基于分层深度强化学习的RIS辅助车载边缘计算中信息年龄与能效优化

兰 军,贾向东*,寇志龙,包红丽,梁文艳,武婧婧
(西北师范大学计算机科学与工程学院,甘肃兰州 730071)

摘 要: 随着第五代(5G)和第六代(6G)移动通信技术的发展以及智能交通系统(Intelligent Transportation Systems, ITS)的成熟,车联网(Internet of Vehicles,IoV)逐渐成为智慧交通的重要支撑. 车载边缘计算(Vehicular Edge Computing, VEC)通过在基站(Base Station, BS)或路侧单元(Roadside Unit,RSU)部署边缘服务器,为车载终端提供低时延计算服务. 然而,车辆高速移动导致的信道衰落、能量受限及任务动态变化,使系统难以兼顾信息时效性与能量效率. 智能反射面(Reconfigurable Intelligent Surface,RIS)能够通过相位可控反射重构传播环境,为VEC系统提供提升链路可靠性和能效的新途径. 本文针对RIS辅助VEC系统中信息年龄(Age of Information,AoI)与能量消耗的协同优化问题,提出一种基于分层深度强化学习(Hierarchical Deep Reinforcement Learning,HDRL)的多目标优化框架. 首先,本文构建了一个考虑车辆运动特性、三维几何信道和任务动态的系统模型,并建立最小化AoI与能量消耗加权和的非凸优化问题. 其次,本文设计了具有"集中控制—分布协同"特性的分层混合强化学习架构:上层采用双延迟确定性策略梯度算法(Twin Delayed Deep Deterministic Policy Gradient,TD3)实现RIS相位连续优化,下层采用联邦多智能体深度确定性策略梯度算法(Federated Multi-Agent Deep Deterministic Policy Gradient,FMADDPG)实现功率与计算频率的分布式资源分配. 为增强两层间的协同学习,本文提出联合预训练与轨迹嵌入机制:上层TD3控制器预生成RIS相位轨迹供下层FMADDPG策略初始化使用,从而实现跨层感知与加速收敛. 此外,本文从理论上证明了FMADDPG算法在有界状态空间与Lipschitz连续奖励条件下的稳定收敛性. 仿真结果表明,所提HDRL框架在信息新鲜度与能耗权衡方面显著优于软演员评论家算法(Soft Actor-Critic,SAC)、Q值混合网络(Q-value MIXing,QMIX)和块坐标下降(Block Coordinate Descent,BCD)等基准方法. 与SAC算法相比,平均信息年龄降低约15%,系统能量效率提升约29%,在信道估计误差与遮挡概率较高的环境下仍保持稳定性能. 本文的主要创新包括:(1)构建了RIS辅助VEC系统中AoI与能耗的多目标优化模型;(2)提出了结合TD3与FMADDPG的分层强化学习框架,实现集中控制与分布式协同;(3)设计了联合预训练与轨迹嵌入机制,有效提升了算法的收敛速度与策略感知能力. 该研究为RIS辅助车联网的低时延与高能效优化提供了新的智能决策范式,对未来智能交通系统的边缘智能化具有重要参考价值.

关键词: 车联网;车载边缘计算;智能反射面;信息年龄;能量效率;分层深度强化学习
基金项目: 国家自然科学基金(No.62261048,No.61861039);甘肃省高校产业支持计划(No.2025CYZC-014)
中图分类号: TN929.5 **文献标识码:** A **文章编号:** 0372-2112(XXXX)XX-0001-16
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20250415

Age of Information and Energy Efficiency Optimization in RIS-Assisted Vehicular Edge Computing Based on Deep Reinforcement Learning

LAN Jun, JIA Xiang-dong*, KOU Zhi-long, BAO Hong-li, LIANG Wen-yan, WU Jing-jing
(*School of Computer Science and Engineering*, *Northwest Normal University*, *Lanzhou*, *Gansu* 730071, *China*)

Abstract: With the advancement of fifth-generation (5G) and sixth-generation (6G) mobile communication technologies and the continuous development of intelligent transportation systems (ITS), the internet of vehicles (IoV) has gradually become a key foundation for smart transportation. Vehicular edge computing (VEC) provides low-latency computing services for vehicular terminals by deploying edge servers at base station (BS) or roadside unit (RSU). However, the high mobility of vehicles results in severe channel fading, limited energy resources, and dynamic task variations, which make it chal-

收稿日期:2025-05-25;录用日期:2025-10-09;责任编辑:覃怀银
*通讯作者:贾向东

lenging to jointly guarantee information freshness and energy efficiency. Reconfigurable intelligent surface (RIS) technology, capable of reconfiguring the wireless propagation environment through controllable phase reflections, offers a promising solution to improve link reliability and energy efficiency in VEC systems.This paper proposes a hierarchical deep reinforcement learning (HDRL)-based multi-objective optimization framework to jointly optimize the Age of Information (AoI) and energy consumption in RIS-assisted VEC systems. Firstly, a system model is established that considers vehicular mobility, three-dimensional geometric channels, and dynamic task arrivals, and a non-convex optimization problem is formulated to minimize the weighted sum of AoI and energy consumption. Secondly, a hierarchical hybrid reinforcement learning architecture with "centralized control and distributed coordination" is designed. In the upper layer, the twin delayed deep deterministic policy gradient (TD3) algorithm is employed to continuously optimize RIS phase configurations, while the lower layer adopts the federated multi-agent deep deterministic policy gradient (FMADDPG) algorithm to realize distributed power allocation and computation frequency control.To enhance cross-layer learning coordination, a joint pretraining and trajectory-embedding mechanism is proposed, where the upper-layer TD3 controller generates representative RIS phase trajectories for initializing the policies of lower-layer FMADDPG agents. This mechanism effectively improves cross-layer awareness and accelerates convergence. In addition, theoretical analysis proves the stability and convergence of the FMADDPG algorithm under bounded state spaces and Lipschitz-continuous reward conditions.Simulation results demonstrate that the proposed HDRL framework significantly outperforms benchmark methods such as the soft actor-critic (SAC), q-value mixing (QMIX) and block coordinate descent (BCD) algorithms in terms of balancing information freshness and energy efficiency. Compared with the SAC algorithm, the proposed approach reduces the average AoI by approximately 15% and improves energy efficiency by about 29%, while maintaining stable convergence under high channel estimation errors and blockage probabilities.The main innovations of this paper are as follows: (1) a multi-objective optimization model is developed for joint AoI and energy efficiency optimization in RIS-assisted VEC systems; (2) a hierarchical reinforcement learning framework combining TD3 and FMADDPG is proposed to achieve centralized RIS control and distributed resource coordination; (3) a joint pretraining and trajectory-embedding mechanism is designed to improve convergence speed and policy adaptability. This study provides a novel intelligent decision-making paradigm for low-latency and energy-efficient vehicular edge computing and offers valuable insights into the edge intelligence development of future intelligent transportation systems.

Key words: internet of vehicles; vehicular edge computing; reconfigurable intelligent surface; age of information; energy efficiency; hierarchical deep reinforcement learning

Foundation Item(s): National Natural Science Foundation of China (No.62261048, No.61861039);Gansu Province University Industry Support Plan (2025CYZC-014)

1 引言

随着5G/6G移动通信技术的持续演进和智能交通系统(Intelligent Transportation Systems,ITS)的快速发展,车联网(Internet of Vehicles,IoV)已成为构筑未来智慧城市与实现自动驾驶愿景的关键基础[1~3]. 在此背景下,车载边缘计算(Vehicular Edge Computing,VEC)作为一种变革性范式,通过将计算和存储资源下沉至基站(Base Station,BS)或路侧单元(Roadside Unit,RSU),为车载设备提供实时、低延迟的计算服务[4,5]. 该分布式架构有效缓解了传统云计算中海量数据传输至远端中心所带来的高延迟与链路拥塞问题,从而支撑自动驾驶中的即时决策、协同感知及实时交通管理等关键应用[6~8].

然而,VEC系统在实际部署中面临诸多挑战,主要包括高速移动性导致的多普勒频移、有限的计算与能量资源,以及对信息年龄(Age of Information,AoI)的严格要求[9,10]. 此外,无线信道的衰落、多径效应及非视距(Non-Line-of-Sight,NLoS)传播进一步限制了通信可靠性与效率. 为此,智能反射面(Reconfigurable Intelligent Surface,RIS)作为一种新兴技术应运而生[11,12]. RIS由大量无源单元组成,可智能调整电磁波的相位和幅度,从而在不增加额外能耗的情况下重构无线传播环境[13,14],凭借其无源、低成本、易部署等特性,RIS在高动态车载环境中展现出巨大潜力,但同时引入了如何高效优化海量反射单元相位并与通信、计算资源协同的问题. 该问题通常为高维、非凸、强耦合的优化挑战,传统基于模型的优化方法难以应对.

面对VEC系统的高动态性与多维耦合特征,传统凸优化方法依赖精确模型和完美信道状态信息(Channel State Information,CSI),在高维决策空间中往往计算复杂度高且难以实时收敛[15~18]. 近年来,深度强化学习(Deep Reinforcement Learning,DRL)作为一种数据驱动的智能决策范式,已在复杂动态资源管理中展现出卓越性能[2,19~22]. DRL通过与环境交互学习最优策略,无

需精确系统模型即可实现高效决策,从而为RIS辅助的VEC系统提供了一种可行的智能优化方案[23~26].

目前,基于DRL的RIS辅助边缘计算研究主要聚焦于RIS相移、用户调度、功率分配、卸载决策及计算资源分配的优化,以提升系统吞吐量或能效[6,27~29]. 例如,文献[9]提出基于近端策略优化(Proximal Policy Optimizatio,PPO)的DRL方法用于主动RIS辅助频分多址-非正交多址移动边缘计算(Frequency Division Multiple Access-Non-Orthogonal Multiple Access Mobile Edge Computing,FDMA-NOMA MEC)系统,以最大化总计算速率;文献[12]则研究了主动RIS辅助无人机网络的轨迹与相移优化,用以避免计算超时;文献[13]考虑双RIS辅助的VEC网络,通过DRL优化卸载与计算资源以提高总卸载效率. 上述研究验证了DRL在提升RIS辅助MEC性能方面的潜力,但多聚焦于单一指标(如吞吐量或能效),未能综合考虑AoI与能效的权衡,且在多维耦合和动态信道环境下仍存在收敛与协同效率的瓶颈.

随着系统复杂度的提升,单一智能体DRL易遭遇"维数灾难". 因此,多智能体深度强化学习(Multi-Agent Deep Reinforcement Learning,MADRL)与分层深度强化学习(Hierarchical Deep Reinforcement Learning,HDRL)被提出以应对大规模分布式优化问题[30,31]. 文献[6]针对多输入多输出-非正交多址(Multiple-Input Multiple-Output-Non-Orthogonal Multiple Access,MIMO-NOMA)辅助VEC系统提出基于分布式DRL的功率分配方法,利用深度确定性策略梯度(Deep Deterministic Policy Gradient,DDPG)实现在线最优策略学习,虽能应对信道不确定性与隐私约束,但未考虑AoI影响. 文献[18]设计MADRL框架优化多RIS辅助IoV网络中的任务卸载与资源分配,验证了MADRL在大规模IoV中的有效性,但优化目标为一般效用函数而非AoI与能效联合优化. 文献[32]针对同时发射与反射可重构智能反射面(Simultaneously Transmitting and Reflecting Reconfigurable Intelligent Surface, STAR-RIS)辅助NOMA MEC系统提出分解式MADRL方法联合优化多维资源,以最小化长期能耗,但其策略特定于STAR-RIS机制,缺乏对AoI的考虑. 总体而言,如何在高动态车载环境下融合双延迟确定性策略梯度算法(Twin Delayed Deep Deterministic Policy Gradient,TD3)、DDPG等算法的联邦多智能体变体,实现分层协同优化,仍有待深入研究.

随着对实时应用需求的日益增长,AoI作为衡量数据新鲜度的关键指标,在RIS辅助网络中受到了越来越多的关注. 例如,文献[27]也探索了RIS辅助车联网(Vehicle-to-Everything,V2X)通信中具有超可靠低延迟通信(Ultra-Reliable Low Latency Communication,URLLC)保障的能效资源分配,其中AoI是一个隐式或显式需要考虑的指标;文献[28]深入探讨了RIS辅助车载网络中的AoI感知资源分配问题,旨在通过优化通信资源来降低信息年龄;文献[31]则研究了RIS辅助物联网(Internet of Things,IoT)网络中的AoI优化,证明了RIS在改善信息新鲜度方面的有效性. 这些研究证实了RIS在降低AoI方面的潜力,并通过DRL实现动态决策. 然而,鲜有工作能够在一个统一的框架内,同时精确建模并有效地联合优化AoI和能量效率,尤其是在高动态且多耦合变量的RIS辅助VEC场景中.

综上所述,尽管现有研究在RIS辅助边缘计算领域取得了显著进展,但仍存在以下待解决的关键研究空白,成为本文深入探索的出发点:

(1)多数现有研究仅优化单一性能指标(如吞吐量、能效或AoI),未充分考虑车载任务对AoI与系统能效的协同需求. 在VEC场景下,任务实时性与系统能耗之间存在内在冲突,需实现平衡以优化整体性能.

(2)RIS相移、卸载决策、功率分配和本地计算频率等多维变量高度耦合且均为连续动作空间,形成高维非凸优化问题. 传统优化或单一DRL方法在处理复杂动态决策时,常面临收敛缓慢、局部最优及高计算开销等问题,难以满足实时性要求.

(3)尽管HDRL可有效应对大规模复杂问题,但如何在RIS相移与计算资源的分层优化中合理选取算法并实现高效协同仍具挑战. 同时,联邦学习与MADRL的融合在隐私保护、学习效率及RIS协同优化方面仍需深入研究.

本文的主要贡献总结如下:

(1)提出一种RIS辅助的车载边缘计算系统中联合优化AoI与能效的多目标分层强化学习架构. 针对车载环境中低时延与高能效的协同需求,本文将AoI与能量消耗纳入统一优化框架,系统性地构建了以RIS为控制中枢的分层式资源管理机制,突破了传统方法优化目标单一、耦合处理能力有限的局限.

(2)设计了一种混合深度强化学习框架,结合TD3与联邦多智能体深度确定性策略梯度算法(Federated Multi-Agent DDPG,FMADDPG)实现"集中控制-分布协同"的上下层优化结构. 上层采用TD3算法优化连续的RIS相位控制变量,下层通过FMADDPG实现分布式功率与计算频率分配,二者分别应对高维全局控制与分布式资源异构的挑战,提升系统整体适应性与扩展性.

(3)提出上下层策略间的联合预训练机制,实现策略感知增强与训练协同加速. 在不破坏分层结构的前提下,设计了"RIS轨迹池"嵌入机制,使下层FMADDPG在策略初始化阶段能够利用上层TD3生成的先验相位

轨迹，实现策略的软耦合联动，提升早期学习效率并缩短收敛时间.

(4)构建面向车联网场景的三维几何信道与任务动态模型，搭建仿真环境并开展对比实验验证算法有效性. 实验结果表明，本文方法在AoI降低与能耗控制方面显著优于软演员评论家算法(Soft Actor-Critic，SAC)、Q值混合网络(Q-value MIXing，QMIX)、块坐标下降(Block Coordinate Descent，BCD)等典型基准算法，在动态城市交通环境下仍保持良好的鲁棒性和收敛性，验证了所提方法在RIS辅助车载边缘计算系统中的实用性与前瞻性.

本文其余部分结构安排如下：第二节介绍系统模型及问题建模；第三节详细描述所提出的混合DRL优化框架，包括基于TD3的RIS控制器与FMADDPG功率分配器；第四节展示仿真结果及性能分析；第五节总结全文并讨论未来的研究方向.

2 系统模型与问题表述

2.1 系统模型

RIS辅助的车载边缘计算系统如图1所示，由三大核心组成部分构成：车载用户群体、RIS以及部署在基站的边缘计算服务器. 系统中的每辆VU均配备单天线，动态生成计算任务，其到达过程服从泊松分布，即第 k 辆车在时刻 t 的任务到达率记作 $\lambda_k(t)\sim\text{Poisson}(\Lambda_k)$. 其中 Λ_k 是车辆特定的平均任务到达率，体现了不同车辆在业务需求上的差异性. 为应对车联网场景中信道快速衰落和遮挡等挑战，系统引入RIS作为可编程信号反射器，提升链路可靠性与通信性能. 该RIS采用由 m 个可调被动单元构成的均匀矩形阵列，每个单元可在离散集合中灵活调整其相位偏移，具体相位取值为 $\phi_m\in\left\{0,\frac{2\pi}{2^b},\frac{2\pi\cdot 2}{2^b},\cdots,\frac{2\pi\cdot(2^b-1)}{2^b}\right\}$，其中 ϕ_m 为第 m 个RIS单元的相位，b 表示相位调控的比特分辨率. 在本系统中，采用3比特控制精度，从而实现较细粒度的信号调制. 边缘计算服务器部署于多天线基站内部，最大计算性能达到 $F_{\max}=25$ TFLOPS，能够处理来自大量车载用户卸载的计算密集型任务，同时满足智能驾驶场景下对低延迟与高可靠性的严格要求.

2.2 移动模型

在所研究的RIS辅助VEC网络中，车辆运行于动态的城市环境中，其移动模式受到道路基础设施、交通密度及车间交互等因素的影响. 为准确建模车辆运动行为，本文采用离散时间的移动模型，该模型同时考虑了由环境不确定性引起的随机扰动与确定性的运动学规律. 对于任意车辆 k，其在时刻 $t+1$ 的速度与航向角

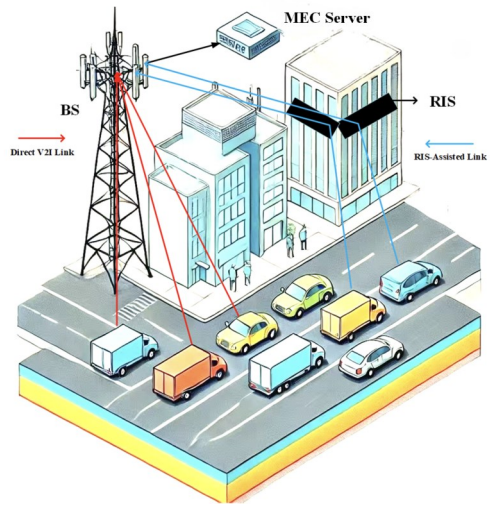

图1 RIS辅助车载边缘计算系统

更新规则如式(1)所示：

$$\begin{cases} v_k(t+1)=v_k(t)+a_k(t)\Delta t+\epsilon_v+\eta_v(t) \\ \theta_k(t+1)=\theta_k(t)+\omega_k(t)\Delta t+\epsilon_\theta+\eta_\theta(t) \end{cases} \tag{1}$$

其中，$v_k(t)$ 与 $\theta_k(t)$ 分别表示车辆 k 在时刻 t 的速度与航向角，$a_k(t)$ 与 $\omega_k(t)$ 分别代表车辆的加速度与角速度，$\epsilon_v\sim\mathcal{N}(0,\sigma_v^2)$、$\epsilon_\theta\sim\mathcal{N}(0,\sigma_\theta^2)$ 表示由于传感器噪声与控制误差所引入的高斯分布不确定性项，而 $\eta_v(t)$ 与 $\eta_\theta(t)$ 则建模了诸如道路状况变化、驾驶员行为不确定性以及交通流波动等外部扰动因素.

为了进一步反映移动性对车载通信系统的影响，本文引入车辆之间的相对运动模型. 具体而言，车辆 k 与车辆 j 之间的相对速度表达如式(2)所示：

$$\begin{aligned} v_{k,j}(t)=&\sqrt{\left(v_k(t)\cos\theta_k(t)-v_j(t)\cos\theta_j(t)\right)^2} \\ &+\sqrt{\left(v_k(t)\sin\theta_k(t)-v_j(t)\sin\theta_j(t)\right)^2} \end{aligned} \tag{2}$$

该相对速度在车辆通信环境中起着关键作用，尤其是在V2X场景中，高速移动导致的多普勒频移显著影响信道的时变特性. 例如，在高速公路场景中，车辆反向行驶会引起剧烈的信道变化；而在城市交叉口，由于存在障碍物和频繁的加减速行为，通信信道的动态性更为复杂. 上述因素均对RIS辅助的VEC系统提出了更高的资源调度与信号处理能力要求，亟需采用自适应的资源分配策略与鲁棒的信号处理技术，以确保通信的可靠性与连续性.

2.3 3D几何信道模型

在RIS辅助VEC系统中，通信通过直接车到基础设施通信(Vehicle-to Infrastructure，V2I)链路和RIS辅助链路进行，其中底层信道特性显著影响系统性能. 这些信道受到环境的三维(Three-Dimensional，3D)几何信

道的影响，包括路径损耗、衰落效应和波束形成特性，所提出的三维几何信道模型如图2所示.

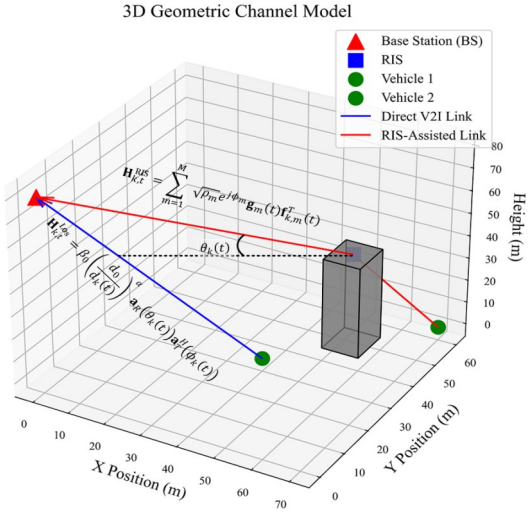

图2　3D几何信道模型

直接V2I链路由两个主要部分组成：车辆到BS链路和RIS到BS链路. 具体地说，车辆k和BS在时隙t的视距(Line-of-Sight，LoS)链路用信道矩阵$\boldsymbol{H}_{k,t}^{\mathrm{LoS}}$表示. 该矩阵模拟了车辆到BS路径和RIS到BS路径的传播特性，不包括RIS引入的反射分量，后者主要遵循LoS传播模型. 该直接链路的路径损耗如式(3)所示：

$$\boldsymbol{H}_{k,t}^{\mathrm{LoS}}=\beta_0\left(\frac{d_0}{d_k(t)}\right)^{\alpha}\mathbf{a}_R(\theta_k(t))\mathbf{a}_T^H(\phi_k(t)) \tag{3}$$

其中，$\beta_0=10^{-3}$为距离$d_0=1$米处的参考路径损耗，$\alpha=2.5$为路径损耗指数，$d_k(t)$表示t时刻车辆k与BS之间的距离，到达BS的仰角和RIS的相移分别用$\theta_k(t)$和$\phi_k(t)$表示. $\mathbf{a}_T^H(\phi_k(t))$为阵列响应向量的共轭转置，表征天线阵方向图的BS阵列响应向量由式(4)给出：

$$\mathbf{a}_R(\theta_k(t))=\left[1,\mathrm{e}^{j\frac{2\pi d}{\lambda}\sin(\theta_k(t))},\cdots,\mathrm{e}^{j(N_R-1)\frac{2\pi d}{\lambda}\sin(\theta_k(t))}\right]^T \tag{4}$$

其中，λ为信号波长，d为BS阵列中相邻天线间距，N_R为BS阵列中单元总数.

为了考虑实际场景中的信道不确定性，本文进一步引入信道估计误差建模机制. 假设观测信道为$\hat{\boldsymbol{H}}_{k,t}^{\mathrm{LoS}}=\boldsymbol{H}_{k,t}^{\mathrm{LoS}}+\epsilon_{k,t}$，其中误差项$\epsilon_{k,t}$服从均值为0、方差为$\sigma_\epsilon^2$的高斯分布$\epsilon_{k,t}\sim\mathcal{N}(0,\sigma_\epsilon^2)$，用以模拟信道估计噪声. 该建模方式可用于评估信道估计误差对资源分配策略的鲁棒性影响.

除了直连链路外，RIS辅助路径通过信号反射增强了信道传播特性. 对应的信道矩阵如式(5)所示：

$$\boldsymbol{H}_{k,t}^{\mathrm{RIS}}=\sum_{m=1}^{M}\sqrt{\rho_m}\,\mathrm{e}^{j\phi_m}g_m(t)\mathbf{f}_{k,m}^T(t) \tag{5}$$

其中，ρ_m表示第m个RIS单元的反射系数，ϕ_m为其施加的相位偏移，$g_m(t)$表示基站与第m个RIS单元之间的LoS信道增益，$\mathbf{f}_{k,m}^T(t)$表征车辆k与RIS之间的Rician衰落信道. RIS通过动态调控各反射单元的相位，有效提升信号接收功率并抑制干扰.

在城市场景中，由于高楼、桥梁等物理遮挡，NLoS链路较为常见. 本文在仿真环境中引入遮挡概率模型：车辆与RIS之间的链路在时隙t时为NLoS的概率为P_{NLoS}，其信道增益将附加遮挡衰减因子$C<1$. 在算法状态空间中，NLoS事件以二值指示变量编码，提升策略对遮挡条件的响应能力.

系统的通信性能主要由每辆车的信干噪比(Signal-to-Interference-plus-Noise Ratio，SINR)决定，车辆k在时刻t的SINR如式(6)所示：

$$\gamma_k(t)=\frac{P_k^{\mathrm{off}}(t)\left\|\boldsymbol{H}_{(k,t)}^{\mathrm{LoS}}+\boldsymbol{H}_{(k,t)}^{\mathrm{RIS}}\right\|_F^2}{\sigma^2+\sum_{j\neq k}P_j^{\mathrm{off}}(t)\left\|\boldsymbol{H}_{(j,t)}^{\mathrm{LoS}}\right\|_F^2} \tag{6}$$

其中，$P_k^{\mathrm{off}}(t)$是车辆k的任务卸载发射功率，$\boldsymbol{H}_{(k,t)}^{\mathrm{LoS}}$与$\boldsymbol{H}_{(k,t)}^{\mathrm{RIS}}$分别为直连信道和RIS辅助信道矩阵，$\sigma^2$为背景噪声功率. 分母中的求和项表示其他车辆产生的累积干扰. 此公式中，分子代表期望接收信号功率，而分母则综合了系统干扰与噪声影响.

2.4　任务动态、AoI、能耗建模

车辆计算任务的动态由任务队列的更新过程决定，任务队列的演化模型如式(7)所示：

$$q_k(t+1)=\left[q_k(t)-\underbrace{f_k^{\mathrm{loc}}(t)\Delta t}_{\text{Local}}-\underbrace{\frac{C_k(t)\Delta t}{D_k}}_{\text{Offloaded}}\right]^+ +\lambda_k(t) \tag{7}$$

其中，$f_k^{\mathrm{loc}}(t)$是车辆本地的计算频率，D_k是任务的数据量，$C_k(t)$为可用通信容量，定义如式(8)所示：

$$C_k(t)=B\log_2\left(1+\gamma_k(t)\right) \tag{8}$$

其中，SINR$\gamma_k(t)$直接决定了数据传输速率和卸载效率，B表示系统下行链路的总可用带宽，用于车辆用户的任务卸载与数据传输. 在仿真部分，我们将其数值设定为B = 20 MHz，任务队列长度$q_k(t)$依据本地计算量、卸载处理量和新任务到达率$\lambda_k(t)$进行更新.

AoI是一种衡量信息新鲜度的性能指标，其核心作用在于描述“接收端所持有的最新状态信息距离真实生成时刻的滞后程度”. 在车联网等动态系统中，用于量化车辆网络中信息的新鲜度，从而为资源调度与能效优化提供决策依据. 车辆k在$t+1$时刻的AoI更新如式(9)所示：

$$\Delta_k(t+1)=\begin{cases}\Delta_{\min}, & \text{if } q_k^{\mathrm{off}}(t)\geq q_k^{\mathrm{th}}\\ \Delta_k(t)+\dfrac{q_k(t)}{q_k^{\max}}, & \text{otherwise.}\end{cases} \tag{9}$$

其中，$q_k^{\max}$表示最大队列长度，q_k^{th}表示卸载阈值. 该模型有效表征了车载通信中信息更新的时效性. 车辆k处的总能耗由三个主要部分组成，如式(10)所示：

$$E_k(t)=\underbrace{\kappa[f_k^{\text{loc}}(t)]^3\Delta t}_{\text{Computing}}+\underbrace{\frac{P_k^{\text{off}}(t)\Delta t}{\eta_{\text{PA}}}}_{\text{Transmission}}+P_{\text{cir}}\Delta t \tag{10}$$

其中，$P_k^{\text{off}}(t)$为卸载功率，P_{cir}为电路功耗. 假设功率放大器效率$\eta_{PA}=35\%$，则传输能量受信号强度和功率转换效率的影响，而计算能量则取决于任务工作量和处理频率.

2.5 多目标优化问题

在RIS辅助的车载边缘计算资源分配背景下，本文将问题建模为一个多目标优化任务，旨在实现AoI与能量消耗之间的最优权衡. 该问题具有以下显著特征：首先，由于同时涉及连续决策变量(如任务卸载功率与本地计算频率)以及离散决策变量(如配置选择)，其数学形式属于非凸混合整数规划，求解难度较高；其次，系统中包含大量车辆、任务及配置选项，导致动作空间维度极高，从而显著增加了计算复杂度；最后，在分布式VEC环境下，各智能体仅能获取自身的局部状态信息，无法全面感知系统全局状态，因此问题呈现出典型的部分可观测特性，需通过分布式或去中心化的决策机制加以应对.

该优化任务的目标是在给定时间段内，最小化所有车辆在所有时隙上的AoI与能量消耗的加权和. 其优化目标函数可表示为：最小化期望的系统总成本，即每个车辆在每个时隙的AoI与能量消耗的加权组合. 该问题受到以下约束条件限制，如式(11)所示：

$$\mathrm{P}_0:\Phi,\{P_k^{\text{off}}(t),f_k^{\text{loc}}(t)\}\ \min\ \mathrm{E}\left[\sum_{t=1}^{T}\sum_{k=1}^{K}\left(\alpha\Delta_k(t)+\beta E_k(t)\right)\right] \tag{11a}$$

$$\text{约束：}\ 0\leqslant P_k^{\text{off}}(t)\leqslant P_{\max} \tag{11b}$$

$$0\leqslant f_k^{\text{loc}}(t)\leqslant f_{\max} \tag{11c}$$

$$\phi_m\in\Phi \tag{11d}$$

$$q_k(t)\leqslant q_k^{\max} \tag{11e}$$

其中，约束(11b)将卸载功率$P_k^{\text{off}}(t)$限制在0到$P_{\max}$之间；约束(11c)将本地计算能力$f_k^{\text{loc}}(t)$限制在0到$f_{\max}$范围内；约束(11d)确保ϕ_m属于可行集合Φ；而约束(11e)则限制任务队列长度$q_k(t)$不超过最大允许长度$q_k^{\max}$.

该问题的提出，为RIS辅助的车载边缘计算系统提供了一个全面的资源优化框架，在明确考虑AoI与能效权衡的基础上进行调度. 然而，由于该问题具备非凸性、高维动作空间等复杂特点，传统的优化方法在实际应用中难以在有限时间内获得高效解. 因此，本文在下一节中提出了一种基于混合深度强化学习的优化框架，将原始问题分解为若干可管理的子任务，以有效应对上述挑战.

3 混合DRL优化框架

3.1 层次分解

为应对RIS辅助车载边缘计算系统中的多目标优化问题，本文提出了混合DRL框架并采用分层解耦策略. 具体而言，问题被划分为两层：上层优化RIS相位配置，下层以联邦方式处理分布式功率分配任务. 该设计有效缓解了高维动作空间的计算复杂性，使各子系统能够独立而协调地优化. 图3展示了框架结构，各模块说明如下：

图中展示了所提出的分层结构：RIS相位调控由TD3模型完成，而功率分配则通过一种FMADDPG模型实现.

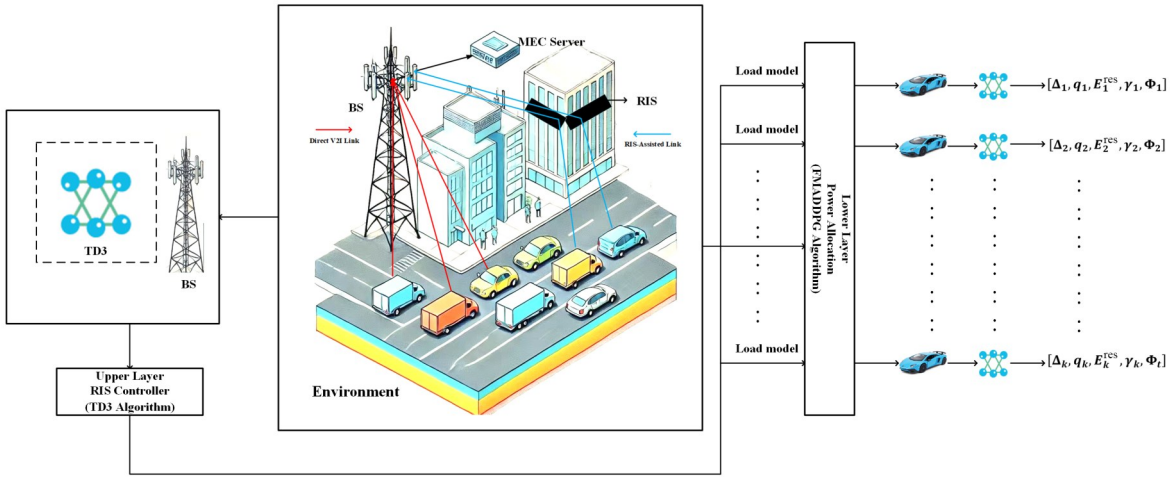

图3 拟议的框架. 该框架包含一个用于RIS phase-shift优化的TD3模型和一个用于功率分配的FMADDPG模型

在上层,RIS相位优化的目标是最大化加权系统容量总和,同时保持相邻时隙间相位配置的时域连续性.其优化问题可表述如式(12)所示:

$$\mathrm{P}_1:\ \max_{\Phi}\sum_{k=1}^{K}\omega_k C_k(\Phi)-\eta\,\|\,\Phi-\Phi_{\mathrm{prev}}\|_F^2 \qquad (12)$$

其中,$C_k(\Phi)$表示车辆k在当前相位配置Φ下的通信容量,ω_k为其对应的权重,Φ_{prev}表示前一时隙的RIS相位矩阵,η则为相位变化平滑度的调节参数,用以确保相位调整过程平稳,避免系统突变.

基站作为智能体与环境交互,并基于TD3算法训练RIS相位优化模型.TD3适用于该层级,因为其在连续动作空间中具有良好的稳定性和收敛性能,尤其适合对RIS相位矩阵这类实数型变量进行精细控制.双Q网络结构与延迟更新机制可有效缓解策略过估计问题,提高学习效率.

在下层,优化目标是对卸载功率和本地计算频率进行合理分配,以最小化任务时延(用AoI衡量)与能量消耗的加权成本,其优化问题表述如式(13)所示:

$$\mathrm{P}_2:\ \min_{\{P_k^{\mathrm{off}},f_k^{\mathrm{loc}}\}}\mathrm{E}[\alpha\Delta_k(t)+\beta E_k(t)] \qquad (13)$$

其中,$\Delta_k(t)$表示车辆k在时刻t的AoI,衡量任务N从生成到成功接收之间的时延;$E_k(t)$表示车辆在本地计算和任务卸载过程中产生的能耗;α与β是控制AoI与能耗之间权衡关系的加权系数.

需要进一步说明的是,原始问题P_0与分层子问题P_1和P_2之间并非松散拆分,而是具有严格的数学对应关系.在给定Φ的情况下,P_0在卸载功率和本地计算频率上的优化目标完全等价于P_2,因此P_2可视为P_0的条件化子问题.由此,定义值函数$V(\Phi)=\min_{\{P_k^{\mathrm{off}}(t),f_k^{\mathrm{loc}}(t)\}}\mathrm{E}[\sum_{t=1}^{T}\sum_{k=1}^{K}(\alpha\Delta_k(t)+\beta E_k(t))]$在约束条件(11b)~(11e)下,$V(\Phi)$与$\mathrm{P}_2$的定义完全一致,因此,原始问题可以改写为$\min_{\Phi}V(\Phi)$,即上层对$\Phi$的优化与下层对功率和频率的最优解相结合.由于系统代价随着车辆可达容量$C_k(\Phi)$的增加而单调减小,可以证明提升RIS相位配置所带来的链路容量将降低整体代价函数值.基于这一性质,本文在上层引入加权容量最大化作为代理目标,并附加平滑正则项以保证相位调整的时域连续性.当权重设计合理时,代理目标的优化方向与最小化$V(\Phi)$的方向一致,从而确保了P_1与P_0在保序性和梯度意义上的一致性.

考虑到下层任务的分布性、动态性及隐私性,本文引入FMADDPG,每个VU在本地训练策略网络,并周期性聚合模型参数,以减少通信开销并增强对异构数据和隐私的支持.为提升上下层协同效率,本文设计了联合预训练机制:上层TD3控制器先生成代表性RIS相位轨迹,并将其嵌入下层FMADDPG状态空间,用于策略初始化和早期收敛,从而建立上下层软耦合,增强训练稳定性与系统泛化能力.尽管上下层优化目标不完全一致,但方向一致,通过策略嵌入与更新实现相互促进,形成可协同的分层多目标优化结构.

通过该分层式DRL框架,系统能够在通信效率与资源利用之间取得良好平衡,同时显著降低了计算复杂度.RIS相位控制与功率分配的独立优化策略,使得系统能够实时适应不断变化的网络环境,提高了对任务负载波动和通信条件变化的鲁棒性.

相比于传统的单层强化学习方法,该框架将全局配置与本地决策解耦,结合TD3与FMADDPG各自的优势,形成"集中控制 + 分布优化"的协同结构,更契合实际部署中对灵活性、响应性与稳定性的综合需求.

在构建好分层结构后,下一步将重点放在上层:RIS相位配置的智能调整.为此,本文引入了一种基于TD3算法的控制器,用于处理连续动作空间下的相位优化问题.

3.2　基于TD3的RIS控制器

为了优化V2X网络中RIS的相位设置,本文引入了一种基于TD3的控制器.作为一种DRL方法,TD3在处理高维连续动作空间方面表现出色,同时能够有效缓解Q值的高估问题.通过引入双Q网络结构和延迟的策略更新机制,TD3提升了学习的稳定性,并增强了动态车联网通信环境中的决策能力.其状态、动作及奖励函数定义如下:

状态(State):控制器的决策依赖于包含RIS相位优化关键信息的系统状态.在每一个时刻t,状态包含前一时刻的RIS相位配置Φ_{prev}、每辆车的历史CSI$H_{k,t-1}$、车辆当前位置v_k以及其到达角(Angle of Arrival,AoA)θ_k.因此,整体状态可表示为:$s_t^{\mathrm{RIS}}=\left[\Phi_{\mathrm{prev}},\{H_{k,t-1}\},\{v_k\},\{\theta_k\}\right]$

动作(Action):在当前状态下,控制器决定一个连续的相位调整动作集$\varphi_m^t\in\mathrm{R}$,其中每一个元素对应RIS中的第m个单元在时刻t的相位调整.目标是在保证相位变化平滑的同时最大化系统性能.所有相位调整组成的动作向量为:$a_t^{\mathrm{RIS}}=[\varphi_1^t,\varphi_2^t,\cdots,\varphi_m^t]\in\mathrm{R}^M$

奖励(Reward):为引导学习过程,奖励函数同时考虑通信容量和相位调整的平稳性.在时刻t的即时奖励定义如式(14)所示:

$$r_t^{\mathrm{RIS}}=\sum_{k=1}^{K}\omega_k C_k(t)-\eta\,\|\,\Phi_t-\Phi_{t-1}\|_2^2 \qquad (14)$$

其中,$C_k(t)$表示第k辆车在时刻t的可达容量,ω_k用于体现不同用户的优先级.第二项惩罚了相位变化过大的情况,其中Φ_t和Φ_{t-1}分别为当前和前一时刻的RIS相位向量,η控制容量最大化与相位平稳性的权衡.

为了加速整体系统的训练收敛过程，本文在TD3策略网络训练初期引入预训练阶段. 该阶段通过与环境的初步交互生成一批代表性的RIS相位配置轨迹，并将其记录为“相位先验数据”. 后续该数据将用于下层FMADDPG模型的预训练阶段，使其能够在策略初始化时充分感知上层优化趋势，从而提升其收敛效率与策略质量. 该机制为上下层策略提供了“软协同”的先验联系，是一种轻量但有效的策略融合方式.

通过不断优化策略，基于TD3的控制器可以实时适应通信环境的变化，学习出最优的RIS相位控制方案，从而提升频谱效率与网络可靠性. 其在连续动作优化与策略学习方面的强大能力，为RIS辅助的车联网通信系统提供了高效的资源管理手段.

3.3 TD3算法实现

在高维参数空间的复杂强化学习环境中，深度神经网络在建模系统交互方面至关重要. TD3算法作为DDPG的改进版本，通过双Q函数估计、延迟策略更新与目标网络同步三项机制，显著提升了连续控制任务的收敛性能. 本研究利用TD3确定RIS单元的最优相位配置，有效应对连续参数优化难题. 该框架由演员（Actor）和评论家（Critic）组成，分别负责策略映射与动作价值评估，并通过延迟同步机制提升策略稳定性.

在训练初始阶段，Actor网络θ_μ和Critic网络θ_Q的参数被初始化，同时目标网络$\theta_{\mu'}$、$\theta_{Q'}$被赋予相同的初始值. 在每个时间步中，智能体根据当前策略$\mu(s_t;\ \theta_\mu)$选择一个动作a_t，并叠加探索噪声Δ_n，随后该动作被作用于环境，系统返回即时奖励及下一个状态s_{t+1}[4,22,24]. 该转移过程被存入经验回放池中，用于后续学习. 此外，为实现上下层策略之间的软协同，本文在预训练阶段引入RIS相位轨迹池P，用于记录智能体在交互过程中生成的状态－动作对（state－action pair），即(s_t, a_t). 这些轨迹并不直接参与TD3策略训练过程，但会作为先验信息嵌入下层FMADDPG智能体的状态表示中，从而提升下层策略对上层行为的感知能力，加快整体收敛速度. 当经验池中积累了足够的样本后，系统从中随机采样小批量数据进行训练. Critic网络首先通Bellman方程计算目标Q值y_i如式（15）所示：

$$y_i = r_i + \gamma \min_{i=1,2} Q_i'(s_{i+1}, \mu'(s_{i+1};\ \theta_{\mu'});\ \theta_{Q'}) \quad (15)$$

其中γ为折扣因子，取两个Q网络中的最小值以抑制过估计偏差.

接下来，Critic网络通过最小化以下损失函数进行更新，如式（16）所示：

$$L(\theta_Q) = \frac{1}{I}\sum_{i=1}^{I}(y_i - Q(s_i, a_i;\ \theta_Q))^2 \quad (16)$$

在更新Critic网络后，Actor网络基于策略梯度法进行优化，目标是最大化Q值的期望. 具体梯度计算如式（17）所示：

$$\nabla_{\theta_\mu} J \approx \frac{1}{I}\sum_{i=1}^{I} \nabla_a Q(s_i, a;\ \theta_Q)|_{s=s_i, a=\mu(s_i)} \nabla_{\theta_\mu}\mu(s_i;\ \theta_\mu) \quad (17)$$

为进一步提高训练稳定性，TD3采用软更新规则来更新目标网络，而非直接复制主网络参数，其更新规则如式（18）、式（19）所示：

$$\theta_{Q'} \leftarrow \tau\theta_Q + (1-\tau)\theta_{Q'} \quad (18)$$

$$\theta_{\mu'} \leftarrow \tau\theta_\mu + (1-\tau)\theta_{\mu'} \quad (19)$$

其中，τ是一个小的正数（通常设置为τ=0.005），用于控制目标网络的更新速率.

整个训练过程将在预设轮数或策略收敛条件下停止. 在每一轮中，智能体不断与环境交互，采集经验，并按照上述步骤更新网络参数. 经过多次迭代后，策略将逐步收敛至最优或近似最优解.

TD3通过引入双Q网络、延迟策略更新与软目标网络更新等机制，大幅增强了学习过程的稳定性，有效降低了Q值方差和过估计问题，进而提升了连续动作空间下的决策可靠性. 在如RIS相位调整这类复杂优化任务中，其性能表现尤为突出. 相关伪代码如算法1所示.

算法1 TD3算法优化RIS相位控制策略

输入：学习率α，折扣因子γ，软更新系数τ，actor网络θ_μ，两个critic网络θ_{Q_1}、θ_{Q_2}，目标网络$\theta_{\mu'}$、$\theta_{Q_1'}$、$\theta_{Q_2'}$

输出：最优RIS相位控制策略μ^*，将观察到的系统状态映射到接近最优的相位配置Φ_t

1. 初始化演员和评论家网络参数为随机值;
2. 设置目标网络参数:$\theta_{\mu'} \leftarrow \theta_\mu, \theta_{Q_1'} \leftarrow \theta_{Q_1}, \theta_{Q_2'} \leftarrow \theta_{Q_2}$;
3. 初始化经验回放池 D;
4. 初始化 RIS 相位轨迹池 P
5. 执行预训练阶段,用于生成RIS轨迹
6. **For** 每轮训练 episode **do**
7. 重置环境,获得初始状态 s_1;
8. **For** 每一时刻 t **do**
9. 生成RIS相位控制动作 $a_t = \mu(s_t;\theta_\mu) + N_t$;
10. 执行动作 a_t,获得奖励 r_t 及下一状态 s_{t+1};
11. 将转移元组 (s_t, a_t, r_t, s_{t+1}) 存入 D;
12. 将当前轨迹(s_t, a_t)存入轨迹池P,供下层FMADDPG状态嵌入
13. **If** D 中样本数量大于 I **then**
14. 从 D 中随机采样 I 个样本构成小批量;
15. 利用式(15)计算目标 Q 值;
16. 根据式(16)最小化损失函数,更新评论家网络;
17. **If** 达到策略更新频率 **then**
18. 根据式(17)更新演员网络;
19. 使用式(18)和式(19)更新目标网络;
20. **End if**
21. **End if**
22. **End for**
23. **End for**

3.4　FMADDPG功率分配器

尽管RIS控制器提升了物理层性能,但系统的高效功率分配仍是关键问题. 为此,本文引入了一种基于FMADDPG的功率分配器,用于分布式车载智能体间的协同功率优化. 该分配器在不共享本地数据的前提下,通过DRL实现对异构环境和操作约束的动态功率调整,从而维持系统的高效通信性能.

在每次迭代t时,智能体k的本地模型参数θ_t^k将加入高斯噪声项$N(0,\sigma_{\mathrm{DP}}^2)$以增强训练过程中的探索能力. 随后,全局模型参数$\theta_{t+1}^G$通过对所有参与智能体的本地更新进行平均计算获得,以实现模型的逐步优化,如式(20)所示:

$$\theta_{t+1}^G=\frac{1}{K}\sum_{k=1}^{K}\theta_t^k+N(0,\sigma_{\mathrm{DP}}^2 I) \tag{20}$$

其中,K表示联邦网络中的智能体总数,σ_{DP}^2为用于探索的本地更新噪声方差.

这种基于联邦学习的方法确保功率分配模型能够持续适应网络状态变化并保护用户隐私. 通过多智能体间的协同优化,FMADDPG在不共享原始数据的情况下实现了动态环境下的高效资源分配,从而提升通信效率并增强系统的安全性与可扩展性. 该模型中的状态、动作和奖励函数定义如下:

状态(State):在每个时间步t,智能体k的状态向量包含若干关键特征,这些特征影响其决策过程,包括功率变化量Δ_k、信道质量q_k、剩余能量E_k^{res},反映其对未来奖励敏感度的折扣因子γ_k,以及上层预训练阶段采集的RIS相位配置轨迹Φ_t,该轨迹来自阶段性构建的轨迹池P. 整体状态可表示为$s_k^t=[\Delta_k,q_k,E_k^{\mathrm{res}},\gamma_k,\Phi_t]$. 该设计在不破坏上下层解耦结构的基础上,实现了策略间的软耦合,提升了学习初期的协同效率.

动作(Action):每个时间步t,智能体选择的动作a_k^t包括两个部分:功率卸载因子$P_k^{\mathrm{off}}(t)$与本地计算因子$f_k^{\mathrm{loc}}(t)$,即$a_k^t=[P_k^{\mathrm{off}}(t),f_k^{\mathrm{loc}}(t)]$.

奖励(Reward):智能体k的奖励函数r_k^t综合考虑了能效与网络吞吐量等多个因素. 如式(21)所示:

$$r_k^t=-\left[\alpha\frac{\Delta_k}{\Delta_{\max}}+\beta\frac{E_k}{E_{\max}}+\lambda\frac{q_k}{q_{\max}}\right] \tag{21}$$

其中,α、β、λ为缩放因子,用于权衡系统的不同目标,$\Delta_{\max}$、$E_{\max}$、$q_{\max}$分别表示AoI、能耗及队列长度的最大值. 该奖励函数设计为负加权和,用以衡量每次决策所带来的代价.

为确保各性能指标在训练中的量纲统一性与数值平衡性,公式中引入了归一化项$\Delta_{\max}$、$E_{\max}$和$q_{\max}$. 实际训练中,这些最大值通过预热阶段统计获取,并在整个训练过程中保持固定,以提升训练稳定性并抑制奖励震荡. 缩放因子α、β和λ的设置参考经验调优与性能敏感性分析,其中$\alpha=1.5$、$\beta=0.1$、$\lambda=0.1$能在多数训练配置下获得良好的收敛性与收敛速度. 该配置在多轮对比实验中展现出良好的鲁棒性与泛化性能,相关内容已在实验部分进行了讨论.

3.5　FMADDPG算法概念

FMADDPG算法是一种面向多智能体环境的强化学习框架,将联邦学习与MADDPG相结合,实现了在保持数据隐私的前提下的协同模型更新. 该去中心化学习范式特别适用于RIS辅助VEC网络中的功率分配问题,在此类网络中,多个VU在动态、部分可观测环境下运行. FMADDPG采用本地策略学习与全局优化相结合的混合架构:每个智能体维护独立的actor-critic网络以学习个性化功率分配策略,而全局critic网络通过聚合多智能体经验提升系统协调能力. 为兼顾实时适应性与收敛稳定性,FMADDPG引入联邦模型聚合机制,即本地智能体周期性上传参数更新,全局模型基于聚合结果优化策略,从而实现高效的可扩展性与数据安全性.

训练过程遵循迭代式强化学习框架. 在每个时间步,智能体k感知其当前局部状态s_k^t,其中包括CSI、AoI及能量约束等. 智能体依据其actor策略选择动作a_k^t,并加入随机探索项以适应动态环境变化. 系统随后反馈奖励r_k^t,并将其转移至下一个状态s_k^{t+1}. 该经验转移元组被存入回放记忆库中,后续以小批量方式采样用于提升训练效率.

策略优化过程分为两个阶段. 在第一阶段,本地评论家网络$Q_k(s,a;\theta_{Q_k})$通过最小化损失函数进行更新,如式(22)所示:

$$L(\theta_{Q_k})=\frac{1}{I}\sum_i\left(y_i^k-Q_k(s_i,a_i;\theta_{Q_k})\right)^2 \tag{22}$$

其中目标值为y_i^k如式(23)所示:

$$y_i^k=r_i^k+\gamma Q_g\left(s_{i+1},\pi_k(s_{i+1};\theta_{\pi_k});\theta_{Q_g}\right) \tag{23}$$

同时,全局评论家网络 按式(24)更新:

$$L(\theta_{Q_g})=\frac{1}{I}\sum_i\left(y_i^g-Q_g(s_i,a_i;\theta_{Q_g})\right)^2 \tag{24}$$

其中目标值y_i^g如式(25)所示:

$$y_i^g=r_i^g+\gamma Q_g\left(s_{i+1},\pi_g(s_{i+1};\theta_{\pi_g});\theta_{Q_g}\right) \tag{25}$$

在第二阶段,智能体应用确定性策略梯度对actor策略进行优化,如式(26)所示:

$$\nabla_{\theta_{\pi_k}}J_k\approx\frac{1}{I}\sum_i\nabla_{a_k}Q_k(s_i,a_i;\theta_{Q_k})\nabla_{\theta_{\pi_k}}\pi_k(s_i;\theta_{\pi_k}) \tag{26}$$

为了提高训练稳定性并防止策略剧烈波动,FMADDPG采用了软更新机制更新目标网络参数,如式(27)、(28)和(29)所示:

$$\theta_{Q_g'}\leftarrow\tau\theta_{Q_g}+(1-\tau)\theta_{Q_g'} \tag{27}$$

$$\theta_{Q_k'} \leftarrow \tau\theta_{Q_k'} + (1-\tau)\theta_{Q_k'} \tag{28}$$

$$\theta_{\pi_g'} \leftarrow \tau\theta_{\pi_g'} + (1-\tau)\theta_{\pi_g'} \tag{29}$$

其中τ为控制参数平滑更新程度的小常数.

通过本地actor-critic更新、联邦学习策略以及全局critic的协同机制,FMADDPG实现了多智能体间的高效资源协调与隐私保护. 其联邦聚合策略进一步提升了大规模VEC网络中的可扩展性. 相关的伪代码如算法2所示.

算法2 基于FMADDPG的联邦功率分配策略优化

输入:学习率α,折扣因子γ,软更新系数τ,每个智能体k的actor与critic网络,全局critic网络θ_{Q_g},本地critic网络θ_{Q_k},RIS相位轨迹池P

输出:优化后的功率分配策略

1. 初始化全局评论家网络θ_{Q_g}及其目标网络$\theta_{Q_g'}$;
2. 初始化每个智能体k的本地actor-critic网络;
3. **For** 每轮训练 episode **do**
4. 重置环境;
5. **For** 每一时刻t **do**
6. 为每个智能体k获取状态s_k^t;
7. 从轨迹池P中提取当前RIS相位信息Φ_t;
8. 每个智能体依据actor策略选择动作a_t^k;
9. 观察每个智能体的奖励r_k^t及下一个状态;
10. 将$(s_t, a_t, r_k^t, s_{t+1})$存入回放缓冲区;
11. **If** 回放缓冲区大小 > I **then**
12. 从D中随机采样一个小批量;
13. 使用式(24)和式(25)更新全局评论家网络;
14. 使用式(22)和式(23)更新本地评论家网络;
15. 依据式(26)使用策略梯度更新本地演员网络;
16. 根据式(27)、式(28)和式(29)执行目标网络的软更新;
17. **End if**
18. **End for**
19. **End for**

3.6 理论收敛性分析

为确保所提出的FMADDPG算法在多智能体分布式环境中的稳定性与收敛性,本文在强化学习经典理论基础上,结合联邦聚合机制和轨迹嵌入设计,建立了理论收敛分析框架. 具体分析如下:

假设1(状态与动作空间有界):每个智能体的局部状态空间与动作空间均为有界紧集;

假设2(奖励函数平滑):每个智能体的即时奖励函数对状态与动作变量均为Lipschitz连续;

假设3(策略光滑性):Actor与Critic网络为连续可导函数,采用满足Robbins-Monro条件的递减学习率,例如设学习率随迭代轮次递减:$\eta_t = 1/t^{\rho}$,其中,常数ρ控制学习率的衰减速度,取值范围为$(0.5, 1]$;

假设4(联邦聚合稳定性):各本地智能体在上传网络参数后,全局Critic模型的聚合误差在时间t满足:$\| \theta_t^G - \frac{1}{K}\sum_{k=1}^{K}\theta_t^k \| \leqslant \epsilon_t$,且随着训练进行$\epsilon_t \to 0$;

假设5(轨迹嵌入有界性):上层TD3预训练阶段生成的RIS轨迹池为有界集合.

在上述条件下,FMADDPG训练过程中每一步的策略梯度估计在期望意义下趋于零,即:

$$\lim_{T\to\infty}\frac{1}{T}\mathrm{E}\left[\| \nabla J(\theta_t)\|^2\right] = 0 \tag{30}$$

当训练步数趋于无穷时,策略网络参数的更新方向将逐渐收敛于某一稳定点,从而实现近似最优策略的收敛. 为进一步验证算法性能,下一节将结合典型车联网场景展开仿真分析.

4 仿真结果

在本节中,本文通过仿真验证所提出混合DRL框架在RIS辅助VEC系统中的有效性. 实验基于真实城市交通场景,采用三维几何信道模型,评估方法在降低AoI与提升能效方面的性能. 系统由一台集成边缘服务器的BS和K=*50*辆车辆组成,基站附近部署了M=100个RIS单元,每个单元的相位控制精度为b=3 bits. 边缘服务器计算能力为$\mathrm{F}_{\max}$=25 TFLOPS,可高效支持任务卸载. 通信信道包括LoS与RIS辅助链路,考虑路径损耗、衰落及波束赋形. 车辆任务到达服从泊松分布,到达率Λ_K=10个任务/秒,计算需求在1~10 GFLOPs均匀分布. 车辆运动遵循第2.2节运动学模型,初始速度为30 km/h,加速度与角速度噪声服从高斯分布. RIS相位由TD3算法优化,车辆发射功率通过FMADDPG动态分配,最大发射功率$P_{\max}$=23 dBm,本地CPU频率上限$f_{\max}$=2.5 GHz.

为进一步评估所提方法在不确定环境下的适应性,本文在上述仿真环境中引入了信道估计噪声$\sigma_\epsilon^2 \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$与链路遮挡概率$P_{\mathrm{NLoS}} \in \{0, 0.2, 0.4, 0.6\}$两个扰动因素,并进行了鲁棒性性能测试. 固定非视距条件下附加衰减因子C=0.6,遮挡事件以二值指示变量形式嵌入状态空间以提升策略感知能力. 每个遮挡概率对应一个仿真场景,用于模拟不同环境复杂度下的链路遮挡程度. 例如,在P_{NLoS}=0.4的情形下,表示车辆与RIS之间的链路在40%的时间内处于非视距状态,从而评估算法在中等遮挡环境下的性能鲁棒性. 仿真与神经网络参数如表1和表2所示.

为了全面评估所提算法在多用户、复杂信道、高动态任务条件下的性能表现,本文选取了三种具有代表性的对比方法,涵盖强化学习领域的先进算法与传统优化方法. 具体而言,SAC基于最大熵理论,具备良好的策略探索性与训练稳定性,适用于与本文上层RIS相位优化任务结构相近的高维控制问题;QMIX是一种分解式多智能体强化学习算法,适用于状态共享但动作解耦的任务,能验证本文下层FMADDPG联邦机制在分

布式功率分配中的策略协调性优势;BCD代表经典非强化学习类优化方法,广泛应用于无线通信与MEC任务中,用作解析优化基准以反映工程可达下界性能. 此外,还设置了两个对照基准:随机RIS+随机资源分配(Resource Allocation,RA),对RIS相位矩阵和车辆资源分配均采用随机配置;无RIS+随机资源分配,完全去除RIS组件,仅对频谱与功率采用随机资源分配.

表1　仿真参数

参数	符号	数值
车辆数量	k	50
智能反射面单元数量	M	100
RIS相位控制精度	b	3 bits
边缘服务器计算能力	$F_{\max}$	25 TFLOPS
路径损耗因子(α)	α	2.5
参考路径损耗(β_0)	β_0	10^{-3}
任务到达率(Λ)	Λ	10 tasks/second
任务计算需求		1~10 GFLOPs
初始车速		30 km/h
加速度/角速度噪声	$\sigma_v^2,\sigma_\theta^2$	$\sigma_v^2=1,\sigma_\theta^2=0.1$
最大发射功率	$P_{\max}$	23 dBm
本地计算频率上限	$f_{\max}$	2.5 GHz
功率放大器效率	η_{PA}	35%
噪声功率	σ^2	10^{-10} W
任务队列最大长度	$q_k^{\max}$	100 tasks
任务队列阈值	q_k^{th}	50 tasks
最大信息年龄	$\Delta_{\max}$	100 ms
最大能耗限制	$E_{\max}$	1000 J
信道估计噪声	σ_ϵ^2	{0, 0.1, 0.2, 0.3, 0.4, 0.5}
遮挡概率	P_{NloS}	{0, 0.2, 0.4, 0.6}
附加遮挡衰减因子	C	0.6
下行链路的总可用带宽	B	20 MHz

表2　神经网络参数

参数	数值
相位控制正则化系数	0.1
学习率衰减系数	0.5
目标网络更新系数	0.005
折扣因子	0.99
Actor 网络学习率	0.001
Critic 网络学习率	0.001
经验回放缓冲区大小	1 000 000
批处理大小(Batch Size)	256
策略更新频率	每 2 步更新一次
目标网络更新频率	每 2 步
总训练轮数(Episodes 数)	1 000
隐藏层数量	2
每个隐藏层的神经元数量	256

图4展示了在TD3算法执行RIS相位优化过程中,不同RIS单元数量对应的累积奖励值变化趋势. 可以观察到,随着RIS单元数的增加,系统获得的奖励值也呈上升趋势. 由于奖励函数与车辆用户(Vehicular Users,VUs)的信息传输速率相关,该结果间接说明RIS单元密度对通信性能的显著影响. 然而,在实际部署中,盲目增加RIS单元数量并不可取,因为这将导致显著的计算开销. 因此,在性能增益与计算复杂度之间应权衡考虑,以实现最优系统性能.

图5展示了所提算法与多种对比方法在训练过程中的累积奖励变化趋势. 从整体性能来看,所提算法在训练初期即表现出快速上升,并在300轮左右实现稳定收敛,最终奖励稳定在–2以上,显著优于SAC、QMIX强化学习基准方法,表明其具备更强的学习能力与策略优化效果. 相比之下,SAC与QMIX虽然也表现出一定的学习能力,但在收敛速度与最终性能方面均存在明显差距,曲线波动性较大,稳定性不足. 传统优化算法BCD表现更为劣势,整体奖励水平较低,难以适应高动态环境下的资源优化需求. 同时,随机分配策略(随机RIS+随机RA和无RIS+随机RA)始终维持在较低的奖励水平,进一步验证了引入RIS与智能资源分配机制对于提升系统性能的必要性.

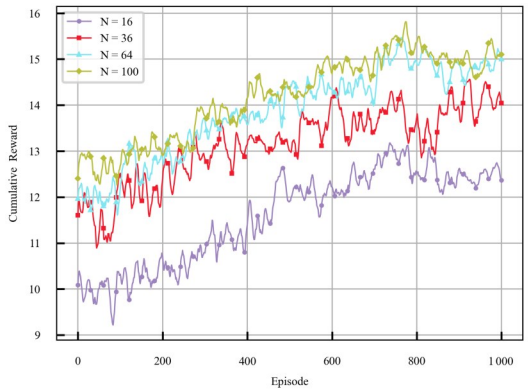

图4　RIS相移优化过程中的累积奖励值比较

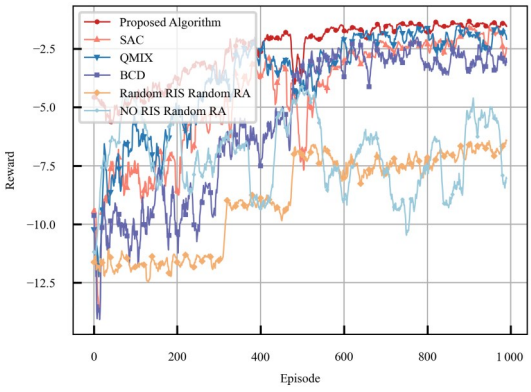

图5　VU功率分配训练的奖励收敛

图6展示了不同算法在训练过程中的AoI变化趋势,用以衡量各方法对任务新鲜度的保障能力.从结果可观察到,所提算法在整个训练过程中始终维持在较低的AoI水平,最终趋于稳定在5 ms左右,显著优于SAC、QMIX、BCD以及两种随机分配策略,表现出更强的信息时效保障能力.相比之下,SAC和QMIX在训练初期虽然AoI有所下降,但稳定性差,波动幅度较大,最终性能亦不及所提方法;BCD整体AoI水平偏高,反映其在动态环境下的任务调度能力较弱.随机分配策略在整个过程中表现最差,AoI普遍维持在7 ms以上,且收敛性与稳定性均不足,验证了盲目资源分配在VEC场景中难以满足低时延服务需求.所提算法在联合优化资源与RIS配置的过程中,能够有效抑制AoI的增长,实现任务的快速处理与高时效传输,凸显其在时延敏感型车联网场景中的应用潜力与优势.

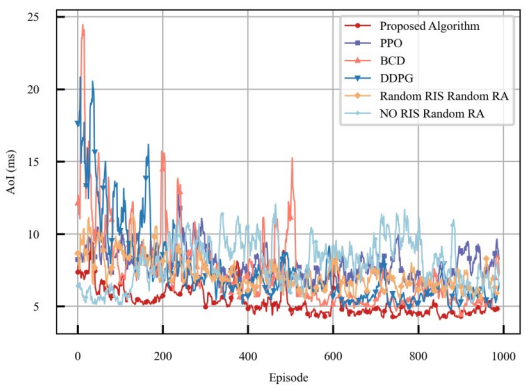

图6 AoI性能比较

图7展示了在不同任务到达率下各算法的总功耗变化趋势,用以评估其在动态负载条件下的能效表现.可以观察到,所提算法在全任务到达率范围内始终保持最低的功耗水平,且增长趋势平缓,体现出卓越的能量控制能力.随着任务到达率的增加,SAC、QMIX与BCD等对比算法的总功耗明显上升,尤其在高负载场景下,SAC的功耗接近1.0 W,表明其在资源调度方面存在能量浪费.而两种随机资源分配策略在全程均表现出最高的能耗,验证了无效资源分配对系统能效的不利影响.所提算法通过联合优化RIS相位与功率分配,在保障系统性能的同时实现能耗最小化,进一步凸显了其在高动态车载边缘计算场景中的应用优势.

图8展示了在8个车辆用户下,不同算法的单用户功耗分布情况,旨在评估各方法在用户侧的能量分配优化能力.结果显示,所提算法在所有用户上的功耗均明显低于SAC、QMIX与传统算法BCD,呈现出稳定且一致的能效优势.尤其在VU4与VU5等高负载用户下,所提方法能将功耗控制在0.2 W以下,而QMIX与BCD对应功耗分别超过0.6 W与0.8 W,说明后者存在

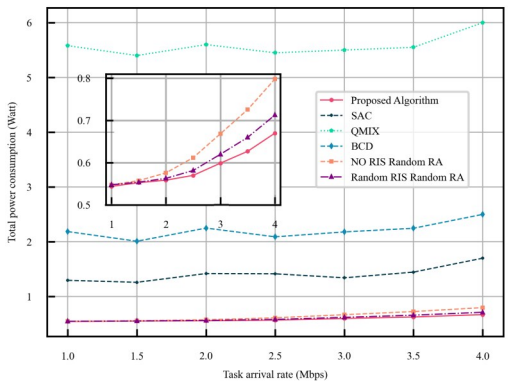

图7 总功耗与任务到达率

功率资源分配不均、策略泛化能力不足等问题.此外,SAC虽然在部分用户(如VU7、VU8)上表现较优,但整体分布不均,部分用户能耗偏高,影响整体网络能效与服务公平性.相较之下,所提算法具备更强的资源自适应性与用户公平性控制能力,能够实现跨用户间的功耗平衡,避免极端能量消耗,有效延长车载终端续航时间.

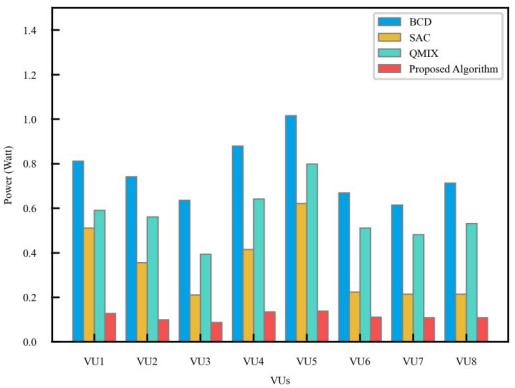

图8 每用户功耗与任务到达率的关系

图9展示了不同算法在用户规模从10扩展至80时的平均功耗变化趋势,以评估其在大规模车联网环境中的扩展性与能效稳定性.从结果可以看出,随着用户数量的增加,所有算法的平均功耗均呈上升趋势,但所提算法始终保持最低水平,且增长斜率最小,体现出优越的能耗控制能力与良好的可扩展性.具体而言,当用户数量达到80时,BCD和QMIX的平均功耗分别超过11 W和9 W,而所提算法仍控制在7 W以下,相比之下降低幅度分别达到36.4%与22.2%.此外,SAC在中小规模下表现尚可,但在用户数增加至60以上后能耗迅速上升,表明其策略泛化能力有限.综上,所提算法在多用户密集接入场景下依然能够保持能效优势,验证了其在RIS辅助大规模VEC系统中的实用性与稳健性.

图10展示了在不同CSI误差方差条件下,各算法

的Average AoI变化趋势,用以评估其对信道不确定性的鲁棒性. 结果表明,随着CSI误差方差从0增大至0.5,所有算法的AoI均有所上升,说明信道估计误差对系统信息新鲜度具有显著影响. 然而,所提算法在全误差区间内始终保持最低的AoI水平,且上升趋势最为平缓,体现出极强的鲁棒性与泛化能力. 以误差方差为0.5为例,所提方法的平均AoI约为1.42 ms,显著低于SAC(1.63 ms)、QMIX(1.8 ms)和BCD(近2.0 ms),表明其在CSI不完美场景下依然能够实现高效可靠的信息调度. 综上,所提算法具备良好的抗CSI扰动能力,能够有效保障车联网中任务的时效性需求,适用于信道动态性强、估计误差大的实际通信环境.

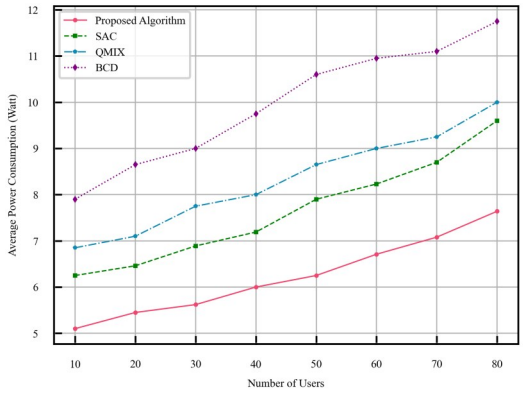

图9　不同用户数下平均功率对比

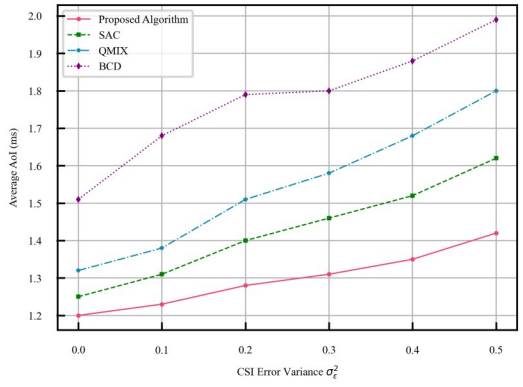

图10　CSI估计误差对AoI的影响曲线图

图11展示了在不同遮挡概率P_{NLoS}条件下,各算法的平均能耗表现,用以评估其在NLoS信道环境下的能效鲁棒性. 结果表明,随着遮挡概率从0增加至0.6,所有算法的平均能耗均有所上升,但所提算法在全范围内始终保持最低能耗,且增长幅度最小,表现出较强的环境适应能力. 当P_{NLoS}=0.6时,BCD算法的平均能耗高达3.5 W,而所提算法仍控制在1.0 W以内,节能效果显著. SAC与QMIX在遮挡概率较低时尚能维持一定能效,但在高遮挡条件下增长迅速,稳定性不足. 综合分析可知,所提算法能够有效应对由遮挡导致的链路质量退化,通过合理调整RIS配置与资源分配策略,实现低能耗运行,展现出在复杂信道环境下的优越能效优化能力.

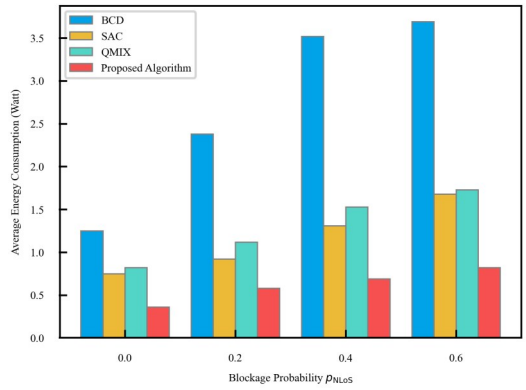

图11　遮挡概率对平均能耗的影响柱状图

图12对比了TD3与DDPG两种下层算法在训练过程中的累积奖励变化趋势,以评估其在连续动作空间下的学习能力与收敛性能. 结果显示,TD3在训练初期波动较大,但在约200轮后逐渐趋稳,并最终收敛至更高的奖励水平,维持在–2以上,整体表现优于DDPG;而DDPG虽然在前期呈现上升趋势,但在600轮之后出现收敛停滞,且最终奖励显著低于TD3. 造成这一现象的主要原因在于DDPG存在策略过估计和梯度更新不稳定的问题,在长时间训练后容易陷入局部最优,导致性能提升受限. 相比之下,TD3通过引入双重评论网络、目标延迟更新以及策略噪声等机制,有效缓解了过估计偏差与策略退化,从而在收敛速度与最终性能上均显著优于DDPG,验证了其在复杂车载边缘计算场景中作为下层资源优化算法的优越性与鲁棒性.

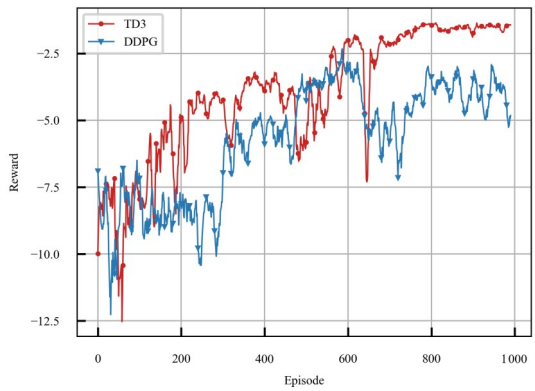

图12　下层TD3与下层DDPG的性能比较

图13展示了在完全相同的上层RIS控制、相同状态-动作表征与统一训练超参数设定下,对所提分层框架与分层SAC、分层QMIX等算法在训练过程中累积奖励的比较结果. 可以看到,所提方法在约200~300个

episode内迅速收敛并维持在较高且平滑的奖励区间，最终稳态奖励显著优于其余三种分层算法；分层SAC虽在早期呈现较快上升趋势，但受熵正则和策略保守性的影响，其稳态奖励仍低于所提方法；分层QMIX在中期出现较明显的振荡与平台期，反映出值函数分解在多智能体非平稳环境下的协调困难；而分层BCD收敛最慢且最终性能最低，进一步说明基于启发式确定性优化在连续、高维控制空间中的局限性.

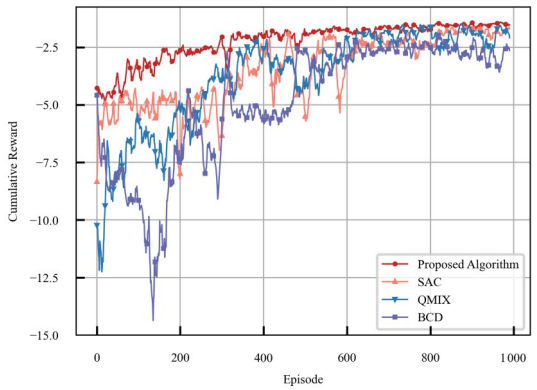

图13 不同分层优化架构算法的累积奖励收敛性能比较

5 结论

本文针对RIS辅助的车载边缘计算系统，围绕信息时效性与能效的协同优化问题，提出了一种分层深度强化学习框架. 该框架通过上下层解耦策略，将高维耦合优化问题划分为RIS相位配置与功率资源分配两个子任务，分别采用TD3与FMADDPG算法进行建模与训练，并通过轨迹池机制实现跨层策略的软耦合，从而提升了训练效率与系统感知能力. 仿真结果表明，所提方法在多用户高动态场景下较现有主流算法在降低AoI与控制能耗方面均具有显著优势.

从理论层面来看，本文的研究具有以下价值：一是拓展了分层强化学习在RIS辅助VEC系统中的适用性，为高维非凸优化问题提供了可扩展的求解范式；二是创新性地引入跨层轨迹嵌入机制，在分布式资源分配过程中增强了全局策略感知，有效缓解了分层训练的不一致性；三是将联邦学习理念融入多智能体资源管理任务，兼顾性能优化与隐私保护，具备良好的理论意义与实际应用前景.

尽管取得了一定成果，本文仍存在局限性：一是算法训练依赖集中式仿真环境，缺乏面向大规模真实部署的轻量化设计；二是系统模型未考虑任务优先级与多任务并发等复杂调度约束，场景适应性仍需提升；三是上下层间虽具协同性，但协同博弈机制尚不完善，策略稳定性与泛化性能有待进一步增强. 未来研究可结合图神经网络与注意力机制以提升模型的全局感知与决策能力，并融合动态任务卸载、缓存及频谱调度等多维资源优化问题，构建更加通用的联合优化框架. 同时，引入联邦安全机制与在线迁移学习策略，有望实现面向实际车联网环境的高效、鲁棒且隐私友好的智能资源调度方案.

参考文献

[1] 许小龙, 杨威, 杨辰翊, 等. 车联网边缘计算环境下基于流量预测的高效任务卸载策略研究[J]. 电子学报, 2025, 53(2): 329-343.
XU X L, YANG W, YANG C Y, et al. Efficient task offloading based on traffic prediction in IoV-enabled edge computing[J]. Acta Electronica Sinica, 2025, 53(2): 329-343. (in Chinese)

[2] 王为念, 苏健, 陈勇, 等. 基于多智能体深度强化学习的车联网频谱共享[J]. 电子学报, 2024, 52(5): 1690-1699.
WANG W N, SU J, CHEN Y, et al. Multi-agent reinforcement learning enabled spectrum sharing for vehicular networks[J]. Acta Electronica Sinica, 2024, 52(5): 1690-1699. (in Chinese)

[3] 李国权, 胡航, 王玥涛, 等. STAR-RIS辅助的CR-SWIPT系统安全波束成形算法[J]. 电子学报, 2024, 52(12): 4002-4008.
LI G Q, HU H, WANG Y T, et al. Secure beamforming algorithm for STAR-RIS assisted cognitive radio systems with SWIPT[J]. Acta Electronica Sinica, 2024, 52(12): 4002-4008. (in Chinese)

[4] ZHANG C, ZHANG W J, WU Q, et al. Distributed deep reinforcement learning-based gradient quantization for federated learning enabled vehicle edge computing[J]. IEEE Internet of Things Journal, 2025, 12(5): 4899-4913.

[5] TANG F X, KAWAMOTO Y, KATO N, et al. Future intelligent and secure vehicular network toward 6G: Machine-learning approaches[J]. Proceedings of the IEEE, 2020, 108(2): 292-307.

[6] ZHU H B, WU Q, WU X J, et al. Decentralized power allocation for MIMO-NOMA vehicular edge computing based on deep reinforcement learning[J]. IEEE Internet of Things Journal, 2022, 9(14): 12770-12782.

[7] HE J L, YU K Q, SHI Y M, et al. Reconfigurable intelligent surface assisted massive MIMO with antenna selection[J]. IEEE Transactions on Wireless Communications, 2022, 21(7): 4769-4783.

[8] DI RENZO M, ZAPPONE A, DEBBAH M, et al. Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road

ahead[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(11): 2450-2525.

[9] AHN J, MEHMOOD MUGHAL D, KIM S H, et al. Computation rate maximization in active RIS-assisted hybrid FDMA-NOMA MEC systems: A deep reinforcement learning approach[J]. IEEE Wireless Communications Letters, 2025, 14(5): 1346-1350.

[10] LIU Y W, LIU X, MU X D, et al. Reconfigurable intelligent surfaces: Principles and opportunities[J]. IEEE Communications Surveys&Tutorials, 2021, 23(3): 1546-1577.

[11] JI Z L, QIN Z J, PARINI C G. Reconfigurable intelligent surface aided cellular networks with device-to-device users[J]. IEEE Transactions on Communications, 2022, 70(3): 1808-1819.

[12] MEI H B, YANG K, LIU Q, et al. 3D-trajectory and phase-shift design for RIS-assisted UAV systems using deep reinforcement learning[J]. IEEE Transactions on Vehicular Technology, 2022, 71(3): 3020-3029.

[13] XIE Y B, SHI L, LI Z H, et al. Efficient task offloading in double roadside RIS-assisted vehicular edge computing networks using deep reinforcement learning[J]. IEEE Transactions on Vehicular Technology, 2025, 74(7): 11353-11365.

[14] NGUYEN K K, TRAN T X, POMPILI D, et al. Reconfigurable intelligent surface-assisted multi-UAV networks: Efficient resource allocation with deep reinforcement learning[J]. IEEE Journal of Selected Topics in Signal Processing, 2022, 16(3): 358-368.

[15] ZHAO J J, YU L, CAI K Q, et al. RIS-aided ground-aerial NOMA communications: A distributionally robust DRL approach[J]. IEEE Journal on Selected Areas in Communications, 2022, 40(4): 1287-1301.

[16] HUANG C, MO R, YUEN C. Reconfigurable intelligent surface for wireless communication: Potential, challenges, and research directions[J]. IEEE Communications Magazine, 2020, 28(2): 136-143.

[17] ZENG Y, ZHANG R, LIM T J. Wireless communications with unmanned aerial vehicles: Opportunities and challenges[J]. IEEE Communications Magazine, 2016, 54(5): 36-42.

[18] HAZARIKA B, SINGH K, BISWAS S, et al. Multi-agent DRL-based task offloading in multiple RIS-aided IoV networks[J]. IEEE Transactions on Vehicular Technology, 2024, 73(1): 1175-1190.

[19] 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.

[20] ZENG L, LIU X, ZHANG W. Joint beamforming design for RIS-assisted multiuser MISO system with discrete phase shifts[J]. IEEE Wireless Communications Letters, 2021, 10(5): 1052-1056.

[21] WANG X, CHEN M, TANG J. Energy-efficient resource allocation in RIS-assisted VEC networks using DRL[J]. IEEE Internet of Things Journal, 2022, 9(3): 1800-1812.

[22] LIU-YANG F, OUYANG W, CHEN L. Intelligent resource management for vehicular networks using multi-agent deep reinforcement learning[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(7): 4123-4135.

[23] ZHENG-LI S, ZHOU P, LIU K. RIS-assisted secure communication in V2X: A game-theoretic approach[J]. IEEE Transactions on Vehicular Technology, 2023, 72(2): 1230-1244.

[24] XU Z, ZHANG X, LI Y, et al. Federated deep reinforcement learning for computation offloading in VEC with non-IID data[J]. IEEE Transactions on Vehicular Technology, 2023, 72(9): 11234-11248.

[25] ZHANG K, MA Y, WANG H, et al. Digital twin empowered industrial Internet of Things: A survey[J]. IEEE Internet of Things Journal, 2021, 8(8): 13789-13804.

[26] PAN Y, WANG H, SHEN X, et al. Federated learning for edge intelligence: A survey[J]. IEEE Communications Magazine, 2021, 59(1): 46-51.

[27] FENG L, LI W J, LIN Y X, et al. Joint computation offloading and URLLC resource allocation for collaborative MEC assisted cellular-V2X networks[J]. IEEE Access, 2020, 8: 24914-24926.

[28] QI K W, WU Q, FAN P Y, et al. Deep-reinforcement-learning-based AoI-aware resource allocation for RIS-aided IoV networks[J]. IEEE Transactions on Vehicular Technology, 2025, 74(1): 1365-1378.

[29] ZHANG R, WU Q. Optimization for intelligent reflecting surface assisted wireless communication: A survey[J]. IEEE Communications Magazine, 2020, 58(1): 26-32.

[30] LIANG F, LIU C, DU J, et al. Intelligent reflecting surface meets machine learning: A survey[J]. IEEE Transactions on Wireless Communications, 2022, 29(1): 114-121.

[31] SUN Q M, NIU J P, ZHOU X W, et al. AoI and data rate optimization in aerial IRS-assisted IoT networks[J]. IEEE Internet of Things Journal, 2024, 11(4): 6481-6493.

[32] LU B S, FANG J L, HONG X M, et al. Task offloading in dynamic energy splitting STAR-RIS assisted NOMA-

MEC systems with decomposition based multi-agent DRL[J]. IEEE Transactions on Vehicular Technology, 2025, 74(8): 13091-13103.

作者简介

兰 军 男,2001年4月出生于甘肃省平凉市. 现为西北师范大学计算机科学与工程学院硕士研究生. 主要研究方向为智能反射面、信息年龄、车联网与资源分配.
E-mail: 2024211162081@nwnu.edu.cn

贾向东 男,1971年8月出生于甘肃省定西市. 教授,博士. 发表学术论文200余篇. 主要研究方向为移动与无线通信关键理论与技术.
E-mail: jiaxd@nwnu.edu.cn

寇志龙 男,2000年5月出生于山西省太原市. 现为西北师范大学计算机科学与工程学院硕士研究生. 主要研究方向为通感一体化网络无线通信、相对论及其应用.
E-mail: 2024211162193@nwnu.edu.cn

包红丽 女,2004年2月出生于甘肃省陇南市. 现为西北师范大学计算机科学与工程学院研究生. 主要研究方向为通感一体化、无人机通信.
E-mail: 2024211162044@nwnu.edu.cn

梁文艳 女,2001年10月出生于山西省太原市. 现为西北师范大学计算机科学与工程学院硕士研究生. 主要研究方向为智能反射面辅助无线通信.
E-mail: 2024211162027@nwnu.edu.cn

武婧婧 女,2001年12月出生于甘肃省武威市. 2024年毕业于西北师范大学计算机科学与工程学院. 现为西北师范大学计算机科学与工程学院硕士研究生. 主要研究方向为通感一体化、智能反射面辅助无线通信等.
E-mail: 2024211162038@nwnu.edu.cn