计算机研究与发展
Journal of Computer Research and Development
DOI: 10.7544/issn1000-1239.202330895
61(9): 2229−2245, 2024

基于多目标深度强化学习的车车通信无线资源分配算法

李 可[1] 马 赛[1] 戴朋林[1] 任 婧[2] 范平志[3]
[1](西南交通大学计算机与人工智能学院 成都 611756)
[2](电子科技大学信息与通信工程学院 成都 611731)
[3](西南交通大学信息科学与技术学院 成都 611756)
(keli@swjtu.edu.cn)

Wireless Resource Allocation Algorithm Based on Multi-Objective Deep Reinforcement Learning for Vehicle-to-Vehicle Communications

Li Ke[1], Ma Sai[1], Dai Penglin[1], Ren Jing[2], and Fan Pingzhi[3]
[1] (*School of Computing and Artificial Intelligence*, *Southwest Jiaotong University*, *Chengdu* 611756)
[2] (*School of Information and Communication Engineering*, *University of Electronic Science and Technology of China*, *Chengdu* 611731)
[3] (*School of Information Science and Technology*, *Southwest Jiaotong University*, *Chengdu* 611756)

Abstract Due to the dynamic uncertainty, diversified service types and scarcity of wireless communication resources in the context of vehicle-to-everything, we explore the challenge of ensuring the requirement for multiple quality of service and the effective utilization of wireless resources in the scenario where V2N (vehicle-to-network) and V2V (vehicle-to-vehicle) links coexist and share spectrum in C-V2X (cellular vehicle-to-everything) networks. First, a multi-objective optimization problem is presented to model the decision-making process of channel selection and power control in C-V2X. The problem considers the impact of dynamic changes in the network environment, aiming to make a balance between the performance of the V2V link (i.e., age of information, delay, and capacity) and the capacity of the V2N link. On this basis, V2V wireless resource allocation algorithm based on multi-objective deep reinforcement learning is also proposed for training neural networks to solve the problem. Through the trained neural network model, the Pareto frontier of the multi-objective optimization problem can be obtained. Simulation results demonstrate that the proposed algorithm can achieve the near-optimal performance for different communication links. Compared with four representative algorithms, the age of information for V2V link is reduced by 12.0% to 17.2%, the V2N link capacity is increased by 11.4% to 21.6%, the V2V link transmission success rate is increased by 4.6% to 13.9%, and the decision delay time is reduced by 10.6% to 20.3%.

Key words cellular vehicle-to-everything; wireless resource allocation; age of information; deep reinforcement learning; multi-objective optimization

摘 要 针对车联网动态不确定特性、业务类型多元化以及无线通信资源稀缺性,研究了蜂窝车联网车与网络(vehicle-to-network, V2N)和车与车(vehicle-to-vehicle, V2V)链路共存且共享频谱场景下保证业务

收稿日期: 2023-11-01; **修回日期**: 2024-05-16
基金项目: 国家重点研发计划项目(2020YFB1807800); 国家自然科学基金项目(62202392, 62172342, U20A20156); 网络与数据安全四川省重点实验室项目(NDS2022-1); 四川省自然科学基金项目(2023NSFSC0459, 2022NSFSC0944); 河北省优秀青年基金项目(F2022105003)
This work was supported by the National Key Research and Development Program of China (2020YFB1807800), the National Natural Science Foundation of China (62202392, 62172342, U20A20156), the Project of Network and Data Security Key Laboratory in Sichuan Province (NDS2022-1), the Natural Science Foundation of Sichuan Province (2023NSFSC0459, 2022NSFSC0944), and the Excellent Youth Fund of the Natural Science Foundation of Hebei Province (F2022105003).

多指标需求和无线资源有效利用的问题. 首先建立多目标优化问题模型来表示蜂窝车联网信道选择和功率控制的决策过程, 该问题考虑了网络环境动态变化的影响, 旨在实现优化目标 V2V 链路的性能(即信息年龄、延迟以及传输速率)和 V2N 链路传输速率之间的权衡. 在此基础上, 提出了基于多目标深度强化学习的车车通信无线资源分配算法进行神经网络训练和问题求解. 通过训练好的神经网络模型可以得到多目标优化问题的帕累托前沿. 仿真实验表明, 所提出算法能够有效地权衡不同通信链路可实现的性能. 与 4 种有代表性的算法比较, V2V 链路信息年龄降低 12.0%~17.2%, V2N 链路传输速率提升 11.4%~21.6%, V2V 链路传输成功率提高 4.6~13.91 个百分点, 决策延迟时间降低 10.6%~20.3%.

关键词 蜂窝车联网; 无线资源分配; 信息年龄; 深度强化学习; 多目标优化

中图法分类号 TP393

车联网[1-3]作为一种多领域交叉的新兴网络, 涉及信息通信、交通、汽车等领域, 引起了国内外工业界与学术界的广泛关注. 蜂窝车联网(cellular vehicle-to-everything, C-V2X)技术是实现车联网中车与车(vehicle-to-vehicle, V2V)、车与基础设施(vehicle-to-roadside infrastructure, V2I)、车与网络(vehicle-to-network, V2N)以及车与人(vehicle-to-pedestrian, V2P)等全方位连接和通信的新一代信息通信技术, 如图 1 所示, 其中 V2N 是指车辆通过接入网或核心网与云平台连接, 云平台与车辆之间进行数据交互, 提供车辆所需要的各类应用服务[4], 如车辆导航、车辆远程监控、紧急救援、信息娱乐服务等, 而 V2V 通信侧重于车辆之间提供低延迟、高可靠、严时效的实时信息传输服务[5-6].

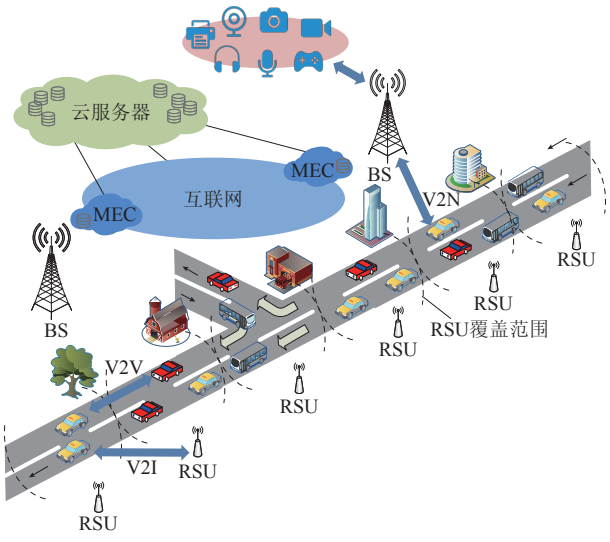


Fig. 1 Architecture of C-V2X

图 1 蜂窝车联网架构

与传统的无线蜂窝网络相比, 车联网具有高动态、时空关联、不确定特性以及严格的服务质量(quality of service, QoS)要求, 这使得 C-V2X 通信面临着诸多特有的挑战. 从空间维度来看, 多个用户同时存在于同一网络中, 并相互竞争有限的无线通信资源, 适当协调用户的传输行为, 处理随机噪声、衰落和干扰的联合影响是必要的; 从时间维度来看, 由于车联网的动态不确定特性, 难以获得准确的信道状态信息(channel state information, CSI), 需要自适应、快速准确地作出传输决策; 从业务类型来看, 不同类型链路需要支持不同 QoS 要求的应用. 因此, 针对车联网动态不确定特性、业务类型的多元化以及无线通信资源稀缺的特点, 研究 V2N 和 V2V 链路资源协同共享以保证 C-V2X 车联网业务的多指标需求和无线资源的有效利用, 是当前车联网资源分配亟需解决的问题[7-8].

鉴于上述问题, 从多目标优化(multi-objective optimization, MO)的角度研究了典型多用户 C-V2X 通信网络的高效传输设计. 特别地, 信息年龄(age of information, AoI)是一种有效的度量信息新鲜度的方法, AoI 描述了自最新状态更新被生成以来所经过的时间[9], 与传统网络性能指标(如延迟、可靠性或传输速率)不同, AoI 被视为一种时效性性能指标. 与现有的大多数研究工作通常仅优化单一目标不同, 进一步提出了以 V2V 链路的 AoI 为目标之一的 MO 问题[10]. 在动态复杂环境下, 确定跨多个时隙的 V2V 信道和功率, 以保证 V2N 与 V2V 通信的 QoS 要求. 主要贡献有 3 个方面:

1)考虑到车联网中 V2N 与 V2V 通信的 QoS 需求差异, 提出了一个新的多目标优化无线资源分配(multi-objective optimization for wireless resource allocation, MO-WRA)问题, 在 V2N 和 V2V 链路共存且共享频谱的复杂蜂窝车联网情况下确定信道选择和功率控制, 以实现不同链路优化目标之间的权衡, 同时保证 V2V 通信链路的 AoI.

2)由于 MO-WRA 问题涉及动态不确定环境下信道状态信息不准确、时间相关的非凸目标和约束

以及相互影响的目标，这导致了极大的决策空间. 结合进化学习，进一步设计了基于多目标深度强化学习的 V2V 资源分配算法，通过训练好的神经网络模型可以得到 MO-WRA 问题的帕累托前沿.

3）为了应对大规模 V2V 通信，加速决策网络提取关键环境状态信息，引入注意力机制以优化深度神经网络，提升神经网络训练速度，增强其实时决策能力.

1 相关工作

近年来，蜂窝车联网资源分配问题得到了广泛研究. 文献 [11] 设计了资源分配算法，该算法在传输可靠性和排队时延的约束下，使 V2V 链路的总吞吐量最大化. 文献 [12] 提出了用于 V2I 和 V2V 链路共存的网络调度和功率控制算法，以提高系统吞吐量. 文献 [13] 提出了一种双时间尺度资源分配算法，该算法基于大时间尺度道路交通信息减小 V2V 链路传输的最大时延. 文献 [14] 也提出了一种基于 V2V 和 V2I 通信的随机模型，该模型结合车辆移动性、信道争用和衰落的影响，提高了通信和计算的可靠性. 尽管上述资源分配策略通过有效的资源分配来提升网络性能，但它们主要关注传统性能指标，如吞吐量、可靠性和时延等，而无法准确衡量接收端的信息新鲜度.

目前学术界提出了“信息年龄 AoI”的概念[9]. AoI 是一种有效度量信息新鲜度的性能指标，被定义为接收端获取的最新数据包自产生时刻到当前接收时刻所经过的时间. 通过资源分配以优化车联网中的 AoI 性能已成为当前的研究热点. 现有研究主要通过控制信息发送频率避免网络拥塞和降低网络传输时延以最小化系统平均 AoI. 文献 [15] 采用李雅普诺夫（Lyapunov）优化方法设计了一种分布式年龄感知数据收集算法，该算法包括基于阈值的源车辆采样策略，可以更加及时地收集状态更新. 文献 [16] 结合平均场理论来分析虚拟传感器网络的网络 AoI，充分考虑了车辆网络的社会特征和潜在的无线通信过程，进一步联合优化源节点处信息更新速率和传感器处的传输概率以最小化平均 AoI. 文献 [17] 利用极值理论和 Lyapunov 优化方法，考虑到 AoI 极端事件发生概率极低的情况，提出了一种感知 AoI 的资源分配算法，以保证超可靠的低时延通信. 但是该算法假设每个车辆用户对中发射机和接收机之间的关联是固定的，这种假设可能会简化模型，但同时也可能忽略了车联网环境中的动态变化，从而限定了对车联网动态性的真实反映. 现有的 AoI 相关的研究大多采用传统的排队论建立理论模型，这种方法在处理大规模动态复杂场景存在局限. 而利用机器学习进行数据驱动建模以更准确地捕捉和模拟车联网中的动态行为，尚有很大的探索空间.

当前研究大多基于一个理想化的假设，即能够获得全局信道状态信息，然而，在车辆高速移动的场景下，信道条件的快速变化使得获取精确的信道状态信息变得极为困难[18]；此外，尽管通过传统信息论方法获得较为准确的信道状态信息[19-20]，但由于计算成本高，仍然难以满足动态车联网环境对于实时应用的迫切需求. 深度强化学习（deep reinforcement learning, DRL）融合了深度学习的感知和强化学习的决策 2 种特性，既可从高维原始数据中直接获取动态环境特征，又具有传统动态规划和马尔可夫决策过程的理论保障以使得网络能够通过智能体与环境的交互来学习动态资源分配策略.

文献 [21] 提出了一种基于信赖域策略优化的车联网联合频谱和功率分配算法，重点研究了系统平均 AoI 的最小化问题. 在参考文献 [22] 中，设计了基于多智能体强化学习的分布式资源分配算法，以最小化系统平均 AoI. 文献 [23] 研究了一种考虑车辆数据包传输模式选择的资源分配问题，并将双时间尺度深度强化学习与谱聚类相结合以提高模型的鲁棒性. 文献 [24] 针对 V2V 复用 V2I 链路的频谱问题提出了基于深度 Q 网络（deep Q-network, DQN）的算法. 虽然该算法考虑了多个优化目标，但是仅考虑各目标权重参数给定的情况，实质上仍是单目标优化. 实际情况下，不同通信链路可能具有不同的传输能力和传输需求，存在多个不一致甚至冲突的优化目标. 在没有预先设定目标权重参数的情况下，对任何目标的优化往往不可避免地会以至少 1 个其他目标的性能下降为代价. 这种现象是多目标优化领域的典型特征，其中需要在多个通常相互冲突的目标之间寻找平衡点[10]. MO 理论已应用于无线通信网络[25]、移动边缘计算系统[26]、车联网[27]. 同样是车联网场景，文献 [27] 提出了一种基于多目标优化理论的顺序传输决策算法，该算法采用了 Lyapunov 优化理论与加权切比雪夫（Chebyshev）方法，在保证不同消息的 QoS 的同时最大化链路的能量有效性. 针对现有文献的分析表明，MO 理论与深度强化学习结合的研究工作尚未得到充分的探索.

与已有工作不同，本文针对蜂窝车联网复杂、动态且不确定场景，进行了深入的探讨. 除了涵盖传统

网络性能指标,本文还创新性地将 AoI 纳入优化目标,提出了多目标优化的资源分配问题,并结合多目标进化学习设计了多目标深度强化学习算法,该算法能够实时进行决策,优化 V2V 无线资源的分配,以满足多样化的 QoS 要求.

2 系统模型与问题描述

本节所涉及的主要符号如表 1 所示.

Table 1 Main Notations Summary

表 1 主要符号汇总

符号	解释
M	V2N 链路集合
K	V2V 链路集合
V	车辆集合
γ_m^{c}	第m条 V2N 链路的信干噪比
g_m^{c}	第m条 V2N 链路的信道功率增益
$\tilde{g}_{k,m}^{\mathrm{v}}$	第k条 V2V 链路对m条 V2N 链路的干扰功率增益
P_m^{c}	第m条 V2N 链路的传输功率
P_k^{v}	第k条 V2V 链路的传输功率
$\rho_{m,k}$	第m条 V2N 链路和第k条 V2V 链路是否共用信道
C_m^{c}	第m条 V2N 链路的传输速率
C_k^{v}	第k条 V2V 链路的传输速率
γ_k^{v}	第k条 V2V 链路的信干噪比
I_k^{c}	第k条 V2V 链路受到 V2N 链路的干扰
I_k^{v}	第k条 V2V 链路受到其他 V2V 链路的干扰
g_k^{v}	第k条 V2V 链路的信道功率增益
B_k^t	在时隙t第k条 V2V 链路的剩余负载量
U_k^t	在时隙t第k条链路的传输延迟容限
L_k^t	在时隙t第k条 V2V 链路的数据包延迟
$A_{i,j}^t$	在时隙t车辆i发送的数据在车辆j接收处的 AoI

2.1 网络模型

本文研究的网络场景如图 2 所示,由 1 个基站(base station, BS)和位于基站通信覆盖范围内的车辆组成,车辆与基站、车辆与车辆之间可以相互通信.假设基站具有计算和缓存能力.具体而言,车辆网络包括$m_{\max}$条 V2N 链路,以集合$M=\{1,2,\cdots,m,\cdots,m_{\max}\}$表示 V2N 链路序号,以及$k_{\max}$条 V2V 链路,以集合$K=\{1,2,\cdots,k,\cdots,k_{\max}\}$表示 V2V 链路序号.

在设计的系统中,可以将 BS 通信范围建模为 2 维欧氏空间ψ,在该范围内包含n个车辆,以集合$V=\{1,2,\cdots,n\}$表示车辆序号.每个车辆$i\in V$由$\{x_i,y_i,o_i,v_i,V_i^{\mathrm{N}},V_i^{\mathrm{T}},A_i\}$表示.其中$x_i$和$y_i$为欧氏空间坐标,$o_i$为

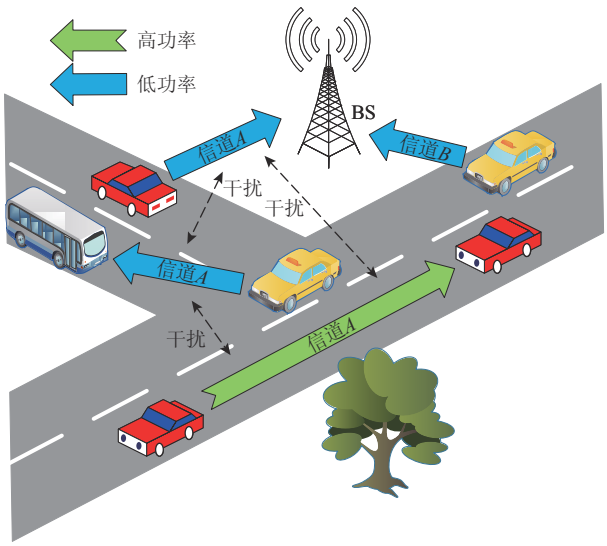


Fig. 2 An example of network scenario

图 2 网络场景示例

车辆行驶的方向,v_i为车辆速度,车辆邻居集合表示为V_i^{N},车辆i的目标通信车辆集合表示为V_i^{T},车辆i发送的数据在其他车辆接收处的 AoI 表示为$A_i=\{A_{i,1},A_{i,2},\cdots,A_{i,n}\}$,用以表征车辆之间所传输信息的新鲜程度.

2.2 通信模型

假设每条 V2N 链路已被预先分配不同的正交子信道以消除网络中 V2N 链路之间的干扰,即第m条 V2N 链路占用第m个子信道,保证了链路之间的无干扰.为提高频谱利用率,假设 V2N 子信道可以被 V2V 链路共享,车辆的收发机采用单天线,当第k条 V2V 链路共享第m条 V2N 链路的子信道时,这条 V2V 链路的接收端可能会受到来自相同子信道的其他 V2V 链路以及 V2N 链路的发射端的干扰.因此系统中可能出现 3 种干扰,分别为:V2N 占用的子信道对共享该子信道的 V2V 的干扰,简称 C2V(cellular user-to-vehicle)干扰;V2V 占用的子信道对使用该子信道的 V2N 的干扰,简称 V2C(vehicle-to-cellular user)干扰;V2V 占用的子信道对占用相同子信道的其他 V2V 用户对的干扰,简称 V2V 干扰.

为了便于建模,将连续时间离散化,用t来表示离散化后的时隙,其中每个时隙的持续时间为t_0.在每个时隙t,基站需要为车辆用户对分配传输信道和发射功率,传输信道集合O和发射功率集合P分别表示为$O=\{O_1,O_2,\cdots,O_{\max}\}$和$P=\{P_1,P_2,\cdots,P_{\max}\}$.

进一步地,定义g_k^{v}为第k条 V2V 链路的信道功率增益,g_m^{c}为第m条 V2N 链路的信道功率增益,$\tilde{g}_{m,k}^{\mathrm{c}}$表示第$m$条 V2N 链路对复用该链路的第$k$条 V2V 链路

的干扰功率增益；$\tilde{g}^{\mathrm{v}}_{k,k'}$表示复用相同 V2N 链路的第$k$条 V2V 链路对第$k'$条 V2V 链路的干扰功率增益. 上述信道功率增益和干扰功率增益均由快衰落和慢衰落组成. 快衰落部分的主要成因是多径效应，慢衰落部分的主要成因包括路径损耗和阴影衰落.

以第m条 V2N 链路的信道功率增益g^{c}_m为例，其计算公式可表示为

$$g^{\mathrm{c}}_m = h^{\mathrm{c}}_m \alpha^{\mathrm{c}}_m = h^{\mathrm{c}}_m \beta_m d^{-\chi}_{m,\mathrm{B}}, \tag{1}$$

其中h^{c}_m是快衰落部分，其服从瑞利（Rayleigh）分布，不同的信道下的快衰落是独立同分布的，α^{c}_m是慢衰落部分，β_m是具有标准差ξ的对数正态阴影衰落，$d_{m,\mathrm{B}}$是信号发射机和接收机之间的欧氏距离，χ是路径损耗分量的衰减指数.

以 V2N 链路占用的子信道对共享该子信道的 V2V 链路的干扰为例，其干扰功率的计算公式可表示为

$$I^{\mathrm{c}}_k = \sum_{m=1}^{m_{\max}} \rho_{m,k} P^{\mathrm{c}}_m \tilde{g}^{\mathrm{c}}_{m,k}, \tag{2}$$

其中P^{c}_m是第m条 V2N 链路的发射功率，$\rho_{m,k}$表示第m条 V2N 链路和第k条 V2V 链路是否共用信道，满足式（3）：

$$\rho_{m,k} = \begin{cases} 1, & \text{第}k\text{条V2V链路重用第}m\text{条V2N链路的子信道,} \\ 0, & \text{其他.} \end{cases} \tag{3}$$

对于第m条 V2N 链路而言，信干噪比γ^{c}_m（signal-to-interference-plus-noise ratio, SINR）可表示为

$$\gamma^{\mathrm{c}}_m = \frac{P^{\mathrm{c}}_m g^{\mathrm{c}}_m}{\sigma^2 + \sum\limits_{k=1}^{k_{\max}} \rho_{m,k} P^{\mathrm{v}}_k \tilde{g}^{\mathrm{v}}_{k,m}}, \tag{4}$$

其中σ^2表示加性高斯白噪声功率，P^{c}_m和P^{v}_k分别表示第m条 V2N 链路和第k条 V2V 链路的发射功率.

对于第k条 V2V 链路，其 SINR 可表示为

$$\gamma^{\mathrm{v}}_k = \frac{P^{\mathrm{v}}_k g_k}{\sigma^2 + I^{\mathrm{c}}_k + I^{\mathrm{v}}_k}. \tag{5}$$

根据上述 V2N 链路和 V2V 链路的 SINR，可以得出 V2V 链路复用 V2N 链路时，第m条 V2N 链路的传输速率C^{c}_m和第k条 V2V 链路的传输速率C^{v}_k的表达式分别为

$$C^{\mathrm{c}}_m = W \times \mathrm{lb}\left(1+\gamma^{\mathrm{c}}_m\right), \tag{6}$$

$$C^{\mathrm{v}}_k = W \times \mathrm{lb}\left(1+\gamma^{\mathrm{v}}_k\right), \tag{7}$$

其中W表示信道带宽.

因此，第k条 V2V 链路的有效传输概率p_k为

$$p_k = Pr\left\{R^{\mathrm{v}}_k \geqslant R_{\mathrm{T}}\right\} = Pr\left\{\frac{1}{T_{\mathrm{d}}}\sum_{t=1}^{T_{\mathrm{d}}} C^{\mathrm{v},t}_k \geqslant \frac{B_0}{U_0}\right\}, \tag{8}$$

其中R^{v}_k表示第k条 V2V 链路的有效传输速率，R_{T}表示有效传输速率阈值，$C^{\mathrm{v},t}_k$表示在时隙t第k条 V2V 链路传输速率，B_0表示数据包的大小，U_0表示传输时延约束，时长为T_{d}个时隙.

对于 V2V 链路而言，还需要关注其延迟性能. 假设只考虑数据包的传输延迟. 定义U_0为数据包生成时的延迟容限，U^t_k表示在时隙t第k条 V2V 链路的延迟容限，更新公式为

$$U^{t+1}_k = U^t_k - t_0, \tag{9}$$

其中t_0为 1 个时隙的长度. 如果在时隙t的前一个时隙内，第k条 V2V 链路完成了数据包传输，那么该数据包的延迟L^t_k可表示为

$$L^t_k = U_0 - U^t_k. \tag{10}$$

如果在时隙t有$U^t_k \leqslant 0$，则说明该数据包的传输时间超过了延迟容限，视为传输失败，不再继续传输该数据包.

2.3 AoI 演进模型

本节介绍 AoI 以及 V2V 链路平均 AoI 的定义及其计算方式.

定义B^t_k为时隙t第k条 V2V 链路传输数据包的剩余负载量，B^t_k的更新公式为

$$B^{t+1}_k = \begin{cases} B_0, & B^t_k < C^{\mathrm{v},t}_k t_0, \\ B^{t+1}_k - C^{\mathrm{v},t}_k t_0, & \text{其他,} \end{cases} \tag{11}$$

其中B_0为初始负载量. 若在当前时隙t内完成了数据包的传输，则在下一个时隙$t+1$将该用户的负载重置为初始负载量B_0；否则，在B^{t+1}_k的基础上减去时隙t内传输的数据量$C^{\mathrm{v},t}_k t_0$.

集合A_i表示车辆i发送的数据在其他车辆接收处的 AoI，所以集合A_i中元素的个数即为车辆数量n，其中$A_{i,j}$表示 V2V 链路中车辆i发送的数据在车辆j接收处的 AoI. $A^t_{i,j}$表示在时隙t车辆i发送的数据在在车辆j接收处的 AoI，$A^{t+1}_{i,j}$的计算如式（12）所示：

$$A^{t+1}_{i,j} = \begin{cases} L^{t+1}_{i,j}, & B^t_{i,j} \leqslant C^{\mathrm{v},t}_{i,j} t_0,\ d^{t+1}_{i,j} \leqslant d_{\mathrm{c}}, \\ A^t_{i,j} + t_0, & B^t_{i,j} > C^{\mathrm{v},t}_{i,j} t_0,\ d^{t+1}_{i,j} \leqslant d_{\mathrm{c}}, \\ 0, & d^{t+1}_{i,j} > d_{\mathrm{c}}, \end{cases} \tag{12}$$

其中$L^{t+1}_{i,j}$表示数据包的延迟，$d^{t+1}_{i,j}$表示车辆i和车辆j之间的欧氏距离，d_{c}表示车辆的通信距离. 当车辆i和车辆j位于通信距离内，如果前一个时隙内有数据包传输成功，$A^{t+1}_{i,j}$为该数据包的延迟，否则$A^{t+1}_{i,j}$随时隙数不断累加；当车辆i和车辆j位于通信距离外，$t=0$.

V2V 链路平均 AoI 可表示为

$$\bar{A}^t = \frac{1}{cnt}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(A_{i,j}^t + A_{j,i}^t\right), \tag{13}$$

其中cnt表示 AoI 非 0 值的个数，可由式(14)计算得到：

$$cnt = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{i,j}^t, \tag{14}$$

其中$\alpha_{i,j}^t \in \{0,1\}$表示车辆$i$和车辆$j$是否位于通信距离内，若$d_{i,j}^{t+1} \leqslant d_{\mathrm{c}}$，则$\alpha_{i,j}^t = 1$，否则$\alpha_{i,j}^t = 0$.

2.4 MO-WRA 问题描述

本节提出多目标优化的无线资源分配 MO-WRA 问题，基站根据车辆传输的状态信息，给该车辆分配信道和发射功率，以实现优化目标，如式(15)所示：

$$\begin{cases} \underset{\{P_k^{\mathrm{v}},\rho_{m,k}\}}{\text{maximize}} \left(\sum_{t=1}^{\infty}V^t, \sum_{t=1}^{\infty}Y^t, \sum_{t=1}^{\infty}Z_k^t, \sum_{t=1}^{\infty}S^t\right) \\ \text{s.t.} \begin{cases} \mathrm{C}_1: L_k^t \leqslant U_0, \forall k \in K, \forall t \in \mathbb{N}_+; \\ \mathrm{C}_2: p_0 \leqslant p_k, \forall k \in K; \\ \mathrm{C}_3: P_k^{\mathrm{v}} \in P, \forall k \in K; \\ \mathrm{C}_4: \rho_{m,k} \in \{0,1\}, \forall m \in M, \forall k \in K. \end{cases} \end{cases} \tag{15}$$

约束C_1表示 V2V 链路延迟约束；约束C_2表示 V2V 链路的有效传输概率约束；约束C_3表示 V2V 链路发射功率约束；约束C_4表示 V2V 链路是否复用 V2N 链路子信道.

式(15)中$P_k^{\mathrm{v}}, \rho_{m,k}$分别表示第$k$条 V2V 链路的发射功率和频谱分配. 针对优化目标中的 V2N 链路传输速率、V2V 链路传输速率、V2V 链路延迟以及 V2V 链路平均 AoI 这 4 个指标，分别定义V^t为时隙t的 V2N 链路传输速率效用函数，Y^t为时隙t的 V2V 链路传输速率效用函数，Z_k^t为时隙t的第k条 V2V 链路的延迟效用函数，S_k^t为时隙t的第k条 V2V 链路的平均 AoI 效用函数. V^t, Y^t, Z_k^t, S_k^t分别表示为

$$V^t \triangleq \sum_{m=1}^{m_{\max}} \frac{C_m^{\mathrm{c},t} - C_{\min}^{\mathrm{c}}}{C_{\max}^{\mathrm{c}} - C_{\min}^{\mathrm{c}}}, \tag{16}$$

$$Y^t \triangleq \sum_{k=1}^{k_{\max}} \frac{C_k^{\mathrm{v},t} - C_{\min}^{\mathrm{v}}}{C_{\max}^{\mathrm{v}} - C_{\min}^{\mathrm{v}}}, \tag{17}$$

$$Z_k^t \triangleq \exp\left(-L_k^t\right), \tag{18}$$

$$S_k^t \triangleq \exp\left(-\left(\bar{A}^t - \bar{A}^{t-1}\right)\right). \tag{19}$$

3 多目标深度强化学习算法

针对多目标优化的无线资源分配 MO-WRA 问题，本节进一步设计了基于注意力机制的多目标近端策略优化算法(multi-objective proximal policy optimization algorithm based on attention mechanism, MOPPO-AM)，如 3.1 节所述，该算法包含 2 个阶段. 在第 1 阶段，将 MO-WRA 问题划分为N个单目标优化的子问题，并将该子问题表述为马尔可夫决策过程，如 3.2 节所述. 通过基于注意力机制的近端策略优化算法(proximal policy optimization algorithm based on attention mechanism, PPO-AM)进行训练，使用的神经网络模型如 3.3 节所述. 将所有训练好的子问题神经网络模型集作为第 2 阶段的初始种群，如 3.4 节所述. 在第 2 阶段，使用进化学习找到多目标主问题的帕累托前沿，如 3.5 节所述.

3.1 MOPPO-AM 算法

MOPPO-AM 算法训练过程如算法 1 所示，该算法通过近端策略优化(proximal policy optimization, PPO)[28]算法来训练基于卷积块注意力模块(convolutional block attention module, CBAM)[29-30]的模型. 然后，将训练好的子问题模型集作为进化学习的初始种群. 进化学习使用近端变异[31]产生后代，基于非支配排序遗传算法Ⅱ(nondominated sorting genetic algorithm Ⅱ, NSGA-Ⅱ)[32]的非支配水平和拥挤距离排序选择模型，最终得到多目标优化问题的一组非支配解，也称为帕累托前沿.

算法 1. MOPPO-AM 算法.

输入：子问题模型集$\Pi = \{\pi_1, \pi_2, \cdots, \pi_N\}$，权重向量集合$\Lambda = \{\lambda_1, \lambda_2, \cdots, \lambda_N\}$，进化学习中的最大迭代次数$I_{\max}$，权重向量个数为$N$；

输出：优化后的模型集$\Pi^* = \{\pi_1^*, \pi_2^*, \cdots, \pi_N^*\}$.

① 随机初始化Π中的模型参数；

② for $i = 1, 2, \cdots, N$ do

③ 使用 PPO-AM 算法训练策略π_i(3.4 节)；

④ end for

⑤ 使用 PPO-AM 算法输出的策略作为进化学习的初始种群；

⑥ for $i = 1, 2, \cdots, I_{\max}$ do

⑦ 使用进化学习进化子问题模型(3.5 节)；

⑧ end for

⑨ return $\Pi^* = \{\pi_1^*, \pi_2^*, \cdots, \pi_N^*\}$.

3.2 马尔可夫决策过程建模

单目标优化子问题表述为马尔可夫决策过程，通常由如下所示的五元组来表征：

$$(S, A, R, \gamma, p), \tag{20}$$

其中S是状态空间，A是动作空间，R是奖励函数，

$\gamma \in (0,1)$为折扣因子，其体现了智能体对即时奖励和未来奖励的权衡，转移概率p是智能体执行某个动作后从一个状态转移到下一个状态的概率.

状态$s_k^t \in S$ 表示在时隙t的第k条链路的状态，状态总共包含 3 类信息. 其中：第 1 类信息为整体的信道和干扰信息，包含 V2V 链路的信道信息G^t、V2N 链路的信道信息H^t、前一个时隙的干扰功率信息I^{t-1}，以及前一个时隙邻居车辆选择的信道信息N_k^{t-1}；第 2 类信息为与资源分配相关的数据包状态信息，包含待发送的信息的剩余负载B_k^t以及延迟容限U_k^t；第 3 类信息是 V2V 链路关联的 AoI A_k^t. 综上所述，s_k^t表达式为

$$s_k^t = \left\{G^t, H^t, I^{t-1}, N_k^{t-1}, O_k^t, U_k^t, A_k^t\right\}. \tag{21}$$

动作$a_k^t \in A$ 包含信道选择O_k^c和发射功率选择P_k^c 2 个维度.

$$a_k^t \in \{1,2,\cdots,3\times|O|\}. \tag{22}$$

第 1 个维度信道数量$|O|$是有限的，第 2 个维度发射功率包含 3 个等级，因此动作空间大小为$3\times|O|$. 在具体的实现过程中，使用$[1,3\times|O|]$范围内的整数对信道进行编号. 对于具体的动作a_k^t，使用$a_k^t \% |O|$表示选择的信道编号，使用$\left\lfloor a_k^t/|O| \right\rfloor$表示选择的发射功率在发射功率列表中的序号.

奖励函数$R\left(s_k^t, a_k^t\right)$ 表示在状态s_k^t且采取动作a_k^t的情况下，即时奖励的期望值. 在时隙t的第k条链路的即时奖励r_k^t的设计为

$$r_k^t = \lambda_V V^t + \lambda_Y Y^t + \lambda_Z Z_k^t + \lambda_S S^t. \tag{23}$$

在奖励函数中，使用链路传输速率作为正向奖励，以此衡量当前决策对其余链路的干扰影响程度. V2V 链路延迟计算由式(10)给出. 对于 V2V 链路平均 AoI，将其进行差分处理，即将相邻时隙 V2V 链路平均 AoI 的变化量$\bar{A}^t - \bar{A}^{t-1}$作为惩罚项参与奖励函数，由式(19)给出. 通过差分处理，可以减少即时奖励的方差，从而提高训练的稳定性. 与此同时，差分处理实际上将本工作的优化目标从直接最小化 V2V 链路平均 AoI 转换为最小化平均 AoI 的增长，这种转换可以更好地引导智能体的训练过程.

3.3 神经网络模型

为了提升算法实时决策能力，算法所采用的神经网络模型引入注意力机制，这有利于网络提取重要的状态信息. CBAM 具有空间和通道注意力，可以关注关键特征，忽略无用特征，使网络更加灵活地适应不同任务和场景，以实现更迅速、更灵活的决策. 卷积神经网络生成特征图之后，CBAM 通过 2 个独立的注意力机制，分别从通道和空间维度对特征图进行加权，以实现自适应特征强化. 这个轻量级的通用模块可以集成到各种卷积神经网络中进行端到端训练，如图 3 所示.

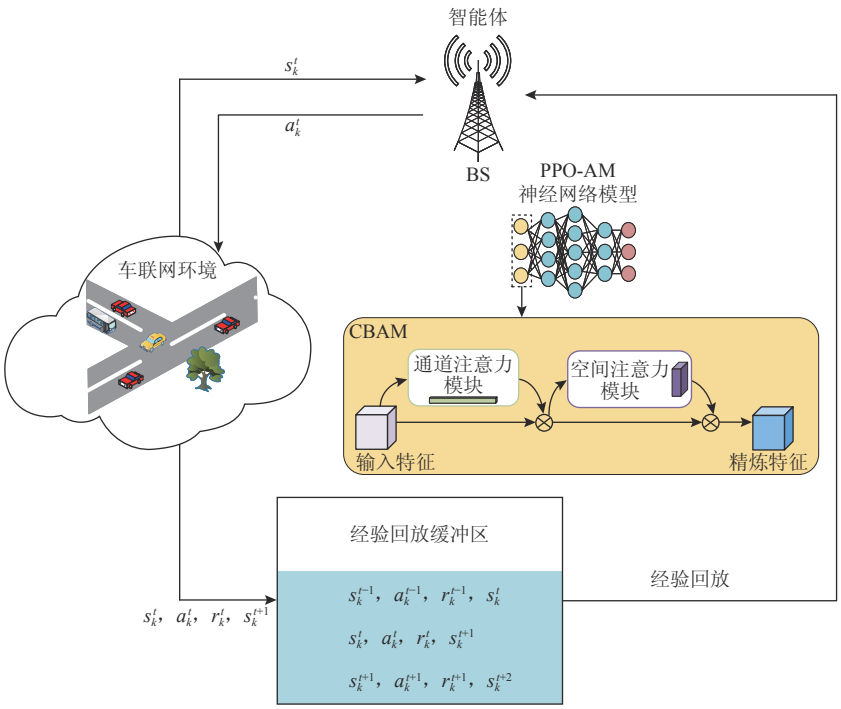


Fig. 3 Framework of PPO-AM

图 3 PPO-AM 框架

给定输入特征图$\boldsymbol{F}$，通道注意力模块推断通道注意力向量，衡量每个通道的重要性. 空间注意力模块推导出3维空间注意力地图，帮助模型更好地理解和利用输入中的空间信息，从而提高了网络的性能和效率. 加权后的特征图为

$$\boldsymbol{F}' = M_{\mathrm{c}}(\boldsymbol{F}) \otimes \boldsymbol{F}, \tag{24}$$

$$\boldsymbol{F}'' = M_{\mathrm{s}}(\boldsymbol{F}') \otimes \boldsymbol{F}', \tag{25}$$

其中M_{c}和M_{s}分别表示基于通道的和基于空间的注意力，⊗表示逐元素乘法，$\boldsymbol{F}'$和$\boldsymbol{F}''$分别表示进行了通道注意力和空间注意力后的输出特征图. 通道注意力模块关注输入数据中有意义的内容，表示为

$$M_{\mathrm{c}}(\boldsymbol{F}) = \sigma(MLP(avgpool(\boldsymbol{F})) + MLP(maxpool(\boldsymbol{F}))), \tag{26}$$

其中$\sigma(\cdot)$表示Sigmoid函数，$maxpool(\cdot)$表示最大池化，$avgpool(\cdot)$表示平均池化，$MLP(\cdot)$表示多层感知器.

空间注意力模块关注输入数据中更有意义的位置，是对通道注意力的补充，表示为

$$M_{\mathrm{s}}(\boldsymbol{F}) = \sigma\left(conv\left(\left[avgpool(\boldsymbol{F}); maxpool(\boldsymbol{F})\right]\right)\right), \tag{27}$$

其中$conv(\cdot)$表示3维卷积层.

3.4 基于PPO-AM的单目标子问题训练算法

在第1阶段，利用PPO-AM算法训练单目标子问题模型. PPO-AM是一种在线深度强化学习算法. 在线学习意味着智能体通过与环境的交互更新策略，而不是像批量学习那样积累一些经验，然后进行一次性的更新. 更新策略的过程是连续进行的，智能体通过不断地与环境交互和从经验池中获得轨迹信息以逐步地改进其策略. 在PPO-AM算法的每次迭代中，基站作为智能体，具有计算和缓存能力，能够使用当前的策略与环境交互，收集经验数据. 然后，利用这些数据，不断逼近状态价值函数和动作价值函数以寻找最优的资源分配策略，如图3所示.

PPO-AM算法是基于策略优化的PPO算法实现的. 其中PPO算法是一种基于策略优化的深度强化学习算法，是在Actor-Critic框架的基础上发展起来的，主要用于训练智能体在环境中采取最优动作策略. 2个策略网络和1个价值网络组成了PPO算法，2个策略网络即新策略和旧策略，比较新策略和旧策略之间的差异，并根据这种差异来确定策略参数的更新方向，这有助于限制策略更新的大小，维持训练的稳定性. PPO算法旨在最大化策略的累积奖励，如式(28)所示，该算法引入状态价值函数、动作价值函数和优势函数更新Actor网络和Critic网络，以找到最优任务选择策略.

$$J(\boldsymbol{\theta}) = E_t\left[\min\left(\frac{\pi_{\boldsymbol{\theta}}(a_t|s_t)}{\pi_{\boldsymbol{\theta}_{\mathrm{old}}}(a_t|s_t)}\hat{A}_t, clip\left(\frac{\pi_{\boldsymbol{\theta}}(a_t|s_t)}{\pi_{\boldsymbol{\theta}_{\mathrm{old}}}(a_t|s_t)}, 1-\boldsymbol{\varepsilon}, 1+\boldsymbol{\varepsilon}\right)\hat{A}_t\right)\right], \tag{28}$$

其中$J(\boldsymbol{\theta})$是目标函数，表示期望累计奖励. $\pi_{\boldsymbol{\theta}}(a_t|s_t)$是在状态$s_t$下采取动作$a_t$的策略函数. $\pi_{\boldsymbol{\theta}_{\mathrm{old}}}(a_t|s_t)$是旧策略函数，即在更新前的策略. $\hat{A}_t$是优势函数，表示在状态s_t下采取动作a_t的优势. $clip$函数确保新策略和旧策略之间的比值在预定义的范围$[1-\varepsilon, 1+\varepsilon]$内. 这有助于防止策略更新过大，从而提高训练的稳定性，ε是用于限制策略更新幅度的超参数.

状态价值函数是指在状态s_t下，遵循策略$\pi_{\boldsymbol{\theta}}$能够获得的期望奖励，如式(29)所示：

$$V_{\pi_{\boldsymbol{\theta}}}^{\boldsymbol{\phi}}(s_t) = E_{\pi_{\boldsymbol{\theta}}}\left[\sum_{i=0}^{\infty}\gamma^i r_{t+i+1}|s=s_t\right], \tag{29}$$

其中$r_{t+i+1} = \sum_{k=1}^{k_{\max}} r_k^{t+i+1}$. γ是折扣因子，用于计算累计奖励，将未来的奖励进行折现. 折扣因子决定了未来奖励的重要性，较小的折扣因子会降低对未来奖励的重视，使智能体更倾向于采取即时奖励更高的动作.

动作价值函数是指在状态s_t下，执行动作a_t之后，遵循策略$\pi_{\boldsymbol{\theta}}$能够获得的期望奖励，如式(30)所示：

$$Q_{\pi_{\boldsymbol{\theta}}}(s_t, a_t) = E_{\pi_{\boldsymbol{\theta}}}\left[\sum_{i=0}^{\infty}\gamma^i r_{t+i+1}|s=s_t, a=a_t\right]. \tag{30}$$

优势函数是指当前状态下采取某个动作相对于采取平均策略的优势，计算方式如式(31)所示：

$$\hat{A}_t = Q_{\pi_{\boldsymbol{\theta}}}(s_t, a_t) - V_{\pi_{\boldsymbol{\theta}}}^{\boldsymbol{\phi}}(s_t). \tag{31}$$

优势函数$\hat{A}_t$越大，意味着当前动作相对于平均水平更好. 在训练过程中，智能体倾向于更频繁地选择具有更大优势的动作. 这可能表明在当前状态下采取该动作更有可能获得更高的奖励或更好的长期奖励.

网络参数$\boldsymbol{\theta}$通过式(32)进行更新.

$$\boldsymbol{\theta} = \boldsymbol{\theta}_{\mathrm{old}} + \delta\hat{\boldsymbol{g}}_{\mathrm{actor}}, \tag{32}$$

其中$\boldsymbol{\theta}$和$\boldsymbol{\theta}_{\mathrm{old}}$分别代表新旧策略的参数. δ代表参数学习率，表示参数更新的快慢. $\hat{\boldsymbol{g}}_{\mathrm{actor}}$是策略梯度，用以更新参数的依据，如式(33)所示：

$$\hat{\boldsymbol{g}}_{\mathrm{actor}} = \nabla_{\boldsymbol{\theta}} L^{clip}(\boldsymbol{\theta}) = E_t\left[\nabla_{\boldsymbol{\theta}}\left(\min\left(\frac{\pi_{\boldsymbol{\theta}}(a_t|s_t)}{\pi_{\boldsymbol{\theta}_{\mathrm{old}}}(a_t|s_t)}\hat{A}_t, clip\left(\frac{\pi_{\boldsymbol{\theta}}(a_t|s_t)}{\pi_{\boldsymbol{\theta}_{\mathrm{old}}}(a_t|s_t)}, 1-\varepsilon, 1+\varepsilon\right)\hat{A}_t\right)\right)\right]. \tag{33}$$

函数$clip$的作用是将新旧策略的比值限制在区

间$[1-\varepsilon,1+\varepsilon]$内，避免了更新步长过大引起的不稳定性，增强了算法的收敛性. 具体地，当优势函数$\hat{A}_t$为正值时，需要增大新旧策略的比值，而比值大于$1+\varepsilon$时，将不提供额外的激励；当优势函数$\hat{A}_t$为负值时，需要减少新旧策略的比值，而比值小于$1-\varepsilon$时，将不再提供额外的激励.

对于 Critic 网络，网络参数$\boldsymbol{\phi}$通过式(34)更新：

$$\boldsymbol{\phi}' = \boldsymbol{\phi} - \eta \hat{\boldsymbol{g}}_{\text{critic}}, \tag{34}$$

其中$\boldsymbol{\phi}$和$\boldsymbol{\phi}'$分别代表 Critic 网络更新前后的网络参数，η代表参数学习率. 关于策略梯度$\hat{\boldsymbol{g}}_{\text{critic}}$，利用均方误差进行计算，如式(35)所示：

$$\begin{aligned}\hat{\boldsymbol{g}}_{\text{critic}} &= \nabla_{\boldsymbol{\phi}} L^{\text{BL}}(\boldsymbol{\phi}) = \\ &E_t \nabla_{\boldsymbol{\phi}} \left(\left(\sum_{i=0}^{\infty} \gamma^i r_{t+i+1} - V_{\boldsymbol{\phi}}(s_t) \right)^2 \right) = \\ &E_t \left[2 \left(\sum_{i=0}^{\infty} \gamma^i r_{t+i+1} - V_{\boldsymbol{\phi}}(s_t) \right) \nabla_{\boldsymbol{\phi}} V_{\boldsymbol{\phi}}(s_t) \right].\end{aligned} \tag{35}$$

PPO-AM 算法伪代码如下所示：

算法 2. PPO-AM 算法.

输入：车辆状态和信道状态的集合S，Actor 网络参数$\boldsymbol{\theta}$，Critic 网络参数$\boldsymbol{\phi}$，最大训练次数J，每局的时隙数T，V2V 链路数$k_{\max}$，奖励折扣因子γ，学习率α，经验回放缓冲区D，经验回放缓冲区大小$D_{\max}$，网络参数更新次数L_{epoch}；

输出：训练完成的策略$\pi_{\boldsymbol{\theta}}$.

① 随机初始化 Actor 网络及其参数、Critic 网络及其参数；
② for $j=1$ to J do
③ 初始化环境；
④ $\pi_{\boldsymbol{\theta}_{\text{old}}} \leftarrow \pi_{\boldsymbol{\theta}}$；
⑤ for $t=1$ to T do
⑥ for $k=1$ to $k_{\max}$ do
⑦ $s_k^t \leftarrow getState(k)$；
⑧ 通过 Actor 网络得到策略$\pi_{\boldsymbol{\theta}}$，进而得到动作$a_k^t$；
⑨ 当前 V2V 链路执行动作a_k^t，计算奖励r_k^t；
⑩ 将数据$\{s_k^t, a_k^t, r_k^t, s_k^{t+1}\}$存入$D$；
⑪ end for
⑫ end for
⑬ if $|D| = D_{\max}$
⑭ for l to L_{epoch}
⑮ 计算$Q_{\pi_{\boldsymbol{\theta}}}$；/*式(30)*/
⑯ 计算$V_{\pi_{\boldsymbol{\theta}}}^{\boldsymbol{\phi}}(s_t)$，$\hat{A}_t$；/*式(29)(31)*/
⑰ 更新 Actor 网络参数$\boldsymbol{\theta}$和 Critic 网络参数$\boldsymbol{\phi}$；/*式(32)~(35)*/
⑱ end for
⑲ end if
⑳ end for

3.5 基于进化学习的多目标主问题训练算法

在第 2 阶段，首先将第 1 阶段训练的子问题模型集作为初始种群. 在每一代，每个单独的模型通过近端变异生成 1 个后代，称为子模型，该后代采用 SM-G-SUM(safe mutation through gradients-summed gradient variant)变异算子[33]将缩放的高斯扰动添加到模型参数$\boldsymbol{\theta}$中，如式(36)所示：

$$\boldsymbol{\theta}' = \boldsymbol{\theta} + \frac{x}{\boldsymbol{\tau}}, x \sim \mathcal{N}(0,\mu), \tag{36}$$

其中μ是变异幅度超参数，$\boldsymbol{\tau}$是子模型网络参数的敏感度向量[33].

具体来说，子问题模型集Π作为初始种群输入到算法 3 中，Π中每个模型作为进化学习中的“个体”. 然后，计算种群中个体的适应度向量(目标向量)，并用适应度向量标记个体. 最后，执行基于多目标的选择. 具体地，采用 NSGA-Ⅱ根据适应度向量的非支配水平和拥挤距离对所有个体进行排序，并保留N个个体. 算法 3 是进化学习的伪代码. 基于进化学习的训练算法过程如图 4 所示.

算法 3. 进化学习.

输入：子问题模型集$\Pi = \{\pi_1, \pi_2, \cdots, \pi_N\}$，其中模型$\pi_i$有参数$\boldsymbol{\theta}_i$，训练批大小为$B_e$，最大迭代次数为$I_{\max}$，权重向量个数为$N$，非支配策略集合$EP$；

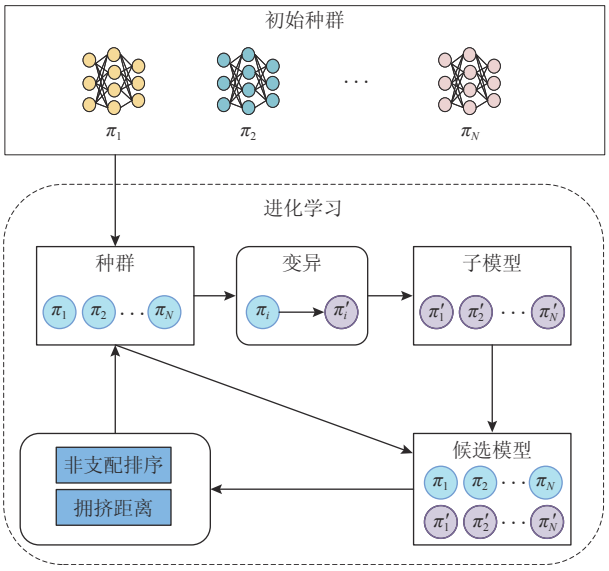


Fig. 4 Overview of training algorithm based on evolutionary learning

图 4 基于进化学习的训练算法概述

输出：优化后的模型集 $\Pi^*=\{\pi_1^*,\pi_2^*,\cdots,\pi_N^*\}$ 和非支配策略集合 EP.

① 初始化非支配策略集合 $EP=\varnothing$;
② for $i=1,2,\cdots,N$ do
③ $\boldsymbol{f}_{\pi_i}=Evaluate(\pi_i)$;
④ end for
⑤ for $g=1,2,\cdots,I_{\max}$ do
⑥ for $i=1,2,\cdots,N$ do
⑦ 对训练数据随机采样
$G_j=RandomInstance(),\forall j\in\{1,2,\cdots,B_e\}$;
⑧ 计算敏感度向量 $\boldsymbol{\tau}$;
⑨ $\boldsymbol{\theta}_i'=\boldsymbol{\theta}_i+\dfrac{x}{\boldsymbol{\tau}}$;
⑩ 产生带有参数 $\boldsymbol{\theta}_i'$ 的后代 π_i'，并且
$\boldsymbol{f}_{\pi_i'}=Evaluate(\pi_i')$;
⑪ end for
⑫ 根据适应度向量的非支配水平和拥挤距离对个体进行排序，并选择 N 个个体作为下一代;
⑬ 根据 π_i' 更新非支配策略集合 EP;
⑭ end for
⑮ return $\Pi^*=\{\pi_1^*,\pi_2^*,\cdots,\pi_N^*\}$.

4 实验设置与结果分析

本节包括实验环境设置、实验对比算法、实验评估指标和实验结果分析.

4.1 实验环境设置

实验基于 TensorFlow 2.7 框架和 Python 3.9 环境，采用 NVIDIA GeForce GTX 1 650 GPU 和 Intel Core i5-10400 CPU. 仿真实验中的系统道路模型参考 3GPP TR 36.885[34] 的城市案例设计，如表 2 所示. 环境中相关参数的设定参考 3GPP TR36.885 中曼哈顿参数，如表 3 所示. 参考该模型，设计得到的道路模型如图 5 所示.

4.2 对比算法

本文设置如下的对比算法以验证所提出算法的性能：

1)随机化资源分配算法 RP(random policy). 在每个时隙，智能体为车辆用户对随机分配传输信道和发射功率.

2)低延时高可靠性优化的深度强化学习资源分配算法 LHP(low-latency high-reliability policy)[24]. 该算法利用深度强化学习 DQN 算法，在 V2V 与 V2N 共享资源的前提下，通过资源分配实现 V2V 链路的延迟、可靠性以及 V2N 链路传输速率的优化.

3)基于 AoI 的低延时高可靠性优化的深度强化学习资源分配算法 LHP-A(low-latency high-reliability policy based on AoI). 该算法是在文献 [24] 原有目标基础上添加了 AoI 的优化目标，使得智能体在决策时额外考虑了链路传输信息的时效性.

4)动态邻近感知资源分配算法 DPP(dynamic proximity-aware policy)[35]. 该算法利用多对 1 匹配博弈算法实现 V2V 链路资源分配以最小化 V2V 链路的延迟.

Table 2 Road Model Parameters

表 2 道路模型参数

参数	城市案例
车道数	每个方向 2 个车道，每条街道共计 4 个车道
车辆数量	60
车道宽度/m	3.5
道路网格大小	433 m×250 m
仿真区域大小	1 299 m×750 m
平均车速/(m·s^{-1})	15
车辆直行概率	0.5
车辆左转概率	0.25
车辆右转概率	0.25

Table 3 Simulation System Parameters

表 3 仿真系统参数

参数	取值
载波频率/GHz	2
信道带宽/MHz	1.5
V2V 链路负载/Kb	30
快衰落模型	Rayleigh 衰落
阴影衰落标准差/dB	3
阴影衰落分布	对数正态分布
基站天线高度/m	25
基站天线增益/dBi	8
基站接收噪声/dB	5
车辆天线高度/m	1.5
车辆天线增益/dBi	3
车辆接收噪声/dB	9
V2N 传输功率/dBm	23，10，5

4.3 评估指标

仿真实验所采用的性能指标如下所示：

1)V2N 传输速率 C^c. V2N 传输速率描述了 V2N 链路在单位时间内传输的数据量. V2N 传输速率越

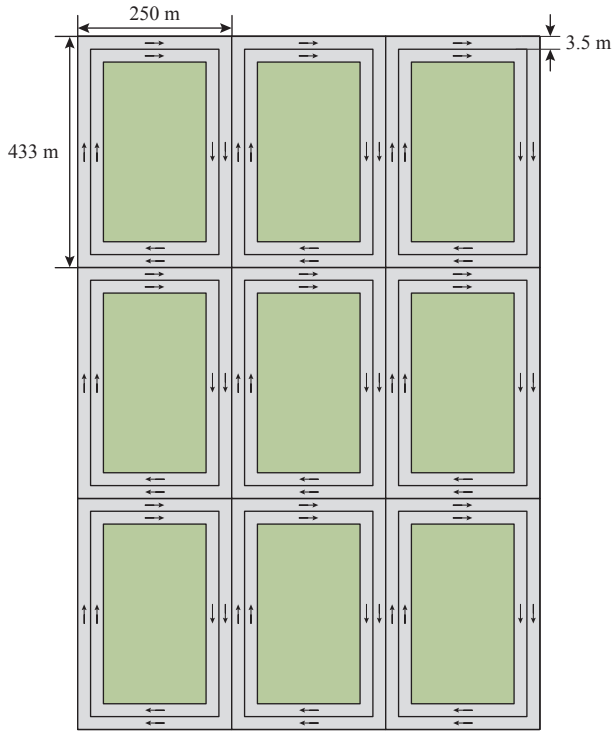


Fig. 5 Schematic diagram of system road model

图 5 系统道路模型示意图

高，表示系统在单位时间内能够传输信息就越多. 这一指标可由式(6)计算得到.

2)V2V 链路平均 AoI$\bar{A}^t$. V2V 链路平均 AoI 表征车辆获取数据的时效性. V2V 链路平均 AoI 值越低，表示链路传输的信息的时效性越强，即传输的信息越新鲜. 这一指标可由式(13)计算得到.

3)V2V 链路传输成功率p_{success}. V2V 链路传输成功率是指 V2V 链路传输过程中，满足延迟约束的数据包所占的比例. V2V 链路数据包传输的成功率越高，表示传输越可靠、满足延迟约束的效果越好. 计算方式如式(37)所示：

$$p_{\text{success}} = \frac{1}{k_{\max}} \sum_{k=1}^{k_{\max}} p_k. \tag{37}$$

4)决策延迟时间t_{d}[36]. 决策延迟时间是衡量智能体响应速度的性能指标，是指从发起请求到决策所经历的平均等待时间，计算方式如式(38)所示：

$$t_{\text{d}} = \frac{1}{n_{\text{d}}} \sum_{i=1}^{n_{\text{d}}} \left(t_i^{\text{s}} - t_i^{\text{e}}\right), \tag{38}$$

其中t_i^{s}和t_i^{e}分表代表第i次决策的发起时刻和决策时刻，n_{d}表示决策的总次数.

4.4 算法性能对比

1)收敛性

图 6 展示了子问题各算法训练过程. 随着训练轮数的增加，各算法的累计奖励均逐渐增大. 其中 LHP-A 算法在 400 轮左右达到收敛，PPO 算法在 350 轮左右达到收敛，PPO-AM 算法在 200 轮左右就已收敛，较前二者分别提速约 50.0% 和 42.9%，收敛后的平均累计奖励分别优化约 5.82% 和 19.41%. 这主要归功于 PPO-AM 算法引入了注意力机制，该机制使得模型在训练过程中能够更加精准地聚焦重要的状态信息，从而加速了对环境关键特征的识别与学习. 注意力机制有效地促进了模型快速捕捉到对累积奖励优化有正面影响的特征，进而加快了整个算法的收敛性.

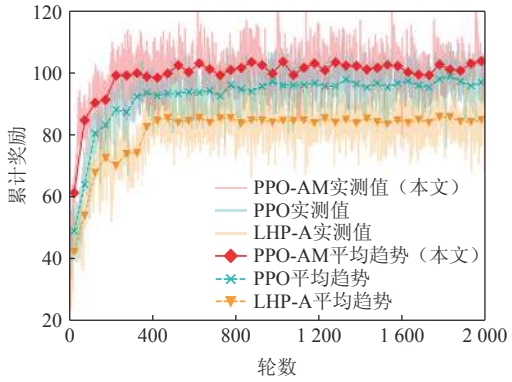


Fig. 6 Training process for the subproblem

图 6 子问题训练过程

2)不同车辆数量下各算法性能对比

车辆数量设置为 20, 40, 60, 80, 100, 120, 基于表 2 和表 3 中的参数，以探究在不同交通密度条件下各算法的性能表现，如图 7~9 所示.

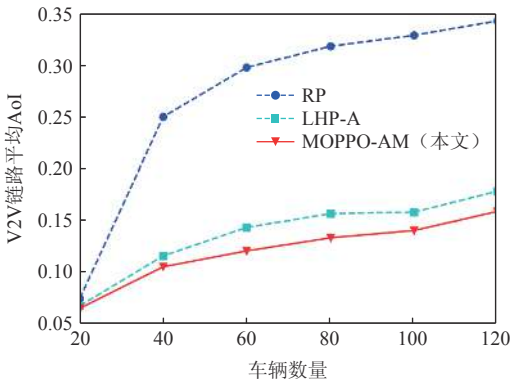


Fig. 7 $\bar{A}^t$ comparison of V2V links with different number of vehicles

图 7 不同车辆数量下 V2V 链路平均 AoI 对比

图 7 展示了不同车辆数量下各算法的 V2V 链路平均 AoI. 就整体而言，AoI 会随着车辆数量的增多呈上升趋势，原因是车辆数量增大会不可避免地导致资源块的竞争变得更加激烈，数据包的传输时间变长，导致链路的 AoI 会变大. 就不同车辆数量而言，

当车辆数量为 20 辆时，3 种算法的 AoI 性能差别不大，这是因为车辆数量较少时，资源竞争并不明显. 当车辆数为 40 辆时，资源竞争开始显现，导致 RP 算法的性能显著下降，其 AoI 开始急剧上升. 相比之下，LHP-A 算法虽然也表现出 AoI 的增长，但其增长速度相对较慢，而 MOPPO-AM 算法显示出在资源受限环境下更为稳健的性能. 当车辆数量继续增加时，MOPPO-AM 算法仍处于较优水平，AoI 增长速度最小. 就平均性能而言，MOPPO-AM 算法的 AoI 较 RP 算法减少 54.4%，较 LHP-A 算法减少 12%，说明 MOPPO-AM 算法在满足 V2V 链路传输信息时效性方面存在优势.

图 8 展示了不同车辆数量下各算法的 V2V 链路传输成功率. 随着车辆数量的增加，各算法的 V2V 链路传输成功率均出现了不同程度的下降趋势，MOPPO-AM 算法下降趋势与其他算法相比较为平稳，但在车辆数为 80 时，成功率出现了较大幅度的下降，原因是车辆数量过多导致不能满足时间约束的数据传输增多. 车辆数量从 40 直到更多，MOPPO-AM 算法相较对比算法的优势更为明显，与 LHP-A 算法相比，V2V 链路传输成功率平均高出 2.1 个百分点，最多可达 5.1 个百分点. 此外，与 LHP 算法和 DPP 算法相比，MOPPO-AM 的 V2V 链路传输成功率分别高出 3.1 个百分点和 4.6 个百分点. 这表明在对比算法中，MOPPO-AM 算法在 V2V 链路传输成功率方面处于较优水平，在保证 V2V 链路传输稳定性方面存在优势.

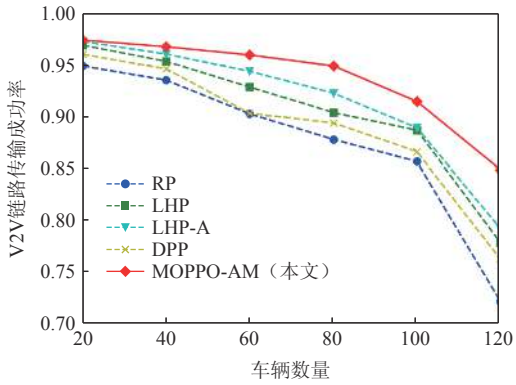


Fig. 8 $p_{success}$ comparison of V2V links with different number of vehicles

图 8 不同车辆数量下 V2V 链路传输成功率对比

图 9 展示了不同车辆数量下各算法的 V2N 链路传输速率. 在不同车辆数量的环境模拟中，MOPPO-AM 算法均处于较优水平：较 RP 算法平均提高约 103%，较 DPP 算法提高约 16.7%，较 LHP 算法和 LHP-A 算法分别提高了 12.1% 和 16.4%.

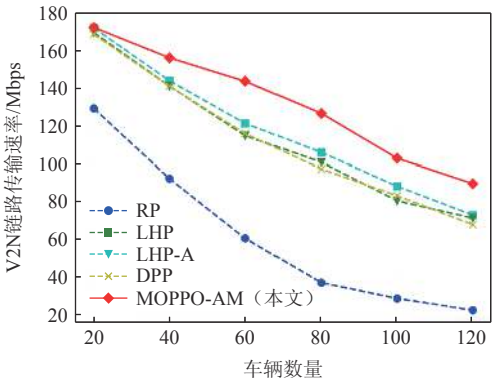
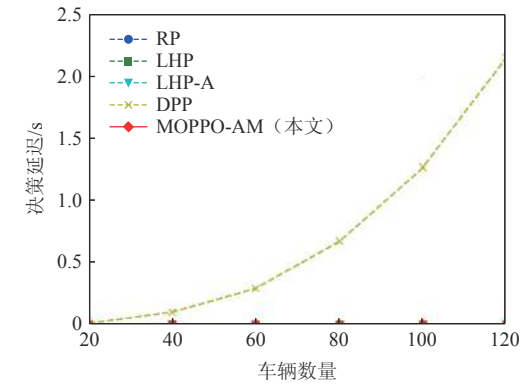


Fig. 9 C^c comparison of V2N links with different number of vehicles

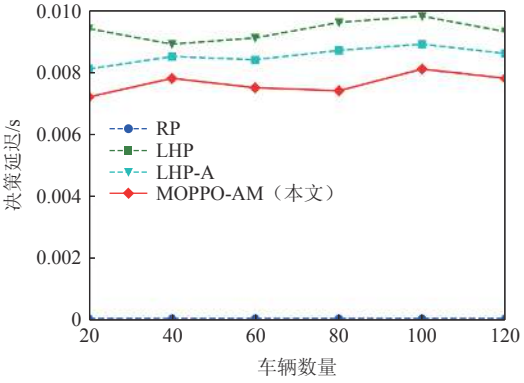
图 9 不同车辆数量下 V2N 链路传输速率对比

综合图 7~9 可知，在环境车辆数量不断增加的情况下，MOPPO-AM 算法可以将车辆数量对性能数据的影响降到最低，由此说明 MOPPO-AM 算法能够更好地适应不同车辆数量的交通场景.

图 10 展示了不同车辆数量下各算法决策延迟时间. 由图 10(a)可知，在决策延迟时间方面，传统算法 DPP, RP 和深度强化学习算法之间存在明显差异. 随着车辆数量的不断增加，DPP 算法的决策延迟时间



(a) 传统算法DPP, RP与深度强化学习算法



(b) 传统算法RP与深度强化学习算法

Fig. 10 t_d comparison with different number of vehicles

图 10 不同车辆数量下决策延迟时间对比

呈现出显著增长的趋势，而深度强化学习算法的决策延迟时间则表现出较好的稳定性. 这表明深度强化学习算法在处理大规模动态场景，尤其是车辆密集的交通环境时，相较于 DPP 算法具有明显的优势. 深度强化学习算法能够更加迅速且准确地做出决策，这不仅提升了决策的响应速度，也有助于实现高效、智能的交通管理. 图 10(b)展示的是除 DPP 算法以外的其他算法决策延迟时间在不同车辆数量下的变化曲线. RP 算法所做出的决策都是随机的，无需与环境交互，故决策时间最短. 其他 3 种算法的决策延迟时间均呈现稳定态势. LHP 算法和 LHP-A 算法的决策延迟时间稳定在 8.53 ms 和 9.57 ms 左右；MOPPO-AM 算法的决策延迟时间稳定在 7.63 ms 左右，在对比算法中处于最优水平，较 LHP 算法和 LHP-A 算法决策时间分别缩短 10.6% 和 20.3% 左右.

3)不同链路负载下各算法性能对比

V2V 链路负载设置为 26 Kb, 28 Kb, 30 Kb, 32 Kb, 34 Kb, 基于表 2 和表 3 中的参数评估在不同链路负载下各算法的性能表现，如图 11~13 所示.

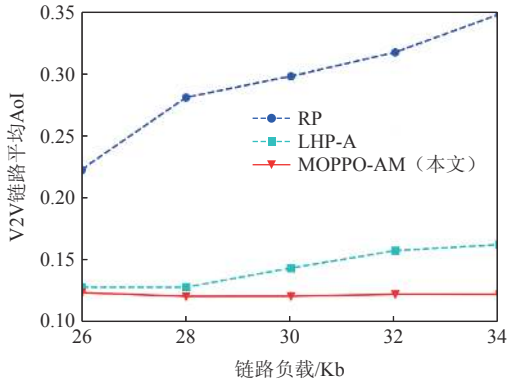


Fig. 11 $\bar{A}^t$ comparison of V2V links with different link loads

图 11 不同链路负载下 V2V 链路平均 AoI 对比

随着 V2V 链路负载持续增加，各算法的性能指标均呈现出变差的趋势，这是因为 V2V 链路负载的增大意味着链路传输任务消耗的资源也随之增多，在相同传输速度情况下，完成 V2V 链路传输任务所耗费的时间更多. 此外，V2V 链路负载的增加还加剧了链路间对有限资源的竞争，导致性能指标的进一步恶化.

图 11 展示了不同链路负载下各算法的 V2V 链路平均 AoI.2 种深度强化学习算法在不同的 V2V 链路负载的情况下均优于 RP 算法. 在链路负载为 26 Kb 和 28 Kb 时，MOPPO-AM 算法和 LHP-A 算法的 AoI 性能接近. 但是，随着 V2V 链路负载不断增加，LHP-A 算法的 AoI 有缓慢增大的趋势，而 MOPPO-AM 算法的 AoI 始终维持在 0.12 左右. 在 AoI 方面，MOPPO-AM 算法较 LHP-A 算法平均提高 15%，较 RP 算法平均提高 58%.

图 12 展示了不同链路负载下各算法 V2V 链路传输成功率. 随着 V2V 链路负载的不断增加，5 种算法的 V2V 链路传输成功率均呈现下降趋势. 其中 RP 算法的表现最差，平均 V2V 链路传输成功率只有 0.83；LHP-A 算法、LHP 算法和 DPP 算法次优，分别为 0.94, 0.93, 0.91；MOPPO-AM 算法的表现最好，V2V 链路传输成功率的下降速度最慢，平均水平可以达到 0.96.

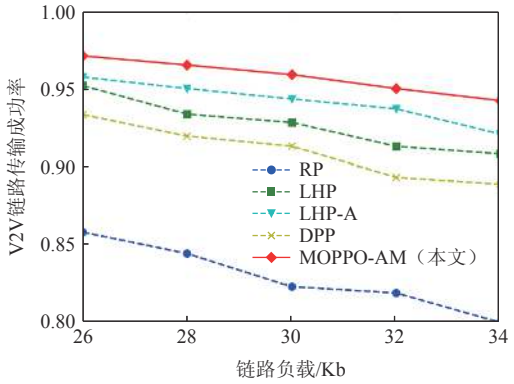


Fig. 12 p_{success} comparison of V2V links with different link loads

图 12 不同链路负载下 V2V 链路传输成功率对比

图 13 展示了不同链路负载下各算法 V2N 链路传输速率. 随着 V2V 链路负载的不断增加，V2V 链路所分配的资源逐渐增多，这加剧了对 V2N 链路的干扰，使得 V2N 链路传输速率也随之降低. 但 MOPPO-AM 算法具有更高的 V2N 链路传输速率，平均值为 140.16 Mbps，较次优的 LHP 算法平均提高 12.79%.

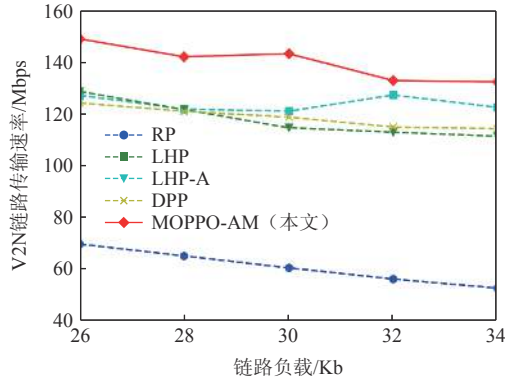


Fig. 13 C^c comparison of V2N links with different link loads

图 13 不同链路负载下 V2N 链路传输速率对比

综合图 11~13 可知，在 V2V 链路负载逐渐增加的情况下，MOPPO-AM 算法仍具有较低的 AoI、较高的 V2V 链路传输成功率和 V2N 链路传输速率，说明 MO

PPO-AM 算法可以更好地完成多种负载大小的传输任务, 且在完成传输任务的同时最小化对 V2N 链路的干扰.

4)不同信道带宽下各算法性能对比

信道带宽设置为 1.00 MHz, 1.25 MHz, 1.50 MHz, 1.75 MHz, 2.00 MHz, 基于表 2 和表 3 中的参数, 以评估在不同信道带宽下各算法的性能表现, 如图 14~16 所示.

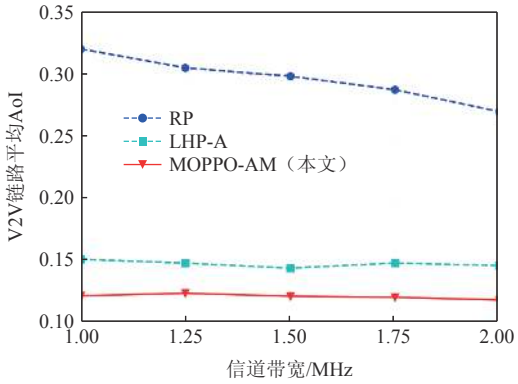


Fig. 14 $\bar{A}^t$ comparison of V2V links with different channel bandwidths

图 14 不同信道带宽下 V2V 链路平均 AoI 对比

随着信道带宽的不断增加, 各算法的性能指标呈现出稳步提升的趋势, 这是因为信道带宽的增加意味着可分配的传输资源变多, 在 V2V 链路负载和延迟约束保持不变的情况下, V2V 链路可分配到的传输资源变多, 链路传输速率会有所提升, 完成传输任务所需要的时间相应变短, 所以性能数据呈现好转趋势.

图 14 展示了不同信道带宽下各算法的 V2V 链路平均 AoI. 虽然 MOPPO-AM 算法的 AoI 下降趋势不比 RP 算法的明显, 但 MOPPO-AM 算法的 AoI 在各种信道带宽的情况下都处于最优水平, 平均值为 0.122, MOPPO-AM 算法较 RP 算法平均减少 58.83%, 较 LHP-A 算法平均减少 17.72%.

图 15 展示了不同信道带宽下各算法的 V2V 链路传输成功率. 随着信道带宽的不断增加, 各算法的 V2V 链路传输成功率均出现了不同程度的上升, 其中 LHP 算法的涨幅最大, 上涨了 13.91 个百分点, MOPPO-AM 算法涨幅最小, 上涨 5.44 个百分点. 但 MOPPO-AM 算法平均表现仍处于较优水平, 平均 V2V 链路传输成功率为 96.16%, 较涨幅最大的 LHP 平均提高 7.5 个百分点.

图 16 展示了不同信道带宽下各算法的 V2N 链路传输速率. 在信道带宽较低的情况下, 其他对比算法的 V2N 链路传输速率普遍偏低, 但 MOPPO-AM 算法仍具有较高且稳定的 V2N 链路传输速率. MOPPO-

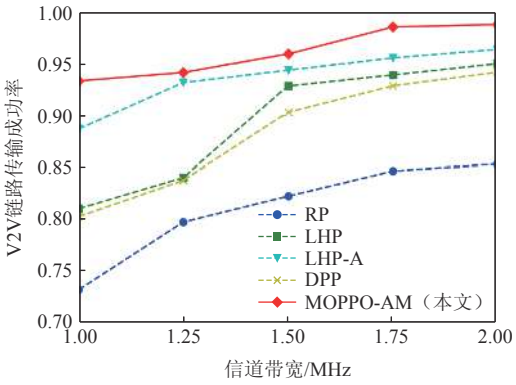


Fig. 15 p_{success} comparison of V2V links with different channel bandwidths

图 15 不同信道带宽下 V2V 链路传输成功率对比

AM 算法在所有信道带宽下 V2N 链路传输速率的平均值为 139.01 Mbps, 在所有对比算法中处于最优水平, 较 RP 算法平均提高 118.89%, 较 DPP 算法平均提高 21.57%, 较 LHP 算法和 LHP-A 算法分别平均提高 13.2% 和 11.4%.

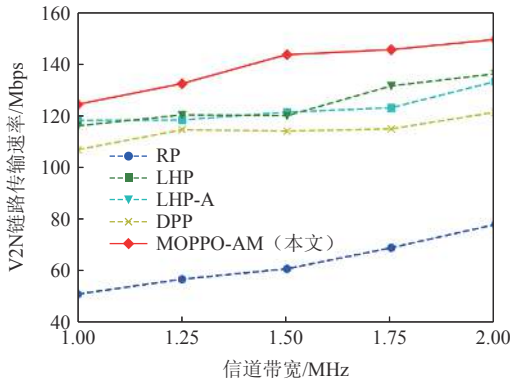


Fig. 16 C^c comparison of V2N links with different channel bandwidths

图 16 不同信道带宽下 V2N 链路传输速率对比

综合图 14~16 可知, 在信道带宽较高、可分配的传输资源较充足的情况下, 各算法的性能指标之间的差距不大. 在传输资源相对匮乏的情况下, MOPPO-AM 算法的性能也出现了下降的趋势, 与其他算法相比, MOPPO-AM 算法性能下降幅度较小且比较稳定, 在信道带宽为 1.00 MHz 和 1.25 MHz 时尤为明显. 由此可见, MOPPO-AM 算法对信道带宽的变化具有更强的鲁棒性和适应性.

综上所述, 实验结果充分证明了 MOPPO-AM 算法在学习能力和环境感知能力方面的出色表现. 该算法不仅能够有效地处理多目标优化问题, 还能通过注意力机制处理与任务最相关的状态信息, 从而在资源受限或竞争激烈的环境中展现显著优势, 并

得出3个结论：

1)注意力机制加速收敛. 基于注意力机制的PPO-AM算法能够更精准地聚焦关键状态特征，减少模型处理不相关信息的复杂度，提高数据利用效率，从而显著加快了收敛速度，有效缩短了训练周期. 同时，PPO通过截断函数 *clip* 限制策略更新的幅度，保证了训练过程的稳定性.

2)多目标优化均衡模型训练效果. 相较于传统的单目标优化算法，基于进化学习的多目标优化MOPPO-AM算法在多个关键性能指标上展现出显著优势，包括V2N链路传输速率、V2V链路传输速率、V2V链路延迟、V2V链路平均AoI以及决策延迟时间. 这是因为MOPPO-AM算法可以均衡V2V链路和V2N链路不同的优化目标，能够更好地满足车联网场景中业务类型的多样化需求.

3)大规模动态复杂场景的高效决策能力. MOPPO-AM算法结合了多目标优化和深度强化学习的优势，显著提升了智能体的探索能力和快速响应能力. 特别是在通信资源竞争加剧的情况下，MOPPO-AM算法依然展现出良好的鲁棒性和适应性，确保了决策过程的高效性和稳定性.

5 结束语

针对蜂窝车联网环境中V2V链路和V2N链路共享无线资源以满足不同性能指标的问题，建立了多目标优化无线资源分配数学规划模型，设计了一种基于进化学习的多目标深度强化学习决策框架求解该问题. 仿真结果表明，该算法保证了智能体在与环境不断交互的过程中快速学习V2V无线资源分配策略，有效解决了动态不确定蜂窝车联网环境下的资源分配问题，旨在实现优化目标V2V链路的性能(即信息年龄、延迟以及传输速率)和V2N链路传输速率之间的权衡. 研究成果不仅提高了蜂窝车联网管控的自动化与自主化效率，而且简化了管控流程与降低人员管理成本，也适用于其他大规模动态复杂网络部署与管理. 后续将研究基于多目标联邦强化学习算法，以解决车联网数据隐私保护场景下的无线资源分配问题.

作者贡献声明：李可负责指导选题、问题建模、算法设计、撰写与修改论文；马赛负责搜集文献资料、实现论文算法、整理实验数据、撰写论文；戴朋林负责指导实验实施、修改论文；任婧和范平志负责网络架构设计和修改论文.

参 考 文 献

[1] Noor-A-Rahim M, Liu Zilong, Lee H, et al. A survey on resource allocation in vehicular networks[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 23(2): 701−721

[2] Kuang Boyu, Li Yuze, Gu Fangming, et al. A review of security research in telematics: Threats, countermeasures and future outlook[J]. Journal of Computer Research and Development, 2023, 60(10): 2304−2321(in Chinese)
(况博裕, 李雨泽, 顾芳铭, 等. 车联网安全研究综述: 威胁、对策与未来展望[J]. 计算机研究与发展, 2023, 60(10): 2304−2321)

[3] Yan Guozhi, Liu Kai, Liu Chunhui, et al. Edge intelligence for Internet of vehicles: A survey[J/OL]. IEEE Transactions on Consumer Electronics, 2024[2024-03-25].https://ieeexplore.ieee.org/abstract/document/10474509

[4] Liu Bingyi, Jia Dongyao, Lu Kejie, et al. Infrastructure-assisted message dissemination for supporting heterogeneous driving patterns[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(10): 2865−2876

[5] Guo Chongtao, Wang Xijun, Liang Le, et al. Age of information, latency, and reliability in intelligent vehicular networks[J]. IEEE Network, 2023, 37(6): 109−116

[6] Gyawali S, Xu Shengjie, Qian Yi, et al. Challenges and solutions for cellular based V2X communications[J]. IEEE Communications Surveys & Tutorials, 2020, 23(1): 222−255

[7] Li Fangwei, Zhang Haibo, Wang Zixin. MEC-based V2X cooperative caching and resource allocation in vehicular networking[J]. Journal on Communications, 2021, 42(2): 26−36 (in Chinese)
(李方伟, 张海波, 王子心. 车联网中基于MEC的V2X协同缓存和资源分配[J]. 通信学报, 2021, 42(2): 26−36)

[8] Liu Zongkai, Dai Penglin, Xing Huanlai, et al. A distributed algorithm for task offloading in vehicular networks with hybrid fog/cloud computing[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2022, 52(7): 4388−4401

[9] Kaul S, Yates R, Gruteser M. Real-time status: How often should one update[C]//Proc of IEEE INFOCOM. Piscataway, NJ: IEEE, 2012: 2731−2735

[10] Cho J H, Wang Yating, Chen I R, et al. A survey on modeling and optimizing multi-objective systems[J]. IEEE Communications Surveys and Tutorials, 2017, 19(3): 1867−1901

[11] Guo Chongtao, Liang Le, Li G Y. Resource allocation for vehicular communications with low latency and high reliability[J]. IEEE Transactions on Wireless Communications, 2019, 18(8): 3887−3902

[12] Nguyen B L, Ngo D T, Dao M N, et al. Scheduling and power control for connectivity enhancement in multi-hop I2V/V2V networks[J]. IEEE Transactions on Intelligent Transporation Systems, 2022, 23(8): 10322−10332

[13] Yang Haojun, Zheng Kan, Zhao Long, et al. Twin-timescale radio resource management for ultra-reliable and low-latency vehicular networks[J]. IEEE Transactions on Vehicular Technology, 2020, 69(1): 1023−1036

[14] Zhou Jianshan, Tian Daxin, Wang Yunpeng, et al. Reliability-optimal cooperative communication and computing in connected vehicle systems[J]. IEEE Transactions on Mobile Computing, 2020, 19(5): 1216−1232

[15] Qin Xiaoqi, Xia Yangyang, Li Hang, et al. Distributed data collection in age-aware vehicular participatory sensing networks[J]. IEEE Internet Things Journal, 2021, 8(19): 14501−14513

[16] Li Zipeng, Xiang Lin, Ge Xiaohu, et al. Age of information modeling and optimization for fast information dissemination in vehicular social networks[J]. IEEE Transactions on Vehicular Technology, 2022, 71(5): 5445−5459

[17] Abdel-Aziz M K, Samarakoon S, Liu Chenfeng, et al. Optimized age of information tail for ultra-reliable low-latency communications in vehicular networks[J]. IEEE Transactions on Communications, 2020, 68(3): 1911−1924

[18] Khan W U, Jamshed M A, Lagunas E, et al. Energy efficiency optimization for backscatter enhanced NOMA cooperative V2X communications under imperfect CSI[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(11): 12961−12972

[19] Wu Weihua, Liu Runzi, Yang Qinghai, et al. Learning-based robust resource allocation for ultra-reliable V2X communications[J]. Transactions on Wireless Communications, 2021, 20(8): 5199−5211

[20] Wu Weihua, Liu Runzi, Yang Qinghai, et al. Robust resource allocation for vehicular communications with imperfect CSI[J]. IEEE Transactions on Wireless Communications, 2021, 20(9): 5883−5897

[21] Peng Nuoheng, Lin Yan, Zhang Yijin, et al. AoI-aware joint spectrum and power allocation for Internet of vehicles: A trust region policy optimization-based approach[J]. IEEE Internet Things Journal, 2022, 9(20): 19916−19927

[22] Parvini M, Javan M R, Mokari N, et al. AoI aware radio resource management of autonomous platoons via multi agent reinforcement learning[C/OL]//Proc of the 17th Int Symp on Wireless Communication Systems. Piscataway, NJ: IEEE, 2021[2023-09-16].https://ieeexplore.ieee.org/document/9562190

[23] Zhang Xinran, Peng Mugen, Yan Shi, et al. Deep reinforcement learning based mode selection and resource allocation for cellular V2X communications[J]. IEEE Internet Things Journal, 2020, 7(7): 6380−6391

[24] Ye Hao, Li G Y, Juang B H F. Deep reinforcement learning based resource allocation for V2V communications[J]. IEEE Transactions on Vehicular Technology, 2019, 68(4): 3163−3173

[25] Han Yan, Tao Xiaofeng, Zhang Xuefei, et al. Delay-aware resource management for multi-service coexisting LTE-D2D networks with wireless network virtualization[J]. IEEE Transactions on Vehicular Technology, 2020, 69(7): 7339−7353

[26] Song Fuhong, Xing Huanlai, Luo Shouxi, et al. A multiobjective computation offloading algorithm for mobile-edge computing[J]. IEEE Internet of Things Journal, 2020, 7(9): 8780−8799

[27] Ma Mengyu, Wang Chao, Li Zuxing, et al. A multi-objective optimization-based transmission design for V2V communication networks[J]. IEEE Transactions on Vehicular Technology, 2023, 72(10): 13081−13093

[28] Schulman J, Wolski F, Dhariwal P, et al. Proximal policy optimization algorithms[J]. arXiv preprint, arXiv: 1707.06347, 2017

[29] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional block attention module[C]//Proc of the European Conf on Computer Vision. Berlin: Springer, 2018: 3−19

[30] Liu Bingyi, Han Weizhen, Wang Enshu, et al. Multi-agent attention double actor-critic framework for intelligent traffic light control in urban scenarios with hybrid traffic[J]. IEEE Transactions on Mobile Computing, 2024, 23(1): 660−672

[31] Bodnar C, Day B, Lió P. Proximal distilled evolutionary reinforcement learning[C]//Proc of the 34th AAAI Conf on Artificial Intelligence. Palo Alt, CA: AAAI, 2020: 3283−3290

[32] Deb K, Pratap A, Agarwal S, et al. A fast and elitist multiobjective genetic algorithm: NSGA-Ⅱ[J]. IEEE Transactions on Evolutionary Computation, 2002, 6(2): 182−197

[33] Lehman J, Chen J, Clune J, et al. Safe mutations for deep and recurrent neural networks through output gradients[C]//Proc of the Genetic and Evolutionary Computation Conf. New York: ACM, 2018: 117−124

[34] Merias P. Study on LTE-based V2X services (release 14), TR 36.885 V14.0. 0[R]. Sophia-Antipolis, France: 3GPP, 2016

[35] Ashraf M I, Bennis M, Perfecto C, et al. Dynamic proximity-aware resource allocation in vehicle-to-vehicle (V2V) communications[C/OL]//Proc of 2016 IEEE Globecom Workshops (GC Wkshps). Piscataway, NJ: IEEE, 2016[2023-09-21].https://ieeexplore.ieee.org/abstract/document/7848885

[36] Lin Kai, Li Chensi, Li Yihui, et al. Distributed learning for vehicle routing decision in software defined Internet of vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(6): 3730−3741



Li Ke, born in 1983. PhD, master supervisor. Member of CCF. Her main research interests include machine learning, distributed system, and Internet of vehicles.

李 可, 1983 年生. 博士, 硕士生导师. CCF 会员. 主要研究方向为机器学习、分布式系统、车联网.



Ma Sai, born in 2000. Master candidate. His main research interests include machine learning and Internet of vehicles.

马 赛, 2000 年生. 硕士研究生. 主要研究方向为机器学习、车联网.



Dai Penglin, born in 1990. PhD, associate professor. Member of CCF. His main research interests include Internet of vehicles, edge computing, and intelligent transportation systems.

戴朋林, 1990 年生. 博士, 副教授. CCF 会员. 主要研究方向为车联网、边缘计算、智能交通系统.



Ren Jing, born in 1982. PhD, assistant researcher. Her main research interests include network architecture, protocol design, network modeling and optimization, and network security.

任　婧, 1982 年生. 博士, 助理研究员. 主要研究方向为网络架构、协议设计、网络建模与优化、网络安全.



Fan Pingzhi, born in 1955. PhD, professor. IEEE fellow. His main research interests include high mobility wireless communications, massive random-access techniques, and signal design and coding.

范平志, 1955 年生. 博士, 教授. IEEE 会士. 主要研究方向为高移动无线通信、大规模随机接入技术、信号设计与编码.