配; 多智能体强化学习

中图分类号: TP393

文献标志码: A

Multidimensional Resource Scheduling for ISCC-IoV Based on Cooperative Sensing

Yang Xinjie¹, Li Ruizhe¹, Mei Jie¹, MA Nan²

1. Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China

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

Abstract: To achieve the joint optimization of communication, sensing, and computing performance in integrated sensing, communication, and computation-based Internet of Vehicles (ISCC-IoV) systems, this paper investigated a vehicle sensing request-driven end-edge cooperative sensing task in multi-roadside unit (RSU) coverage scenarios, and proposes a joint sensing, communication, and computation resource allocation scheme, where a system utility function was formulated to jointly consider sensing data size, task latency, and energy consumption. The joint optimization of RSU selection and multi-dimensional resource allocation was then modeled as a non-convex and NP-hard mixed-integer nonlinear programming problem. To solve this problem, a Dual-Critic Heterogeneous Multi-Agent Reinforcement Learning (DCHMARL) algorithm was proposed, where vehicles and RSU were modeled as heterogeneous agents. By combining a dual-critic mechanism with a hierarchical Actor structure, DCHMARL improves training stability and learning efficiency. Simulation results demonstrate that the proposed resource allocation scheme achieves superior overall communication, sensing, and computation performance under dynamic system conditions. Furthermore, compared with benchmark RL algorithms, DCHMARL achieves higher system utility, lower task failure rates, and faster, more stable convergence.

Key words: Integrated sensing, communication and computation, cooperative sensing, resource allocation, multi-agent

收稿日期: XXXX-XX-XX; 修回日期: XXXX-XX-XX

通信作者: 杨新杰, yangxinjie@nbu.edu.cn

基金项目: 国家自然科学基金委员会, 联合基金项目, U21A20448

Foundation Items: National Natural Science Foundation of China (NSFC) under grant U21A20448

0 引言

随着 5G 移动通信系统和人工智能技术的快速发展,车联网 (internet of vehicles, IoV) 进入高速发展阶段^[1]。IoV 面向道路安全和智能交通管理等应用,具有信息交互和环境感知需求。车辆可通过雷达等设备感知环境信息,并通过通信链路发送给基站、路侧单元 (roadside unit, RSU) 或周围车辆。同时,集成感知功能的多功能基站和 RSU 也可以感知更大范围环境信息并广播给车辆^[2]。在传统 IoV 系统中,感知与通信功能分离实现,硬件成本高昂,且雷达占用额外频谱,降低了系统频谱效率^[3]。因此,通信感知融合 (integrated sensing and communication, ISAC)^[4]技术在 IoV 系统的应用得到广泛关注与深入研究。然而,随着计算密集型 and 时延敏感型车载应用的普及,海量感知数据需要实时处理,对车载设备的计算能力构成了严峻挑战^[5]。因此,研究人员结合 ISAC 与移动边缘计算 (mobile edge computing, MEC)^[6]技术,构建了通感算一体化车联网系统 (integrated sensing, communication and computation based internet of vehicles, ISCC-IoV)^[7],以提升系统资源利用效率。

相比传统无线网络,ISCC-IoV 系统存在通信、感知与计算三类业务,它们的异构资源需求显著增加了系统资源分配与调度的复杂性,而传统资源调度方案往往关注单维资源使用,难以兼顾三者性能目标。因此,设计高效的资源分配方案是 ISCC-IoV 系统亟待解决的挑战和研究热点。在文献[8]中,Zhao 等人考虑了 ISCC-IoV 系统中计算任务的随机性,构建任务计算队列模型,通过联合调度通信资源与计算资源,以最大化任务完成率。Gu 等^[9]研究了 ISCC-IoV 系统中计算卸载与资源分配问题,提出联合优化卸载比例、发射功率和计算资源问题并利用联邦学习求解,实现了系统能耗最小化。文献[10]注意到多个车辆对同一目标的感知数据存在冗余,因此,提出减少冗余数据传输并联合优化了卸载决策和通信资源分配,显著减少了感知数据的总处理时延。然而该工作只是模拟了感知数据,并未对感知业务进行建模和评估。He 等^[11]研究了一个由配备了 MEC 服务器且具有感知功能的基站组成的 ISCC-IoV 系统,设计了一种计算和感

知资源分配方案,在满足任务需求与时延约束的同时最大化了感知性能,但其工作仅考虑将通信与计算性能作为约束条件。Cheng 等^[12]构建了一个集成了通信、感知与边缘计算多功能基站的 ISCC 系统,在满足雷达感知性能的约束下,通过联合优化带宽与计算资源实现了通信与计算性能联合优化。上述研究多关注单一过程或特定任务,缺乏通、感、算三类性能的统一评价与协同优化。因此,部分研究开始尝试联合优化通、感、算性能。例如,文献[13]提出一种面向 ISCC 的无线调度架构,以互信息、传输速率和时延分别评价感知、通信和计算性能,通过联合优化设备关联与子信道分配,实现对通、感、算性能的联合优化。然而,该研究基于单基站场景,未充分考虑多 RSU 或多基站时的动态关联与跨节点资源竞争问题。

上述 ISCC-IoV 研究主要针对单一感知主体 (如单车或基站),其感知范围有限且易受遮挡影响,而基于多节点协作的多源感知融合能够弥补上述局限,提升整体感知能力^{[14][15]}。文献[16]设计了一种车端融合方案,其中,边缘节点 RSU 将预处理后的感知数据传输至车辆,车辆执行数据融合后回传给 RSU,由 RSU 进行计算并广播结果。该方案中数据需多次传输且 RSU 感知数据较大,增加通信时延且导致频谱浪费。因此,文献[17]提出了边缘融合方案,其中,车辆将感知数据卸载至 RSU 进行融合、计算等后续操作,有效降低了传输负载和系统时延。

此外,ISCC-IoV 系统资源分配需要优化多维决策变量,复杂度高,传统凸优化或启发式算法难以实现高效在线决策。而多智能体强化学习 (multi-agent reinforcement learning, MARL) 在动态环境下具有强大的自适应决策能力,近年来逐渐被用来解决类似问题。例如,文献[18]为 ISCC-IoV 系统提出一种基于 MARL 的 RSU 间多跳路由选择方案,以优化卸载决策,减少端到端时延。文献[19]针对 IoV 中的任务迁移与卸载问题,将深度确定性策略梯度 (deep deterministic policy gradient, DDPG) 与决斗双深度 Q 网络 (dueling double deep Q-network, D3QN) 相结合,提出新的 QDPG 算法,进行连续动作和离散动作的混合决策。文献

[20]等为IoV系统提出了一种多智能体近端策略优化 (multi-agent proximal policy optimization, MAPPO) 算法, 将不同决策节点建模为异构智能体, 但该算法基于统一网络结构与策略优化框架, 未针对异构智能体进行差异化设计。

综上, 现有关于 ISCC-IoV 多维资源分配、多节点感知融合以及 MARL 的研究虽然取得了一定进展, 但大多集中于优化特定任务和单一性能, 普遍缺乏对通、感、算资源的协同优化。同时, 多数研究采用单一 RSU 或基站作为计算与感知融合中心节点, 限制了节点选择的灵活性, 难以适应车辆位置动态变化和任务异构需求。此外, 大多基于 MARL 的研究将离散与连续动作在单个智能体内部进行联合建模, 或在异构智能体场景下采用统一的策略优化框架。针对上述研究缺陷, 本文面向多车辆、多 RSU 构成的 ISCC-IoV 系统, 提出了一种基于车辆与 RSU 协作感知的通、感、算多维资源分配方案, 主要贡献如下:

(1) 面向多 RSU 车联网端一边协作感知场景, 构建 RSU 选择与通、感、算资源联合优化问题。车辆根据环境状态动态选择 RSU 进行协作感知与计算卸载, 并由 RSU 完成计算处理。在此基础上, 本文考虑基于感知数据量的感知性能, 基于总时延的通信性能, 以及基于总能耗的计算性能, 构建了一种兼顾多维性能的系统效用模型, 并联合优化 RSU 选择与多维资源分配, 旨在最大化系统效用。

(2) 针对上述优化问题中车辆与 RSU 角色异构以及离散变量与连续变量并存的特点, 提出了一种基于 MARL 的双价值网络异构多智能体强化学习 (dual-Critic heterogeneous MARL, DCHMARL) 算法。DCHMARL 将车辆与 RSU 建模为异构智能体, 构建了一种双价值 (Critic) 结构并联合设计分层策略 (Actor) 结构。此外, 异构智能体采用差异化策略训练方法, 支持在统一框架下协同优化离散 RSU 选择与连续资源分配, 显著提升了训练的稳定性及整体性能。

本文剩余部分组织如下, 第 1 节对系统模型和问题形成进行了描述, 第 2 节提出了 DCHMARL 算法, 第 3 节提供了仿真结果和分析。最后, 第 4 节对本文工作进行了总结。

1 系统模型

本文考虑一个位于城市中心的多车辆、多 RSU 构成的 ISCC-IoV 系统, 如图 1 所示。系统包含一段单向多车道街道, M 个配备 MEC 服务器的 RSU 沿道路均匀分布, 其集合记为 $\mathfrak{R} = \{r_1, \dots, r_m, \dots, r_M\}$, 若干辆车沿道路中心行驶。车辆和 RSU 均支持 ISAC 技术, 并假设感知功能复用上行链路的通信频谱。MEC 服务器拥有计算资源, 能够为车辆提供计算卸载服务, 并可以对感知数据进行计算处理。

本文为以上系统提出了一种端-边协作感知与资源分配方案。假设系统将时间划分为长度固定、足够小的优化时隙 Δt , 每个时隙分为长度固定的感知时隙与通信-计算时隙^[8], 车辆和 RSU 以时分工作方式发送感知和通信信号, 且车辆通过周期性广播共享自身位置。在感知时隙, 车辆根据泊松分布随机产生至多一个环境感知请求。假设在一个给定时隙, 共有 N 辆车生成了感知请求, 记为 $\mathcal{V} = \{v_1, \dots, v_n, \dots, v_N\}$ 。随即, 车辆根据广播信息构建其感知范围 (Range of Sensing, ROS) 内目标车辆的集合, 并使用分配的感知带宽进行多目标感知, 通过接收回波获得车辆感知数据。此外, RSU 使用预留的感知带宽对其 ROS 进行感知, 得到 RSU 感知数据。在通信-计算时隙, 车辆使用分配的上行带宽将车辆感知数据卸载至服务 RSU, 形成 v_n 的感知任务 $task_n = \{T_n^{\max}, S_n\}$, 其中, $T_n^{\max}$ 为感知任务最大容忍时延, S_n 为感知任务的数据量。最后, RSU 为车辆的感知任务分配计算资源, 经过计算处理后将结果发回给车辆。需要注意的是, 若上述过程在最大容忍时延内或在通信-计算时隙结束前未能完成, 则该任务将因超时被系统予以丢弃处理。

本文用二元变量 $\mu_{n,m} \in \{0, 1\}$ 表示车辆与 RSU 间的选择关系, 若车辆 v_n 选择 r_m 作为其服务 RSU, 则 $\mu_{n,m} = 1$, 反之, $\mu_{n,m} = 0$ 。由于每个车辆只能选择一个服务 RSU, 选择关系需满足以下约束:

$$C1: \sum_{m=1}^M \mu_{n,m} = 1, \forall n \quad (1)$$

假设研究的 ISCC-IoV 系统基于 3GPP 标准 NR-V2X (new radio-vehicle to everything), 采用正交频分多址接入 (orthogonal frequency-division multiple access, OFDMA) 技术, 每个车辆使用分配

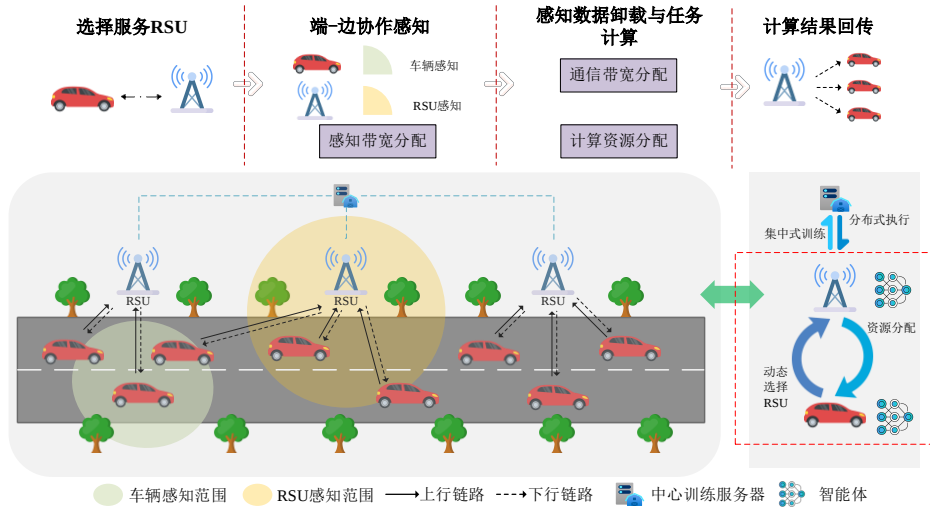


图1 面向端-边协作感知的ISCC-IoV系统模型

的正交信道卸载车辆感知数据。将系统上行总带宽记为 B^{ul} ，则分配给车辆 v_n 的上行带宽 $B_{n,m}^{ul}$ 需满足以下约束：

$$C2: \sum_{m=1}^M \sum_{n=1}^N B_{n,m}^{ul} \leq B^{ul} \quad (2)$$

本文假设感知功能复用上行通信频谱，其中，系统为每个RSU预留固定的正交带宽 B_{rsu}^{sense} ，在剩余带宽内，系统为车辆分配正交的感知带宽 $B_{n,m}^{sense}$ 。因此，感知带宽分配需要满足以下约束：

$$C3: \sum_{m=1}^M \sum_{n=1}^N B_{n,m}^{sense} \leq B^{ul} - M \cdot B_{rsu}^{sense} \quad (3)$$

在RSU处，MEC服务器分配计算资源 $f_{n,m}$ 时需要满足以下约束：

$$C4: \sum_{n=1}^N f_{n,m} \leq F, \forall m \quad (4)$$

需要说明的是，当 $\mu_{n,m} = 0$ 时， r_m 无需为车辆 v_n 分配资源，即 $B_{n,m}^{sense} = B_{n,m}^{ul} = f_{n,m} = 0$ 。

1.1 感知模型

本文假设车辆的ROS为半径为 r 的圆形区域，每辆车配备一个由 L 根收发天线单元构成的均匀线性阵列 (uniform linear array, ULA)。记 $\mathcal{I}_n^v = \{1, \dots, i, \dots, I_n^v\}$ 为车辆 v_n ROS 内的目标车辆集合，在感知时隙， v_n 将集合内车辆视作感知目标并发送雷达探测信号。

根据3GPP标准 TR 38.901 建议的信道模型^[21]，可以将目标车辆 i 与车辆 v_n 之间的雷达回波等效信道定义为：

$$g_{n,i} = \sqrt{\alpha_{RCS} \cdot 10^{\frac{2PL(d_{n,i}) + \zeta_{n,i}^r}{10}}} \tilde{g}_{n,i} \quad (5)$$

其中， $PL(d_{n,i}) = 128.1 + 37.6 \cdot \log_{10}(d_{n,i})$ 为路径损耗， α_{RCS} 为目标的雷达散射截面， $d_{n,i}$ 为两车间距离， $\zeta_{n,i}^r \sim \mathcal{N}(0, \sigma_r^2)$ 为雷达链路阴影衰落， σ_r 为该阴影衰落的标准差， $\tilde{g}_{n,i} \sim \text{CN}(0, 1)$ 为小尺度快衰落。

对于车辆 v_n ，目标车辆 i 可视作角度为 θ_i 的点目标。因此，车辆 v_n 在 θ_i 方向上的接收信干噪比 (signal to noise and interference ratio, SINR) 可表示为：

$$\text{SINR}_{n,i} = \frac{P_{n,i}^{\text{rad}} |g_{n,i}|^2}{n_0 B_{n,m}^{\text{sense}} + \sum_{j \in \mathcal{I}_n^v, j \neq i} P_{n,j}^{\text{rad}} |g_{n,j}|^2} \quad (6)$$

其中， $P_{n,i}^{\text{rad}}$ 为车辆 v_n 在 θ_i 方向上的探测信号功率， n_0 为加性白高斯噪声 (additive white gaussian noise, AWGN) 功率谱密度。

车辆 v_n 在 θ_i 方向上雷达估计信息速率^[22]为：

$$R_{n,i}^{\text{rad}} = \frac{\varepsilon}{2} \log_2 \left(1 + 2\tau_{\text{sense}} (B_{n,m}^{\text{sense}})^3 \sigma_{\text{proc}}^2 \zeta^2 \text{SINR}_{n,i} \right) \quad (7)$$

其中， σ_{proc}^2 为均值为0、服从正态分布的过程噪声的方差， ζ 为一个取决于探测信号的功率谱密度的常数， ε 为雷达占空比， τ_{sense} 为感知时隙长度。

车辆 v_n 感知 $\mathcal{I}_n^v$ 辆车得到车辆感知数据量为：

$$S_n^{\text{veh}} = \sum_{i=1}^{I_n^v} \tau_{\text{sense}} R_{n,i}^{\text{rad}} \quad (8)$$

本文假设RSU配备一个由 K 个收发天线单元

构成的 ULA, 记 $\mathcal{I}_m^r = \{1, \dots, i, \dots, I_m^r\}$ 为 RSU r_m 的目标车辆集合, r_m 在感知时隙向目标车辆发送探测信号。

类似的, 目标车辆 j 可视作角度为 φ_j 的点目标。因此, RSU r_m 在 φ_j 方向上的雷达估计信息速率为:

$$R_{m,j}^{\text{rad}} = \frac{\varepsilon}{2} \log_2 \left(1 + \frac{2\tau_{\text{sense}} (B_{\text{rsu}}^{\text{sense}})^3 \sigma_{\text{proc}}^2 \zeta^2 P_{m,j}^{\text{rad}} |g_{m,j}|^2}{n_0 B_{\text{rsu}}^{\text{sense}} + \sum_{k \in \mathcal{I}_m^r, k \neq j} P_{m,k}^{\text{rad}} |g_{m,k}|^2} \right) \quad (9)$$

其中, $P_{m,j}^{\text{rad}}$ 为 r_m 在 φ_j 方向上的探测信号功率, $g_{m,j}$ 为 r_m 与目标车辆 j 间的雷达回波等效信道, 其定义与公式 (7) 类似, 此处不再赘述。

RSU r_m 感知 I_m^r 辆车得到 RSU 感知数据量为:

$$S_m^{\text{rsu}} = \sum_{j=1}^{I_m^r} \tau_{\text{sense}} R_{m,j}^{\text{rad}} \quad (10)$$

因此, 当 $\mu_{n,m} = 1$ 时, 形成车辆 v_n 的感知任务数据量为:

$$S_n = S_m^{\text{rsu}} + S_n^{\text{veh}} \quad (11)$$

1.2 通信模型

在通信-计算时隙, 当时, 车辆 v_n 向 RSU r_m 通过上行链路卸载感知数据。本文采用 3GPP TR 38.901 建议的信道模型, 将车辆 v_n 到 RSU r_m 的上行链路信道定义为:

$$h_{n,m} = \sqrt{10^{-\frac{\text{PL}(d_{n,m}) + \zeta_{n,m}^c}{10}}} \tilde{h}_{n,m} \quad (12)$$

其中, $d_{n,m}$ 为车辆 v_n 与 RSU r_m 间的距离, $\zeta_{n,m}^c \sim \mathcal{N}(0, \sigma_c^2)$ 为通信信道的阴影衰落, σ_c 为该阴影衰落的标准方差, $\tilde{h}_{n,m} \sim \text{CN}(0, 1)$ 为小尺度快衰落。

由于 OFDMA 信道正交性, RSU r_m 的接收信噪比 (signal to noise ratio, SNR) 为:

$$\gamma_{n,m} = \frac{P_n^{\text{veh}} |h_{n,m}|^2}{n_0 B_{n,m}^{\text{ul}}} \quad (13)$$

其中, P_n^{veh} 为 v_n 的发射功率。因此, 车辆 v_n 与 RSU r_m 间的上行链路传输速率为:

$$R_{n,m}^{\text{ul}} = B_{n,m}^{\text{ul}} \log_2(1 + \gamma_{n,m}) \quad (14)$$

1.3 计算模型

车辆 v_n 通过上行链路将车辆感知数据卸载到的 RSU r_m 的传输时延为:

$$\tau_n^{\text{ul}} = \frac{S_n^{\text{veh}}}{R_{n,m}^{\text{ul}}} \quad (15)$$

车辆 v_n 卸载需要的传输能耗为:

$$E_{n,m}^{\text{ul}} = \frac{S_n^{\text{veh}} P_n^{\text{veh}}}{R_{n,m}^{\text{ul}}} \quad (16)$$

本文假设每个时隙内感知任务相互独立且无排队, RSU 为每个任务分配计算资源以进行多任务并行处理。因此, RSU r_m 处理车辆 v_n 的感知任务的计算时延为:

$$\tau_n^{\text{comp}} = \frac{S_n c_m}{f_{n,m}} \quad (17)$$

其中, c_m 为 RSU r_m 的计算能力, 即处理单位 bit 数据所需要的 CPU 周期。

RSU r_m 处理车辆 v_n 的感知任务的计算能耗为:

$$E_{n,m}^{\text{comp}} = k c_m S_n (f_{n,m})^2 \quad (18)$$

其中, k 为有效开关电容, 与使用的芯片有关。

由于 RSU 完成车辆感知任务计算后返回给车辆的结果数据量较小, 其下行传输的时延和能耗可以忽略不计。

综上, 车辆 v_n 感知任务处理需要的总时延为:

$$\tau_n = \tau_n^{\text{ul}} + \tau_n^{\text{comp}} \quad (19)$$

车辆 v_n 感知任务处理需要的总能耗为:

$$E_n = E_{n,m}^{\text{ul}} + E_{n,m}^{\text{comp}} \quad (20)$$

1.4 问题形成

本文提出基于端-边协作感知的 ISCC-IoV 系统架构, 并重点研究其中的多维资源分配方案, 旨在提升协作感知效能的同时, 有效平衡通信与计算开销, 实现通、感、算多维性能协同优化。因此, 本文以 τ_n 、 S_n 和 E_n 分别作为通信、感知与计算的绩效指标, 并定义综合考虑三者性能的系统效用函数。

车辆 v_n 的效用函数定义如下:

$$U_n = \frac{\alpha S_n}{\gamma \tau_n + \beta E_n} \quad (21)$$

其中, $\alpha > 0$ 、 $\gamma > 0$ 和 $\beta > 0$ 分别为感知任务数据量、总时延、总能耗的比例系数, 将三者转化为无量纲的标量并约束在同一数量级。

进一步, ISCC-IoV 系统的效用函数定义如下:

$$U = \sum_{n=1}^N U_n \quad (22)$$

所提多维资源分配方案的优化目标是在满足各

类约束的前提下,通过联合优化RSU选择与多维资源分配,最大化系统效用。定义 $\mathbf{B}^{\text{ul}} = [B_{n,m}^{\text{ul}}]_{N \times M}$ 为车辆上行带宽分配矩阵, $\mathbf{B}^{\text{sense}} = [B_{n,m}^{\text{sense}}]_{N \times M}$ 为车辆感知带宽分配矩阵, $\mathbf{f} = [f_{n,m}]_{N \times M}$ 为车辆计算资源分配矩阵, $\boldsymbol{\mu} = [\mu_{n,m}]_{N \times M}$ 为车辆与RSU选择矩阵。因此,形成的优化问题可以描述为:

$$\begin{aligned} P1: & \max_{\boldsymbol{\mu}, \mathbf{B}^{\text{ul}}, \mathbf{B}^{\text{sense}}, \mathbf{f}} U \\ \text{s.t. } & C1 - C4 \\ & C5: \mu_{n,m} \in \{0, 1\} \\ & C6: \tau_n \leq T_n^{\text{max}}, \forall n \end{aligned} \quad (23)$$

其中, C1和C5表示车辆只能选择一个服务RSU。C2-C4是分配给车辆的总带宽和总计算资源的限制, C6则表示完成车辆感知任务的总时延不得超过最大时延约束。

2 基于DCHMARL的选择与资源分配方案

形成的优化问题P1是混合整数非线性规划(mixed-integer nonlinear programming, MINLP)且NP-hard,难以通过传统算法求解。因此,本文提出了一种基于MARL的双Critic异构多智能体强化学习算法, DCHMARL。基于ISCC-IoV系统中车辆与RSU的不同职能, DCHMARL将车辆与RSU定义为异构的车辆智能体与RSU智能体,通过异构智能体间的协同决策,实现RSU选择与多维资源分配。

2.1 多智能体环境建模

(1) 状态空间

在所提DCHMARL算法中,将深度强化学习中的回合(episode)定义为系统模型中的时隙,车辆 v_n 在第 t 步(step)的状态定义为:

$$s_{v_n}(t) = \{\tilde{n}, p_n(t), T_n^{\text{max}}, L_n(t-1), D_n(t)\} \quad (24)$$

包含以下内容:

$$\tilde{n} = \frac{n}{N}$$

$$L(t-1) = \{l_1(t-1), l_2(t-1), \dots, l_M(t-1)\} \quad (25)$$

$$D_n(t) = \{d_{n,1}(t), d_{n,2}(t), \dots, d_{n,M}(t)\}$$

其中, $\tilde{n}$ 为车辆 v_n 的归一化身份特征,用于区分不同车辆智能体; $p_n(t)$ 表示车辆 v_n 第 t 步的位置; $L(t-1)$ 代表 M 个RSU广播第 $t-1$ 步的负载信息,

具体地, $l_m(t-1)$ 定义为第 $t-1$ 步选择RSU r_m 的车辆个数; $D_n(t)$ 表示第 t 步车辆 v_n 与 M 个RSU之间的距离,比如, $d_{n,m}(t)$ 表示第 t 步车辆 v_n 与RSU r_m 的间距。

类似地, RSU r_m 第 t 步的状态定义为:

$$S_{r_m}(t) = \{\tilde{m}, P_m(t), \mathcal{T}_m(t)\} \quad (26)$$

包含以下内容:

$$\tilde{m} = \frac{m}{M}$$

$$P_m(t) = \{p_1(t), p_2(t), \dots, p_{L_m}(t)\} \quad (27)$$

$$\mathcal{T}_m(t) = \{T_1^{\text{max}}, T_2^{\text{max}}, \dots, T_{L_m}^{\text{max}}\}$$

其中, $\tilde{m}$ 为RSU r_m 的归一化身份特征,用于区分不同RSU智能体, $L_m(t) = \sum_{n=1}^N \mu_{n,m}(t)$ 表示第 t 步选择RSU r_m 的车辆个数, $P_m(t)$ 表示第 t 步选择RSU r_m 的车辆的位置集合, $\mathcal{T}_m(t)$ 表示这些车辆的任务最大容忍时延集合。

(2) 动作空间

车辆动作反映了车辆智能体的选择决策,因此,所提DCHMARL算法将车辆 v_n 第 t 步的动作定义为:

$$a_{v_n}(t) = \{\mu_{n,1}(t), \mu_{n,2}(t), \dots, \mu_{n,M}(t)\} \quad (28)$$

RSU r_m 第 t 步动作定义为:

$$a_{r_m}(t) = \{B_m^{\text{sense}}(t), B_m^{\text{ul}}(t), F_m(t)\} \quad (29)$$

包含以下内容:

$$B_m^{\text{sense}}(t) = \{B_{1,m}^{\text{sense}}(t), B_{2,m}^{\text{sense}}(t), \dots, B_{N,m}^{\text{sense}}(t)\}$$

$$B_m^{\text{ul}}(t) = \{B_{1,m}^{\text{ul}}(t), B_{2,m}^{\text{ul}}(t), \dots, B_{N,m}^{\text{ul}}(t)\} \quad (30)$$

$$F_m(t) = \{f_{1,m}(t), f_{2,m}(t), \dots, f_{N,m}(t)\}$$

其中, $B_m^{\text{sense}}(t)$ 为RSU r_m 第 t 步感知带宽分配决策, $B_m^{\text{ul}}(t)$ 为上行带宽分配决策, $F_m(t)$ 为计算资源分配决策。

(3) 奖励函数

为引导车辆在满足时延约束的前提下最大化系统效用, 本文将车辆的奖励函数设计为与效用函数一致的形式。具体来说, 若车辆能在最大容忍时延内完成任务, 其奖励定义为自身效用; 若任务超出该时延阈值, 则基于超时时长给以惩罚。因此, 车辆 v_n 的奖励函数可定义为:

$$r_{v_n}(t) = \begin{cases} \frac{\alpha S_n(t)}{\gamma \tau_n(t) + \beta E_n(t)}, & \text{if } \tau_n(t) \leq T_n^{\max} \\ \rho(\tau_n(t) - T_n^{\max}), & \text{if } \tau_n(t) > T_n^{\max} \end{cases} \quad (31)$$

其中, $\rho < 0$ 为超时惩罚系数。

RSU 旨在最大化被服务车辆的整体效用, 同时引入公平性惩罚项以避免资源向单一车辆过度倾斜。因此, RSU r_m 的奖励函数可表示为:

$$r_{r_m}(t) = \sum_{n=1}^N \mu_{n,m}(t) U_n(t) - W_m^p(t) \quad (32)$$

其中, $W_m^p(t) = \omega \frac{1}{L_m(t)} \sum_{n=1}^N \left(\mu_{n,m}(t) U_n(t) - \frac{1}{L_m(t)} \sum_{i=1}^N U_i(t) \right)^2$ 为公平性惩罚项, $\omega > 0$ 为公平性惩罚系数。

可以看出, DCHMARL 根据车辆效用分别设计了车辆和 RSU 的奖励函数以引导多智能体实现最大化奖励, 与 P1 优化目标高度契合。

为简化表达, 用 $(s_{v_n}, a_{v_n}, r_{v_n})$ 、 $(s'_{v_n}, a'_{v_n}, r'_{v_n})$ 、 $(s_{r_m}, a_{r_m}, r_{r_m})$ 和 $(s'_{r_m}, a'_{r_m}, r'_{r_m})$ 分别表示车辆 v_n 和 RSU r_m 在第 t 步以及第 $t+1$ 步的状态、动作和奖励。此外, 用 $s_v = \{s_{v_n}\}$, $a_v = \{a_{v_n}\}$, $r_v = \{r_{v_n}\}$ 和 $s_r = \{s_{r_m}\}$, $a_r = \{a_{r_m}\}$, $r_r = \{r_{r_m}\}$ 分别表示车辆和 RSU 在第 t 步的联合状态、动作和奖励, 用 $s'_v = \{s'_{v_n}\}$, $a'_v = \{a'_{v_n}\}$, $r'_v = \{r'_{v_n}\}$ 和 $s'_r = \{s'_{r_m}\}$, $a'_r = \{a'_{r_m}\}$, $r'_r = \{r'_{r_m}\}$ 分别表示车辆和 RSU 在第 $t+1$ 步的联合状态、动作和奖励。

2.2 DCHMARL 算法框架与训练

在 DCHMARL 中, 车辆智能体的决策将直接影响各 RSU 的负载状态以及车辆间的资源竞争, 具有全局影响性, 而 RSU 智能体只需在车辆选择基础上独立进行资源分配。因此, 与采用单一 Critic 或统一策略优化框架的 MARL 方法不同, DCHMARL 为车辆与 RSU 智能体分别配备了全局 Critic 网络和本地 Critic 网络, 形成基于双 Critic 的差异化评估结构。两类 Critic 具有不同的输入信息和评价目标: 全局 Critic 以所有智能体的状态和动作为输入, 评估车辆 RSU 选择对系统整体性能的影响, 并引导多智能体协同优化; 本地 Critic 仅以单个 RSU 智能体的局部状态和动作为输入, 评估其资源分配决策产生的局部奖励, 有助于稳定 RSU

智能体策略更新并降低联合状态、动作空间带来的训练复杂度, 提升整体训练稳定性。

此外, 与由单个智能体统一生成离散—连续混合动作, 或对不同智能体采用相同策略优化方式的方法不同, DCHMARL 根据车辆与 RSU 的功能差异对混合决策过程进行解耦, 形成一种“离散选择—连续控制”的分层 Actor 结构。具体而言, 车辆智能体进行离散决策以确定服务 RSU, RSU 智能体执行连续资源分配, 同时根据异构智能体网络结构不同采用差异化策略学习方法, 在降低训练复杂度的同时提高训练稳定性。

进一步, 全局 Critic 网络由在线 Critic 网络与目标 Critic 网络组成, 其网络参数分别记为 ω_n 和 ω_n^- 。本地 Critic 网络由在线 Critic 网络与目标 Critic 网络组成, 其参数分别记为 ϕ_m 和 ϕ_m^- 。两类在线 Critic 网络的输出 Q_{ω_n} 和 Q_{ϕ_m} 分别定义为:

$$Q_{\omega_n}(s_v, a_v, s_r, a_r) = E \left[\sum_{t=0}^{\infty} \gamma^t \cdot r_{v_n} \right] \quad (33)$$

$$Q_{\phi_m}(s_{r_m}, a_{r_m}) = E \left[\sum_{t=0}^{\infty} \gamma^t \cdot r_{r_m} \right] \quad (34)$$

其中 $\gamma \in [0, 1)$ 为折扣因子, 用于平衡当前奖励与未来奖励之间的权重。

车辆智能体 v_n 具有在线 Actor 网络 π_{v_n} 和目标 Actor 网络 $\pi_{v_n}^-$, 其参数分别为 θ_{v_n} 和 $\theta_{v_n}^-$; RSU 智能体 r_m 具有在线 Actor 网络 π_{r_m} 和目标 Actor 网络 $\pi_{r_m}^-$, 其参数分别为 θ_{r_m} 和 $\theta_{r_m}^-$ 。车辆智能体 v_n 和 RSU 智能体 r_m 的在线 Actor 网络的输出分别为:

$$a_{v_n} = \pi_{v_n}(s_{v_n}; \theta_{v_n}), a_{r_m} = \pi_{r_m}(s_{r_m}; \theta_{r_m}) \quad (35)$$

分别利用 $J_{\theta_{v_n}}$ 和 $J_{\theta_{r_m}}$ 来评价, 其定义如下:

$$J_{\theta_{v_n}} = E \left[Q_{\omega_n}(s_v, a_v, s_r, a_r) \right] \quad (36)$$

$$J_{\theta_{r_m}} = E \left[Q_{\phi_m}(s_{r_m}, a_{r_m}) \right] \quad (37)$$

网络训练首先更新车辆智能体的全局 Critic 网络与 Actor 网络, 随后更新 RSU 智能体的本地 Critic 网络与 Actor 网络。接下来将分别介绍车辆与 RSU 智能体的训练流程。

(1) 车辆智能体训练

车辆智能体采用双时延深度确定性策略梯度 (twin delayed deep deterministic policy gradient,

TD3) 更新其网络参数。

值得注意的是, 车辆智能体的 RSU 选择是离散动作, 但 TD3 的 Actor 更新需要动作对网络参数可导。因此, 本文采用 Gumbel-Softmax 方法对车辆智能体动作进行处理, 使车辆在前向决策时输出离散独热 (one hot) 动作, 在反向传播时通过连续松弛变量传递梯度。具体而言, 车辆智能体 v_n 的 Actor 网络先输出一个偏好向量 $\mathbf{x}_n = [x_{n,1}, \dots, x_{n,m}, \dots, x_{n,M}]$, 其中, x_m 为车辆 v_n 选择 RSU r_m 的概率, 且满足 $\sum_{m=1}^M x_{n,m} = 1$ 。随后, 对车辆 v_n 选择 RSU r_m 的动作引入 Gumbel 噪声 $g_{n,m}$, 表示如下:

$$g_{n,m} = -\log(-\log u_{n,m}), u_{n,m} \sim U(0,1) \quad (38)$$

通过 Gumbel-Softmax 方法得到连续松弛后的偏好向量为 $\mathbf{y}_n = [y_{n,1}, \dots, y_{n,m}, \dots, y_{n,M}]$, 其中, $y_{n,m}$ 可表示为:

$$y_{n,m} = \frac{\exp((\log x_{n,m} + g_{n,m})/\zeta)}{\sum_{j=1}^M \exp((\log x_{n,j} + g_{n,j})/\zeta)} \quad (39)$$

其中, ζ 为 Gumbel-Softmax 温度系数。

此时, 连续松弛后的偏好向量 $\mathbf{y}_n$ 可对 $\mathbf{x}_n$ 求导, 允许梯度进行反向传播。而前向决策时, 车辆智能体 v_n 输出 one hot 动作 a_{v_n} , 通过如下方式计算:

$$\mu_{n,m}(t) = \begin{cases} 1, & \text{if } m = \arg \max_m (y_{n,m}) \\ 0, & \text{else} \end{cases} \quad (40)$$

基于 TD3 思想, DCHMARL 算法中车辆智能体 v_n 的全局 Critic 网络被设计成由两个独立的 Critic 网络共同组成, 即双 Q 网络, 参数分别为 $\omega_{n,1}$ 和 $\omega_{n,2}$, 输出分别为 $Q_{\omega_{n,1}}$ 和 $Q_{\omega_{n,2}}$ 。每次训练时, 中央控制器从经验回放缓冲区 D 中随机抽取一批数据计算全局在线 Critic 网络的输出, 并取其目标网络的输出 $\bar{Q}_{\omega_{n,1}}$ 和 $\bar{Q}_{\omega_{n,2}}$ 中较小的一个用于计算时序差分 (temporal difference, TD) 目标, 如下:

$$y_n^g = r_{v_n} + \gamma \min_{k=1,2} Q_{\omega_{n,k}}(s'_v, a'_v, s'_r, a'_r) \quad (41)$$

随后, 全局 Critic 网络的两个独立 Critic 网络分别计算各自的损失, 其损失函数定义为:

$$L(\omega_{n,k}) = \mathbf{E} \left[\left(Q_{\omega_{n,k}}(s_v, s_r, a_v, a_r) - y_n^g \right)^2 \right], k = 1, 2 \quad (42)$$

最后, 全局 Critic 网络的参数通过最小化损失函数做如下更新:

$$\omega_{n,k} \leftarrow \omega_{n,k} - v \nabla_{\omega_{n,k}} L(\omega_{n,k}), k = 1, 2 \quad (43)$$

其中, $v > 0$ 为 Critic 网络的学习率。

基于 Gumbel-Softmax 方法, 车辆智能体 v_n 的策略梯度计算如下:

$$\nabla_{\theta_{v_n}} J_{v_n} = \mathbf{E} \left[\nabla_{\theta_{v_n}} \pi_{v_n}(s_{v_n}; \theta_{v_n}) \nabla_{a_{v_n}} Q_{\omega_{n,1}}(s_v, a_v, s_r, a_r) \right] \quad (44)$$

Actor 网络 and 所有目标网络采用延迟更新策略, 即每更新 d 次 Critic 网络, 更新 Actor 网络一次, 并软更新目标网络一次:

$$\theta_{v_n} \leftarrow \theta_{v_n} + \mu \nabla_{\theta_{v_n}} J_{v_n} \quad (45)$$

$$\bar{\omega}_{n,k} \leftarrow \tau \omega_{n,k} + (1 - \tau) \bar{\omega}_{n,k}, k = 1, 2 \quad (46)$$

$$\bar{\theta}_{v_n} \leftarrow \tau \theta_{v_n} + (1 - \tau) \bar{\theta}_{v_n} \quad (47)$$

其中, $\mu > 0$ 为 Actor 网络的学习率, $\tau \in (0, 1)$ 是软更新参数。

(2) RSU 智能体训练

RSU 智能体采用 DDPG 更新其网络参数。RSU 智能体采用车辆智能体训练过程中抽取的同一批数据计算其 TD 目标, 如下:

$$y_m^l = r_{r_m} + \gamma Q_{\phi_m}(s'_{r_m}, a'_{r_m}) \quad (48)$$

本地 Critic 网络的损失函数定义为:

$$L(\phi_m) = \mathbf{E} \left[\left(Q_{\phi_m}(s_{r_m}, a_{r_m}) - y_m^l \right)^2 \right] \quad (49)$$

类似地, 通过最小化损失函数更新本地 Critic 网络的参数 ϕ_m , 如下:

$$\phi_m \leftarrow \phi_m - v \nabla_{\phi_m} L(\phi_m) \quad (50)$$

RSU 智能体 r_m 的 Actor 网络的策略梯度为:

$$\nabla_{\theta_{r_m}} J_{r_m} = \mathbf{E} \left[\nabla_{\theta_{r_m}} \pi_{r_m}(s_{r_m}; \theta_{r_m}) \nabla_{a_{r_m}} Q_{\phi_m}(s_{r_m}, a_{r_m}) \right] \quad (51)$$

基于策略梯度, 对 RSU 智能体的 Actor 网络参数进行如下优化:

$$\theta_{r_m} \leftarrow \theta_{r_m} + \mu \nabla_{\theta_{r_m}} J_{r_m} \quad (52)$$

同样, 每个智能体的目标网络通过软更新方法做如下更新:

$$\bar{\phi}_m \leftarrow \tau \phi_m + (1 - \tau) \bar{\phi}_m \quad (53)$$

$$\bar{\theta}_{r_m} \leftarrow \tau \theta_{r_m} + (1 - \tau) \bar{\theta}_{r_m} \quad (54)$$

为保证 RSU 智能体输出的连续动作满足 C2—

C4, DCHMARL 将 RSU 智能体输出的原始动作向量为三个长度为 N 的子向量, 并分别通过 Softmax 函数进行归一化。因此, 分配给车辆 v_n 的感知带宽比例 $\rho_{n,m}^{\text{sense}}$ 、上行带宽比例 $\rho_{n,m}^{\text{ul}}$ 和计算资源比例 $\rho_{n,m}^{\text{comp}}$ 计算如下:

$$\rho_{n,m}^* = \frac{\exp(z_{n,m}^*)}{\sum_{i=1}^N \exp(z_{i,m}^*)}, * \in \{\text{sense, ul, comp}\} \quad (55)$$

其中, $z_{n,m}^*$ 为 RSU 智能体 r_m 的 Actor 网络输出的原始动作值。由 Softmax 函数性质可得

$$\rho_{n,m}^* \geq 0, \sum_{i=1}^N \rho_{i,m}^* \leq 1, * \in \{\text{sense, ul, comp}\} \quad (56)$$

最后, 将上述分配比例乘以对应的资源总量, 得到实际资源分配量。因此, 实际分配量满足约束 C2—C4。

算法 1 总结了所提 DCHMARL 算法的流程。

算法 1 DCHMARL

- 1) 初始化: 系统环境, 网络参数
- 2) for 每回合:
 - 3) 更新车辆位置和任务请求
 - 4) for 每步:
 - 5) for 每个智能体 v_n :
 - 6) 基于状态 s_{v_n} , 获得动作 a_{v_n}
 - 7) end for
 - 8) for 每个智能体 r_m :
 - 9) 基于状态 s_{r_m} , 获得动作 a_{r_m}
 - 10) end for
 - 11) 执行动作 a_v, a_r 获得奖励 v 和 r
 - 12) 将 $(s_v, a_v, s'_v, s_r, a_r, s'_r)$ 存储到回放池
 - 13) end for
 - 14) 随机从回放池里取样 mini-batch
 - 15) for 每个智能体 v_n :
 - 16) 根据公式 (41) 计算 y_n^g
 - 17) for $k(k=1,2)$:
 - 18) 根据公式 (42) - (43) 更新全局 Critic 网络
 - 19) end for
 - 20) if 回合数 mod $d = 0$:
 - 21) 根据公式 (44) - (45) 更新 Actor 网络
 - 22) 根据公式 (46) - (47) 更新目标网络
 - 23) end if
 - 24) for 智能体 r_m :
 - 25) 根据公式 (48) 计算 y_m^l
 - 26) 根据公式 (49) - (50) 更新本地 Critic 网络
 - 27) 根据公式 (51) - (52) 更新 Actor 网络
 - 28) 根据公式 (53) - (54) 更新目标网络
 - 29) end for
 - 30) end for

2.3 复杂度分析

为全面评价本文所提 DCHMARL 算法的复杂度, 本小节选择以下三种常用于资源分配的 MARL 算法做对比。

1) MADDPG[23]: 算法将车辆和 RSU 作为异构智能体, 均使用了 DDPG 训练。没有引入双 Critic 机制, 车辆和 RSU 智能体独立决策。

2) MDRCO[24]: 车辆选择距离最近的 RSU 作为服务 RSU, 算法中仅设计 RSU 智能体, 采用 Actor-Critic 框架训练 RSU 智能体。

3) MAPPO[20]: 算法将车辆和 RSU 作为异构智能体, 使用了 PPO 进行联合训练。没有引入双 Critic 机制, 车辆智能体基于 Categorical 分布决策, RSU 智能体基于 Dirichlet 分布决策。

用 L_a 和 L_c 分别表示 Actor 网络和 Critic 网络的层数, e_i^a 和 e_i^c 分别表示 Actor 网络和 Critic 网络第 i 层的神经元个数, 当 Actor 网络和 Critic 网络均为全连接时, 其总连接数分别为:

$$\begin{aligned} \mathfrak{C}_a &= \sum_{i=2}^{L_a-1} (e_{i-1}^a e_i^a + e_i^a e_{i+1}^a) \\ \mathfrak{C}_c &= \sum_{i=2}^{L_c-1} (e_{i-1}^c e_i^c + e_i^c e_{i+1}^c) \end{aligned} \quad (57)$$

用 E 表示总训练回合, T 表示每个训练回合下的总步数, B 表示 mini-batch 大小。因此, DCHMARL 与两个对比算法的训练复杂度表示如下。

(1) DCHMARL:

$$O\left(ETB\left(N\left(\mathfrak{C}_a + 2\mathfrak{C}_c\right) + M\left(\mathfrak{C}_a + \mathfrak{C}_c\right)\right)\right) \quad (58)$$

(2) MADDPG:

$$O\left(ETB(N+M)(\mathfrak{C}_a + \mathfrak{C}_c)\right) \quad (59)$$

(3) MDRCO:

$$O\left(ETBM(\mathfrak{C}_a + \mathfrak{C}_c)\right) \quad (60)$$

(4) MAPPO:

$$O\left(ETB((N+M)\mathfrak{C}_a + \mathfrak{C}_c)\right) \quad (61)$$

可以看出, 本文所提 DCHMARL 算法的训练复杂度略高于对比算法, 但其带来的复杂度增长与智能体数量与网络连接规模呈多项式关系, 计算代价在可接受范围。DCHMARL 增加训练开销, 旨在换取更优的策略学习能力和系统优化效果。仅从理论复杂度, 不足以全面评判算法优劣。因此, 下一节将通过仿真实验, 对 DCHMARL 性能做全面评估与对比。

3 仿真分析

本文搭建了图 1 所示的 ISCC-IoV 系统。其中, 研究的车道长 500m, 车道左上角设为坐标原点 (0, 0), RSU 沿 x 轴均匀分布, 且最左边的 RSU 位于原点。此外, 假设感知任务的最大容忍时延 $T_n^{\max}$ 服从均匀分布 $U(0.05, 0.08)$ 秒。其他重要的环境仿真参数及设置如表 1 所示。设定每回合包含 100 步, 训练持续 5000 回合, 每 100 回合进行一次评估, 其他重要的神经网络参数及设置如表 2 所示。

本文仿真实验环境基于 Windows 10、Python 3.12.0 和 PyTorch 2.7.1, 硬件配置为 Intel Xeon Gold 6248 CPU @ 2.50GHz、NVIDIA GeForce RTX 4090 GPU 及 256 GB 系统内存

表 1 系统仿真参数及设置

系统仿真参数	值
总时隙长度 Δt	0.1 s
感知时隙长度 τ^{sense}	0.05 s
道路宽度	3.7 m
车辆总数	5-25
RSU-MEC 数量 M	5
感知请求泊松分布均值	0.5
阴影衰落标准差 σ_r, σ_c	4 dB
车辆发射功率 P_n^{veh}	30 dBm
车辆雷达探测信号功率 $P_{n,i}^{\text{rad}}$	40 dBm
RSU 雷达探测信号功率 $P_{n,m}^{\text{rad}}$	40 dBm
AWGN 功率谱密度 n_0	-174 dBm/Hz
系统上行链路带宽 B^{ul}	3 MHz
MEC 服务器的最大计算资源 F	6 GHz
比例系数 α, β, γ	0.05, 100, 10

3.1 所提 ISCC-IoV 系统资源分配方案性能分析

为验证所提 ISCC-IoV 系统资源分配方案的性

表 2 神经网络参数及设置

神经网络参数	值
经验回放池大小 D	100000
mini-batch 大小 B	64
Critic 网络学习率 ν	0.001
Actor 网络学习率 μ	0.0001
折扣因子 γ	0.99
软更新参数 τ	0.0005

能及有效性, 本文对 $P1$ 的优化函数作为评估标准, 并选取了以下两种常见的资源分配方案进行性能对比。为公平起见, 对比方案也使用了 DCHMARL 算法作为优化问题的求解方法。

1) 对比方案 1^[16]: 该方案以系统能耗为约束, 以最小化时延为优化目标, 但方案未考虑感知性能。

2) 对比方案 2^[9]: 该方案以最小化系统能耗为优化目标, 但未考虑时延约束和感知性能。



图 2 车辆数对资源分配方案性能的影响

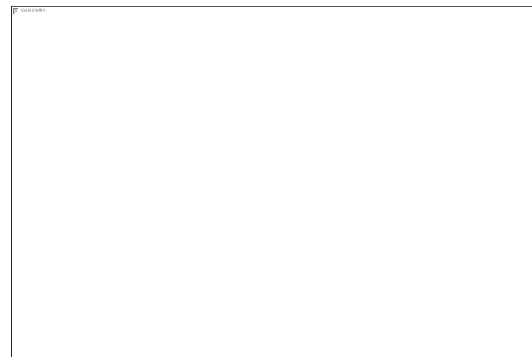


图 3 资源倍数对资源分配方案性能的影响(车辆数=20)

图 2 和图 3 分别展示了不同车辆数和不同资源倍数下各资源分配方案的车辆平均效用值对比, 其中, 资源倍数定义为系统上行带宽和 MEC 服务器

计算资源默认值的倍数。可以看到,随着车辆数增加,更多的车辆竞争有限的资源,因此各方案的车辆平均效用都会降低;而随着资源增加,车辆有更多资源供选择,因此车辆平均效用都会增加。其中,本文所提的多维资源分配方案在不同场景下都实现了最高效用值,证明了在综合考虑通、感、算性能的情况下,所提方案可以实现更优系统效用。此外,对比方案2的性能优于对比方案1,这是因为对比方案1在降低时延的同时可能带来额外的能耗开销;而对比方案2在能耗项上的优化对系统效用的提升更为显著,因此对比方案2可以一定程度上优化通、感、算的总体性能。

3.2 所提 DCHMARL 算法性能分析

为全面评估所提 DCHMARL 算法的性能,本节设计了多组对照消融方案,以验证各核心模块的有效性。

1) DCHMARL-SQ: 该算法取消了车辆智能体的双 Q 网络设计,其余设计同 DCHMARL。

2) DCHMARL-LC: 该算法为车辆智能体和 RSU 智能体均配备本地 Critic 网络,其余设计同 DCHMARL。

3) DCHMARL-NGS: 该算法未对车辆智能体的离散动作进行 Gumbel-Softmax 处理,其余设计同 DCHMARL。

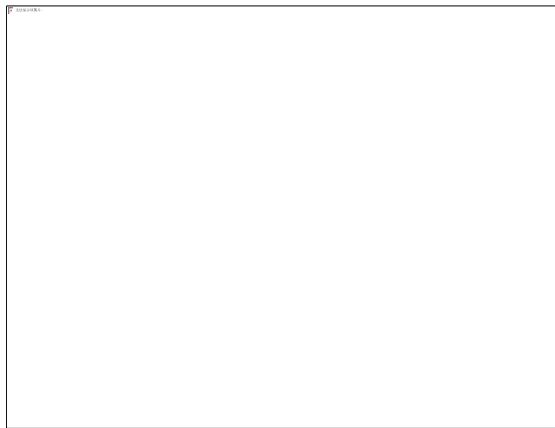


图4 DCHMARL 及消融方案的系统奖励(车辆数=10)

图4对比了 DCHMARL 及消融方案在车辆数为 10 时的系统奖励。可以看出, DCHMARL 获得了最高且较为稳定的系统奖励,说明其各项设计均能有效提高算法性能。其中,未采用 Gumbel-Softmax 处理离散动作的 DCHMARL-NGS 表现出一定的性能退化,验证了该方法在提升混合动作训练稳定性

方面的有效性。同时, DCHMARL-SQ 的系统奖励也有所下降,说明双 Q 网络能够有效缓解价值高估问题。此外,仅采用本地 Critic 网络的 DCHMARL-LC 性能下降最明显,说明车辆动作具有较强的全局影响,仅依赖本地 Critic 网络难以充分评估其影响,验证了双 Critic 结构的必要性。

进一步地,本节选取 MADDPG、MDRCO、MAPPO 三种常见 MARL 基准算法,与所提 DCHMARL 算法开展全面性能对比。



图5 不同算法的系统奖励对比(车辆数=10)

图5对比了不同算法在车辆数为 10 时的系统奖励。可以看出,本文所提 DCHMARL 显示了最快收敛速度,且实现了更高的系统奖励,验证了其在解决复杂的通、感、算多维资源分配问题上的有效性。这得益于 DCHMARL 的车辆智能体通过双 Q 学习有效缓解了高估问题,并通过延迟更新 Actor 网络获得稳定的策略;同时,分层 Actor 设计增强了车辆智能体与 RSU 智能体之间的协同优化能力。此外,MDRCO 在训练初期显示出较高系统奖励,而随着训练进行,系统奖励显著降低并逐渐收敛于较低的值,说明仅依靠 RSU 智能体难以根据经验进行高效学习。MAPPO 和 MADDPG 的系统奖励介于 DCHMARL 与 MDRCO 之间,且 MAPPO 略优于 MADDPG,说明二者可以实现较优解,但容易陷入局部最优。

图6展示了在不同车辆数下各算法的车辆奖励的箱线图。可以看出,随着车辆数的增加,资源的竞争更加激烈,因此,各算法的车辆奖励均降低。本文所提 DCHMARL 实现了最高奖励且分布更集中,表现出最大平均值、最小箱体下降幅度,说明了其在不同环境下具备较强的适应能力和鲁棒性。

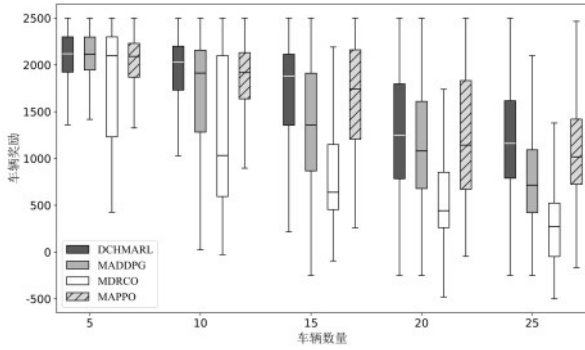


图6 车辆数对车辆奖励性能的影响

此外, MAPPO 表现出仅次于 DCHMARL 的性能, MADDPG 箱体波动较大, 说明其易受多智能体环境复杂度的影响, 而 MDRCO 箱体呈现出其奖励低、波动较大和稳定性不足的缺陷。

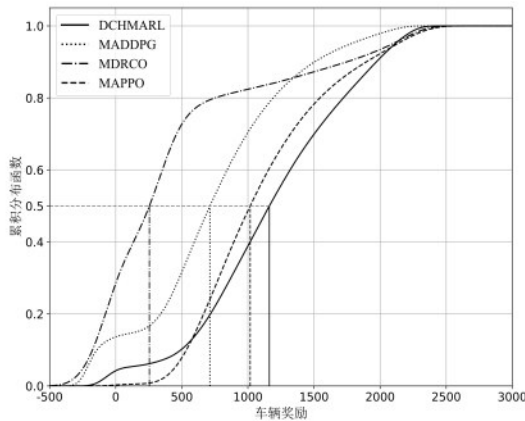


图7 车辆奖励的累积分布函数(车辆数=25)

图7对比了各算法在车辆数为25时车辆奖励的累积分布函数。可以看出, 本文所提 DCHMARL 的曲线靠右, 在相同累积概率下车辆奖励值最高, 验证了其在车辆奖励上的优势。MAPPO 的曲线位于 DCHMARL 左侧, 且明显优于 MADDPG 和 MDRCO, 但在高奖励区间的表现略逊于 DCHMARL。MDRCO 的曲线在低奖励区占比明显, 表明其奖励值主要集中于中低奖励范围, 获得高奖励的能力相对有限。MADDPG 的车辆奖励则介于 MDRCO 与 DCHMARL 算法之间。

图8展示了在不同车辆数下各算法的平均车辆效用。可以看到, 随着车辆数的增加, 车辆平均效用均呈现下降趋势。DCHMARL 在不同场景下均能实现最优性能, 体现了其对复杂环境的适应性和鲁棒性。同时, MAPPO 缺乏对车辆与 RSU 的异构



图8 车辆数对车辆平均效用的影响

决策特征的针对性设计, 因此其车辆平均效用略低于 DCHMARL。在车辆数等于5时, MDRCO 的车辆效用略高于 MADDPG, 这是因为车辆密度很低时, 车辆选择就近 RSU 即可获得较好的传输条件, 从而实现较高的车辆效用。但随着车辆数增多, MADDPG 的车辆智能体与 RSU 智能体的协同决策优势逐渐显现, 表现出更优的整体性能。

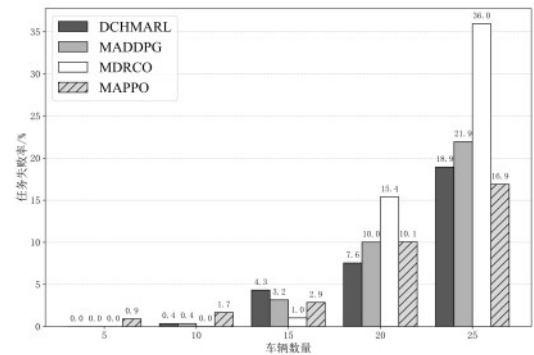


图9 车辆数对任务失败率的影响

图9展示了车辆数对各算法任务失败率的影响。可以看到, 给定资源下, 更多的车辆数导致任务获得资源难度增大, 各算法的任务失败率因此均增加。其中, 车辆数在5-10时, 四个算法的任务失败率差异较小。车辆数为15时, DCHMARL 的任务失败率略高, 而 MDRCO 任务失败率最低。结合图8可知, 这是由于 DCHMARL 更注重性能均衡, 通过接受一定的任务失败率为代价来换取更高的车辆效用。相比之下, MDRCO 的车辆就近选择 RSU, 虽降低了任务失败率, 但其车辆效用相对较低。车辆数进一步增加时, DCHMARL 在保证低任务失败率的同时, 实现最高效用, 体现了其在高负载场景下更强适应能力。

图10和图11分别展示了不同资源倍数时各算



图10 资源倍数对车辆平均效用的影响(车辆数=20)

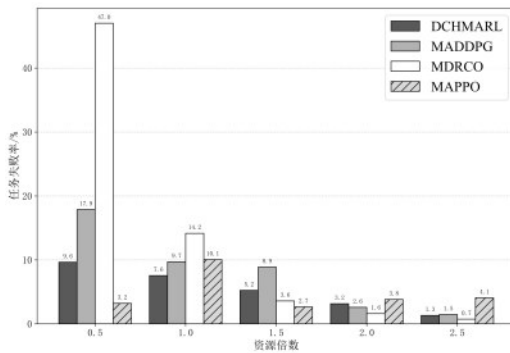


图11 资源倍数对任务失败率的影响(车辆数=20)

法的平均车辆效用和任务失败率。可以看到,随着系统可用资源增加,各算法的车辆效用均呈上升趋势,任务失败率总体下降。当资源倍数为0.5时,DCHMARL与MAPPO均表现出较高的车辆效用与较低任务失败率,表明其在资源稀缺场景下具有较强的适应能力;在资源倍数为1.0-2.5时,DCHMARL的车辆效用增长更为显著,并始终保持最高的车辆效用,而MAPPO车辆效用增幅有限,且任务失败率呈先升后降的波动趋势,反映出其对冗余资源调度效率偏低,鲁棒性较差。尽管其在部分场景下DCHMARL的任务失败率略高于MDRCO或MAPPO,但综合两项指标,DCHMARL可在车辆效用与任务完成率间实现有效权衡。其他两个算法性能欠佳,其中,MADDPG的车辆效用始终高于MDRCO,而后者仅在部分场景下具有较低任务失败率。

4 总结

本文面向多RSU多车辆的ISCC-IoV系统,提出了一种新颖的基于端-边协作感知的多维联合资源分配方案。该方案构建了一个综合考虑通、感、算性能的系统效用函数,并提出了一种双Critic异

构多智能体强化学习算法DCHMARL。仿真结果表明,与常见资源分配方案相比,所提的资源分配方案能够在综合考虑通、感、算性能的情况下获得更高的系统效用。此外,对比传统强化学习算法,DCHMARL能够显著提升系统效用、降低任务失败率,并表现出更优的收敛性与稳定性。

参考文献:

- [1] Mollah M B, Zhao J, Niyato D, et al. Blockchain for the internet of vehicles towards intelligent transportation systems: A survey[J]. IEEE Internet of Things Journal, 2020, 8(6): 4157-4185.
- [2] Cheng X, Duan D, Gao S, et al. Integrated sensing and communications (ISAC) for vehicular communication networks (VCN)[J]. IEEE Internet of Things Journal, 2022, 9(23): 23441-23451.
- [3] Cui Y, Liu F, Jing X, et al. Integrating sensing and communications for ubiquitous IoT: Applications, trends, and challenges[J]. IEEE network, 2021, 35(5): 158-167.
- [4] Liu F, Masouros C, Petropulu A P, et al. Joint radar and communication design: Applications, state-of-the-art, and the road ahead[J]. IEEE Transactions on Communications, 2020, 68(6): 3834-3862.
- [5] Kuang Z, Li L, Gao J, et al. Partial offloading scheduling and power allocation for mobile edge computing systems[J]. IEEE Internet of Things Journal, 2019, 6(4): 6774-6785.
- [6] Xu Y, Ren J, Wang G, et al. A blockchain-based nonrepudiation network computing service scheme for industrial IoT[J]. IEEE Transactions on Industrial Informatics, 2019, 15(6): 3632-3641.
- [7] Wen D, Zhou Y, Li X, et al. A survey on integrated sensing, communication, and computation[J]. IEEE Communications Surveys & Tutorials, 2024.
- [8] Zhao J, Ren R, Zou D, et al. IoV-oriented integrated sensing, computation, and communication: System design and resource allocation[J]. IEEE Transactions on Vehicular Technology, 2024, 73(11): 16283-16294.
- [9] Gu X, Wu Q, Fan P, et al. DRL-based federated self-supervised learning for task offloading and resource allocation in ISAC-enabled vehicle edge computing[J]. Digital Communications and Networks, 2024.
- [10] Huang B, Fan X, Zheng S, et al. Collaborative Sensing-Aware Task Offloading and Resource Allocation for Integrated Sensing-Communication-and Computation-Enabled Internet of Vehicles (IoV) [J]. Sensors, 2025, 25(3): 723.
- [11] He Y, Yu G, Cai Y, et al. Integrated sensing, computation, and communication: System framework and performance optimization[J]. IEEE Transactions on Wireless Communications, 2023, 22(2): 1114-1128.
- [12] Cheng K, Fang X, Wang X. Optimized resource allocation and time partitioning for integrated communication, sensing, and edge computing network[J]. Computer Communications, 2022, 194: 240-249.
- [13] Zhao L, Wu D, Zhou L, et al. Radio resource allocation for integrated sensing, communication, and computation networks[J]. IEEE Transactions on Wireless Communications, 2022, 21(10): 8675-8687.
- [14] Zhang Q, Chen Z, Xia B, et al. Design and optimization of edge computing for data fusion in V2I cooperative systems[C]//2020 IEEE/CIC International Conference on Communications in China (ICCC). IEEE,

- 2020: 466-471.
- [15] Gabb M, Digel H, Müller T, et al. Infrastructure-supported perception and track-level fusion using edge computing[C]//2019 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2019: 1739-1745.
- [16] Liu Q, Luo R, Liang H, et al. Energy-efficient joint computation offloading and resource allocation strategy for ISAC-aided 6G V2X networks[J]. IEEE Transactions on Green Communications and Networking, 2023, 7(1): 413-423.
- [17] Zhao Y, Hou F, Huang J, et al. Delay optimization in vehicular edge computing with sensing information fusion and heterogeneous tasks [C]//2024 IEEE International Conference on Communications Workshops (ICC Workshops). IEEE, 2024: 1888-1894.
- [18] Tang M, Wong V W S. Deep reinforcement learning for task offloading in mobile edge computing systems[J]. IEEE Transactions on Mobile Computing, 2020, 21(6): 1985-1997.
- [19] Wei R, Qin T, Huang J, et al. Resource allocation scheduling scheme for task migration and offloading in 6G Cybertwin internet of vehicles based on DRL[J]. IET communications, 2024, 18(18): 1244-1265.
- [20] Ning X, Zeng M, Hua M, et al. Multiple reconfigurable intelligent surfaces aided vehicular edge computing networks: A MAPPO-based approach[J]. IEEE Transactions on Vehicular Technology, 2024, 73(11): 17496-17509.
- [21] Generation T. Technical specification group radio access network; study on channel model for frequencies from 0.5 to 100 GHz[J]. 3GPP TR 38.901 v16. 1.0, 2019.
- [22] Dou C, Huang N, Wu Y, et al. Sensing-efficient NOMA-aided integrated sensing and communication: A joint sensing scheduling and beamforming optimization[J]. IEEE Transactions on Vehicular Technology, 2023, 72(10): 13591-13603.
- [23] Liu J, Wang Y, Pan D, et al. QoS-aware task offloading and resource allocation optimization in vehicular edge computing networks via MADDPG[J]. Computer Networks, 2024, 242: 110282.

- [24] Zhu X, Luo Y, Liu A, et al. Multiagent deep reinforcement learning for vehicular computation offloading in IoT[J]. IEEE Internet of Things Journal, 2020, 8(12): 9763-9773.



杨新杰（1971—），男，宁波大学信息科学与工程学院副教授、硕士生导师，主要研究方向为下一代移动通信系统架构、移动物联网接入技术、协作中继网络性能等。



李睿哲（2002—），男，宁波大学硕士生，主要研究方向为车联网、通感一体化、移动边缘计算。



梅杰（1991—），男，宁波大学信息科学与工程学院副研究员、硕士生导师，主要研究方向为车联网以及网络智能化。



马楠（1979—），男，北京邮电大学信息与通信工程学院教授，北京无线通信测试技术重点实验室主任，主要研究方向包括无线通信理论、测试技术和大数据应用