

基于 EMA 评估的认知车联网频谱共享研究

陈一凡¹ 陈玲玲^{1*} 王一帆¹

CHEN Yifan CHEN Lingling WANG Yifan

摘要

针对认知车联网 (CIoV) 中频谱资源日益紧张的问题, 文章提出了一种基于指数移动平均 (EMA) 声誉评估的动态频谱定价与分配模型。该模型利用 EMA 量化次用户车辆 (SV) 对系统吞吐量的贡献, 建立声誉值与频谱定价之间的关联, 激励次用户车辆优化资源使用效率。在定价约束条件下, 将频谱分配建模为最优决策问题, 并采用柔性执行者-评价者 (SAC) 算法进行求解。仿真结果表明, 与 Q-learning 和 A2C 算法相比, SAC 算法能够显著提升系统总效益以及主、次用户吞吐量, 验证了所提模型的有效性。

关键词

认知车联网; 频谱共享; 指数移动平均法; 声誉值

doi: 10.3969/j.issn.1672-9528.2026.03.021

0 引言

车联网 (IoV) 是一个复杂的技术生态与应用系统, 其快速发展正面临频谱资源可用性方面的巨大挑战。由于汽车行业智能化和网络化的持续趋势, 所以车联网通信技术至关重要。虽然专用短距离通信 (DSRC) 和蜂窝 V2X (包括 LTE-V2X 和 5G NR-V2X) 等通信标准能够促进安全关键性应用, 但车联网不断增长的多样化服务需求, 包括提高交通效率、全面的车辆信息传播和车载娱乐, 正在加速频谱需求的快速增长。这种日益扩大的差距加剧了频谱稀缺的现状^[1]。

认知无线电技术 (cognitive radio, CR)^[2] 是一种提高传统无线通信中频谱效率的技术, 工作原理是动态频谱访问。通过智能感知、自主学习与电磁环境自主搜索实现, 使次用户车辆 (SV) 可动态接入并利用主用户车辆 (PV) 当前未占用的授权频段。在不对 PV 产生干扰的前提下, 实现机会式频谱复用, 认知无线电 (CR) 可有效缓解有限频谱资源的供需矛盾。

与传统的认知无线网络不同, 认知车辆网络由于车辆的移动性和集成性造成了独有的特性和挑战。世界各地的研究人员正在积极探索认知车辆网络中的频谱资源分配解决方案。例如, 文献 [3] 提出了一种新的认知车联网频谱管理框架, 使车辆能够实时访问 TV 频段中未使用的频谱。在此基础上; 文献 [4] 进一步探索了使用 CR 技术将 DSRC 扩展到 TV 频谱, 通过仿真证明他们的新模型在数据传输方面优于传统系统; 文献 [5] 提出了一种频谱共享方案, 允许服务提供商租赁未充分利用的频段, 从而提高运营效率并增加利润率。该框架保证了卓越的服务, 扩大了对更广泛用户群的访问范围, 并

证明了提高频谱效率和所有相关利益相关者的经济利益的巨大潜力。然而, 这些传统方法大多采用单智能体框架且依赖先验知识, 难以在环境快速变化的场景中实现准确预测和高效决策。

针对认知车联网场景下频谱资源竞争加剧的现状, 以及传统频谱分配算法存在的复杂度较高问题, 为进一步提升 PV 与 SV 参与频谱共享的积极性与分配公平性, 本文创新性地提出一种基于 SAC 算法的频谱高效共享方案。通过构建基于主次用户吞吐量评估的声誉值定价模型, 优化主用户与次用户的总效益指标, 以实现主次用户吞吐量最大化的核心目标。具体而言, 其通过历史交互平衡双方的资源需求, 在解决频谱资源紧张问题的同时, 有效降低算法计算复杂度, 为认知车联网频谱资源的分配提供了新的技术路径。

1 系统模型

1.1 基于指数移动平均的声誉评估设计

本文构建了一个面向认知车联网 (CIoV) 的频谱共享网络模型, 该模型包含若干基站、 I 辆 PV 及 J 辆 SV, 且每辆 SV 仅能占用单一信道资源。如图 1 所示。

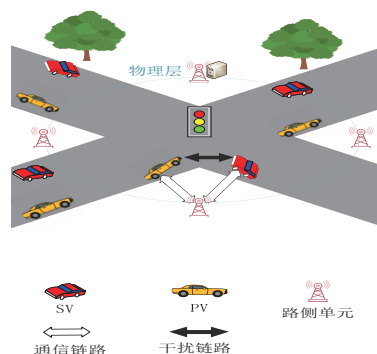


图 1 认知车联网的频谱共享网络模型

本节提出一种基于指数移动平均 (EMA) [6-7] 的声誉值评估模型, 通过量化 SV 在占用 PV 信道且不影响 PV 的情况下 SV 的吞吐量、SV 在不影响 PV 的情况下 PV 的吞吐量。该模型能够实时反映 SV 的最新行为趋势, 以平滑的方式整合历史数据, 关注 SV 在占用信道时对 PV 吞吐量的保护能力及其自身的频谱利用效率。

(1) 评分理念与通信指标

声誉值评估模型关注 SV 占用 PV 信道时对系统整体性能和 PV 的影响。为此, 定义了 4 个关键的通信指标, 作为评估 SV 声誉值的基础, 具体如下:

PV 的信干噪比计算公式 SNR_i 表示为:

$$\text{SNR}_i = \frac{p_i \cdot g_i}{\sigma \cdot B + I_j} \quad (1)$$

式中: p_i 表示 PV 的发射功率; g_i 表示 PV 的信道增益; σ 表示噪声功率谱密度; B 表示 PV 占用的信道带宽; $\sigma \cdot B$ 表示噪声总功率; I_j 表示 SV 对 PV 造成的干扰。

SV 对 PV 造成的干扰公式 I_j 表示为:

$$I_j = \sum_{j=1}^J p_j \cdot g_j \quad (2)$$

式中: p_j 为 SV 的发射功率; g_j 为 SV 的信道增益。

考虑通信过程中的信干噪比时, 根据信道容量公式计算 SV 在不影响 PV 的情况下的 PV 吞吐量, 表示为:

$$C_p = B \log_2 [1 + \text{SNR}_i] \quad (3)$$

式中: SNR_i 表示信道的信干噪比。

SV 的信干噪比计算公式 SNR_j 表示为:

$$\text{SNR}_j = \frac{p_j \cdot g_j}{\sigma \cdot B + I_i} \quad (4)$$

式中: B 表示 SV 占用的信道带宽; I_i 表示 PV 对 SV 造成的干扰。

PV 对 SV 造成的干扰公式 I_i 表示为:

$$I_i = \sum_{i=1}^I p_i \cdot g_i \quad (5)$$

同样, 考虑通信过程中的信干噪比时, 根据信道容量公式计算 SV 在不影响 PV 的情况下的 SV 吞吐量, 表示为:

$$C_s = B \log_2 [1 + \text{SNR}_j] \quad (6)$$

式中: B 表示 SV 占用的信道带宽; SNR_j 表示信道的信干噪比。

综上, 本文考虑 SV 在不影响 PV 的情况下的 PV 吞吐量 C_p 和 SV 吞吐量 C_s 两个通信指标来评估 SV 的声誉值。

(2) 基于指数移动平均 (EMA) 的声誉值评估

EMA 是时间序列分析中常用的平滑技术, 赋予近期数据更高的权重, 而较旧数据点的权重呈指数级衰减 [7]。这使得

EMA 适合用于动态变化的声誉值评估, 因为它能够实时响应 SV 行为的变化。因此, 将上述得到的 PV 吞吐量公式和 SV 吞吐量公式进行标准化处理, 目的是将不同量纲的吞吐量值统一到声誉评估所需的标准区间 [0,1] 之间, 具体表示为:

PV 吞吐量标准化表示为:

$$\hat{C}_p = \frac{C_p - C_p^{\min}}{C_p^{\max} - C_p^{\min}} \quad (7)$$

式中: $\hat{C}_p \in [0, 1]$; $C_p^{\min}$ 和 $C_p^{\max}$ 分别是 PV 吞吐量的最小值和最大值。

SV 吞吐量标准化表示为:

$$\hat{C}_s = \frac{C_s - C_s^{\min}}{C_s^{\max} - C_s^{\min}} \quad (8)$$

式中: $\hat{C}_s \in [0, 1]$; $C_s^{\min}$ 和 $C_s^{\max}$ 分别是 SV 吞吐量的最小值和最大值。

单次交互的性能得分 S_k 表示为 SV_j 每次完成信道占用交互 K 后, 计算本次交互的综合性能得分 S_k 。这个得分反映了此次交互中 SV 对 PV 吞吐量的保护及其自身的吞吐量表现。具体公式为:

$$S_k = \omega_1 \hat{C}_p^{(k)} + \omega_2 \hat{C}_s^{(k)} \quad (9)$$

式中: $\hat{C}_p^{(k)}$ 和 $\hat{C}_s^{(k)}$ 是第 k 次交互中标准化后的 PV 和 SV 吞吐量; ω_1 和 ω_2 是预设的权重, 用于衡量两个指标在声誉值评估中的重要性, 满足 $\omega_1 + \omega_2 = 1$ 。

SV_j 的声誉值 r_j 将通过 EMA 公式进行动态更新。假设 $r_j^{(k-1)}$ 是 SV_j 在第 $k-1$ 次交互后的声誉值, 第 k 次交互后的声誉值 $r_j^{(k)}$ 则计算为:

$$r_j^{(k)} = \alpha_1 \cdot S_k + (1 - \alpha_1) \cdot r_j^{(k-1)} \quad (10)$$

式中: $r_j^{(k)}$ 是 SV_j 在第 k 次交互后的最新声誉值; S_k 是第 k 次交互的综合性能得分; $r_j^{(k-1)}$ 是 SV_j 在第 $k-1$ 次交互时的声誉值; α_1 是平滑因子, 取值范围为 (0,1]。较大的 α_1 值表示模型对最新交互的响应更灵敏, 声誉值波动更大, 遗忘速度快; 较小的 α_1 值表示模型对最新交互的响应更平滑, 声誉值变化更稳定, 但对历史行为的记忆更长。

对于新加入 CIOV 网络的 SV, 因历史交互数据缺少, 其初始声誉值 $r_j^{(0)}$ 可以设定为一个中值。

1.2 基于声誉值的动态频谱资源定价和分配模型

本定价模型的核心设计理念是将次用户在 CIOV 网络中的通信指标表现 [8], 量化为可计算的声誉值 (reputation value, Rep), 将其融入频谱资源的价格中。声誉值高的 SV 将支付较低的频谱使用价格, 而声誉值低的 SV 则需支付较高的价格。具体定价公式为:

$$P_j = P_b \cdot (1 - \alpha_2 \cdot r_j^{(k)}) \quad (11)$$

式中: P_j 为 SV 在申请单位频谱资源时需支付的实际价格; P_b 为单位频谱资源的基础价格; $r_j^{(k)}$ 为 SV 的声誉值, 由此前 EMA 公式得出, 其取值范围为 $[0,1]$; α_2 为声誉影响因子, 取值范围为 $[0,1]$, 该参数决定了声誉值对最终价格的权重。

SV_j 的成本在于获取所需频谱资源所支付的费用。具体计算公式为:

$$Ct_j = -\sum_{i=1}^I \eta_j^i \cdot P_j \cdot s_j^i \quad (12)$$

$$\sum_{i=1}^I s_j^i \leq S, \quad \sum_{i=1}^I s_j^i = D \quad (13)$$

式中: P_j 表示单位频谱资源价格单位, 由 SV_j 的声誉值动态确定; s_j^i 表示 SV_j 向 PV_i 购买的频谱资源量; η_j^i 表示 SV_j 是否向 PV_i 购买频谱资源的二元决策变量; D 表示 SV_j 的频谱资源需求量; S 表示 PV_i 所能提供的最大频谱资源量。

同样, PV_i 作为频谱资源的提供方, 其收入来源于向 SV_j 出售空闲频谱资源。因此, PV_i 的收入函数计算公式为:

$$Rv_i = \sum_{j=1}^J \eta_j^i \cdot (P_j - \varepsilon) \cdot s_j^i \quad (14)$$

式中: ε 表示 PV_i 提供频谱资源所产生的单位成本参数。

1.3 联合优化目标表述

本文目标是在延迟约束下最大化 PV_s 和 SV_s 的系统总效益。对 SV_j 的频谱选择变量 η_j^i 进行了优化, 建立了决策问题 (MDP), 表示为:

$$\begin{aligned} \max_{\eta_j^i} U_{\text{tot}} &= \omega_1 \sum_{j=1}^J Ct_j + \omega_2 \sum_{i=1}^I Rv_i \\ \text{s.t.} \quad &\begin{cases} \eta_j^i \in \{0,1\} \\ 0 \leq S \leq S_{\max} \\ \sum_{i=1}^I s_j^i \leq S \\ \sum_{i=1}^I s_j^i = D \\ P_j^{\min} \leq P_j \leq P_j^{\max} \end{cases} \end{aligned} \quad (15)$$

式中: ω_1 和 ω_2 表示控制参数, 控制成本和收入之间的权重关系; η_j^i 表示 SV_j 是否向 PV_i 购买频谱资源的选择变量; S 表示 PV 所能提供的资源量, 其不超过其最大可用资源; $\sum_{i=1}^I s_j^i$ 表示为 SV 购买的总的频谱资源量不能超过 PV 所能提供的总的资源量 S , 并且 SV 购买的总的频谱资源量要满足其需求 D 。频谱资源的价格 P_j 必须在最低价格 $P_j^{\min}$ 和最高价格 $P_j^{\max}$ 之间。

2 基于 SAC 强化学习的频谱资源分配算法

强化学习在动态认知车联网环境序贯决策问题中展现出独特优势, 其通过智能体与复杂未知环境的持续交互, 借助试错机制动态优化累积奖励最大化的频谱选择策略^[8]。

2.1 马尔科夫决策 (MDP) 过程模型建立

本文将频谱资源交易过程建模为一个具有无限期折扣因

子的连续状态空间马尔可夫决策过程^[9]。该 MDP 系统架构由以下核心组件构成:

(1) 状态空间 S 设计

第 t 个时隙的系统状态被表示为 $S_t \in S$ 。状态 $S(t)$ 是一个联合状态空间, 包括 PV 所能提供的总的资源量 S , SV 的频谱资源需求量 D , 频谱资源价格 P_j 以及实际通信环境中的信干噪比 SNR。因此, 状态空间可以描述为:

$$S_t = [S, D, \text{SNR}, P_j] \quad (16)$$

(2) 动作空间 A 设计

在第 t 个时隙的动作 $a_t \in A$ 是系统状态 S_t 的一系列决策。该决定为 SV 是否向 PV 购买频谱资源的选择变量 η_j^i 。因此, 每个决策周期中的动作可以描述为:

$$a_t = (\eta_j^i) \quad (17)$$

式中: $\eta_j^i=1$ 表示 SV 购买了 PV 的频谱资源, 即 SV 选择占用 PV 的空闲信道; $\eta_j^i=0$ 表示 SV 不购买 PV 的频谱资源, 即 SV 选择不占用 PV 的空闲信道。

(3) 奖励函数 R_t 设计

在强化学习中, 奖励函数的设计直接影响模型的收敛特性。本文创新性地将约束条件以惩罚项形式加入奖励函数设计中, 使算法在优化过程中能够同步满足既定约束要求。若 SV 购买的总的频谱资源量 s_j^i 超过 PV 所能提供的总的资源量 S 且不满足资源量的需求 D 时, 资源供需惩罚项设计为:

$$S_p = \left[\sum_{i=1}^I s_j^i - D \right]^2 + \left[\sum_{i=1}^I s_j^i - S \right]^2 \quad (18)$$

式中: $\left[\sum_{i=1}^I s_j^i - D \right]^2$ 表示为 SV 购买的频谱资源量 $\sum_{i=1}^I s_j^i$ 与其频谱资源需求量 D 之间的差值, 若不满足或超过其频谱资源需求量 D , 都会获得一个惩罚值, 用于鼓励 SV 购买的总的频谱资源量满足资源量的需求 D ; $\left[\sum_{i=1}^I s_j^i - S \right]^2$ 表示为 SV 购买的总的频谱资源量 $\sum_{i=1}^I s_j^i$ 若超过 PV 所能提供的总的资源量 S , 也会获得一个惩罚值。

因此, 本文使用权重 ζ_1 和 ζ_2 对总效益项和惩罚项进行线性求和, 表示不同指标之间的平衡。因此, 奖励函数 R_t 可以表示为:

$$R_t = \zeta_1 \cdot U_{\text{tot}} - \zeta_2 \cdot S_p \quad (19)$$

式中: ζ_1 和 ζ_2 是平衡总效益项 U_{tot} 和惩罚项 S_p 的权重。

2.2 算法框架

SAC 算法的核心运作机制如下: 首先构建双神经网络架构, 一方面搭建 Policy 策略网络用于生成近似行为策略; 另一方面构建 Soft Q 价值网络对策略价值进行评估^[10]。其次, 将认知车联网运行过程中各时刻产生的状态特征、执行动作

及对应奖励值存入经验记忆库, 最终通过优化网络损失函数实施反向训练, 迭代更新网络参数以获取性能最优的资源分配策略。从优化目标来看, 传统强化学习算法旨在寻求最优策略 π , 使训练集的累加奖励值达到最大化。而 SAC 算法在此基础上加入了熵最大化, 其优化目标不仅要求训练集的累积奖励值最大化, 还通过最大化策略熵来增强策略的探索能力, 即在追求高回报的同时保持策略的随机性, 避免陷入局部最优解。用公式表示为:

$$\pi = \operatorname{argmax} \sum_{t=0}^T E_{(s_t, a_t) \sim \rho_{\pi}} [r_t + \alpha \cdot H(\pi(s_t))] \quad (20)$$

式中: H 表示策略熵, 表示当前策略的随机化程度; s_t 表示当前环境的状态; a_t 表示对当前状态做由 Bellman 一般递推方程可以得到强化学习中判断行为策略价值大小的函数表达式为:

$$Q(s_t, a_t) = r(s_t, a_t) + \gamma \cdot E_{s_{t+1}, a_{t+1}} [Q(s_{t+1}, a_{t+1})] \quad (21)$$

式中: s_{t+1} 表示环境的下一个状态; a_{t+1} 表示对应下一个状态的动作。

为实现最大熵, 在实际算法中需要将熵作为奖励的一部分, 在计算每一个状态的 Q 值的同时计算状态对应的熵。计算公式为:

$$Q_s(s_t, a_t) = r(s_t, a_t) + \gamma E_{s_{t+1}, a_{t+1}} [Q_s(s_{t+1}, a_{t+1}) - \alpha \cdot \ln \pi(a_{t+1} | s_{t+1})] \quad (22)$$

表示取自记忆库的下一个状态和取自当前策略的下一个动作的期望值。

2.2.1 策略网络 (Policy Network)

构建作为 Actor 网络的 Policy 网络, 其功能是生成动作策略。该网络以环境中状态信息作为输入, 具备双重输出机制: 一是输出为 A 的结果, 表示在训练集对应的动作选择空间中, 每种动作的选取概率分布; 二是基于上述概率分布进行采样, 最终确定并输出具体执行的动作。用公式表示为:

$$A_t = \pi_{\phi}(s_t) \quad (23)$$

Policy 网络的损失函数具体如下:

$$J_{\pi}(\phi) = E_{s_t \sim D} [\pi(s_t) \sum_{a=1}^A [\alpha \cdot \ln \pi(s_t) - \min(Q_1(s_t), Q_2(s_t))]] \quad (24)$$

式中: 选取 $Q_1(s_t)$ 和 $Q_2(s_t)$ 的原因来缓解 Q 值估计过高的问题。

2.2.2 双 Soft Q 网络

在该算法的价值评估模块中构建了包含当前价值网络与目标价值网络的双 Soft Q 网络架构, 都以观察到的状态信息 S 及动作选择 a 构成的状态-动作对作为输入, 并通过参数化的函数形式输出动作空间中每个候选动作对应的价值估计。目标 Q 值计算为:

$$Q_t[s_t, a_t] = r[s_t, a_t] + \gamma E_{s_{t+1} \sim \rho} \pi(s_{t+1}) \{ \min[Q_1(s_{t+1}), Q_2(s_{t+1})] - \alpha \cdot \ln \pi[s_{t+1}] \} \quad (25)$$

式中: α 表示温度熵系数; $\pi(s_t)$ 表示动作策略 π 下动作空间中每一种动作出现的概率。

Soft Q 网络的损失函数用公式表示为:

$$J_Q(\theta) = \frac{1}{m} \sum_{i=1}^m [Q_{\theta}(s_t, a_t) - Q_t(s_t, a_t)]^2 \quad (26)$$

为计算更准确的行为价值, 采用一种基于温度熵系数 α 的动态调整策略, 即不将其设为定值, 而是根据网络学习情况对温度熵系数进行动态调节, 损失函数表示为:

$$J(\alpha) = -E_{s_t \sim S} [\alpha \cdot (\sum_{a=1}^A \pi(s_t) \cdot \ln \pi(s_t) - K \cdot \ln \frac{1}{A})] \quad (27)$$

式中: A 为行为空间大小; K 为目标熵系数。

3 仿真分析

本章旨在通过 Python 仿真实验, 全面评估所提出的车联网频谱资源分配策略以及基于声誉的定价机制的性能。实验设定 5 辆拥有可用频谱信道的 PV_s, 3 辆占用频谱信道的 SV_s, 通过协商基于声誉值的定价获取空闲频谱资源。系统带宽 $B=1$ MHz, $\sigma=10^{-14}$ W/Hz, 单个 SV 的发射功率为 20 mW, 单个 PV 的发射功率为 40 mW。为证明本文基于声誉值频谱共享方案有效鼓励了 PV 和 SV 之间的共享, 评估了 SAC、Q-learning、A2C 三种算法下系统总效益、PV 吞吐量、SV 吞吐量的对比, 并且还对比了不同的声誉值对 SV 频谱出价的影响。

如图 2 所示, 本文中所利用的 SAC 算法相比较 Q-learning、A2C 算法在系统总效益方面具有很大的提升, 在经过将近 200 次的迭代后, SAC 算法下的系统总效益值趋近于收敛, 并且显著高于 Q-learning、A2C 算法下的系统总效益。

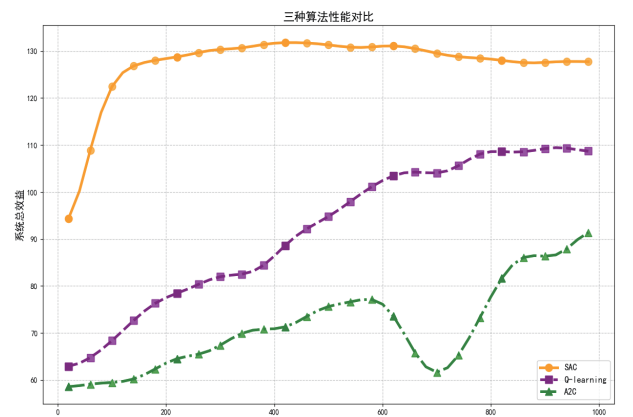


图 2 系统总效益

如图 3 所示, 仿真模拟分析了本文中所利用的 SAC 算法相比较 Q-learning、A2C 算法在 PV 的吞吐量方面具有很大的提升, 可以看出, 考虑 PV 吞吐量的 SAC 算法经过将近 150 次的迭代后稳定上升, 相比较于 Q-learning、A2C 算法具有更高的吞吐量, 并且趋势稳定上升。

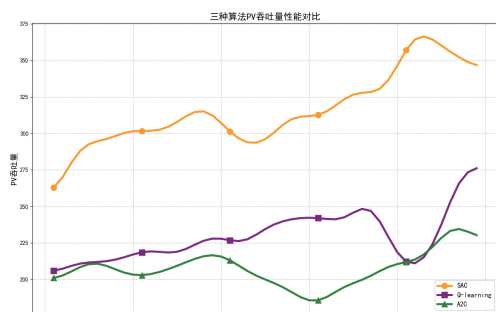


图3 PV吞吐量

如图4所示, 仿真模拟分析了本文中所利用的 SAC 算法相比较 Q-learning、A2C 算法在 SV 的吞吐量方面具有很大的提升, 可以看出, 考虑 SV 吞吐量的 SAC 算法经过将近 200 次的迭代后趋近于收敛, 相比较于 Q-learning、A2C 算法具有更高的吞吐量, 并且趋势较为平缓稳定。

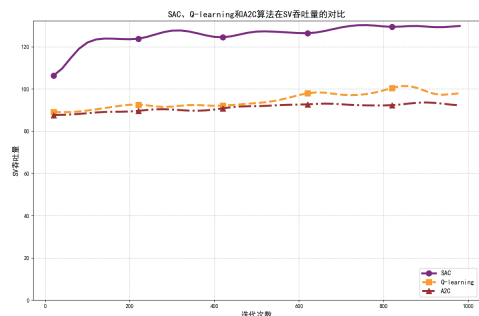


图4 SV吞吐量

图5表明了基于 EMA 的声誉值对 SV 定价方案的影响。数据表明, 声誉值较高的 SV 随着交互次数的增加能够以更低出价购入频谱资源, 验证了声誉值的激励效应, 在此机制下, SV 会主动追求更高声誉值以获取资源交易的定价优势, 证明了基于 EMA 的声誉值体系在 SV 定价方案中的可行性与有效性。

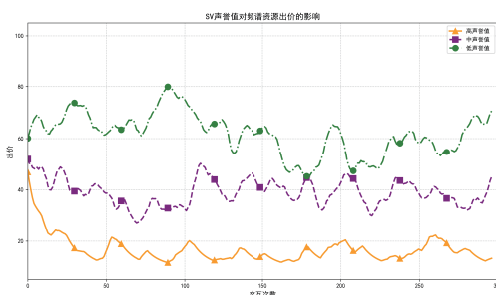


图5 SV不同声誉值的出价

4 结语

本研究提出了一种基于声誉值的动态频谱资源定价与分配模型, 应用于认知车联网环境。通过利用基于 EMA 的声誉评估方法和 SAC 算法, 显著提升了系统总效益和用户吞吐量。所提出的模型有效解决了认知车联网中的频谱资源短缺

问题。实验结果表明, 与 Q-learning 和 A2C 算法相比, 本文采用的 SAC 算法在系统总效益及主用户吞吐量的指标下收敛速度更快, 性能更好。

参考文献:

- [1]LI H N, LI J Y, ZHAO H Y, et al. Blockchain-based incentive mechanism for spectrum sharing in IoV[EB/OL].(2022-04-28) [2026-01-22].<https://doi.org/10.1155/2022/6807257>.
- [2]WANG B B, LIU K J R. Advances in cognitive radio networks: a survey[J]. IEEE journal of selected topics in signal processing, 2011, 5(1): 5-23.
- [3]FELICE M D, CHOWDHURY K R, BONONI L. Analyzing the potential of cooperative cognitive radio technology on inter-vehicle communication[C/OL]//2010 IFIP Wireless Days. Piscataway:IEEE,2010[2026-01-22].<https://ieeexplore.ieee.org/document/5657770>.DOI:10.1109/WD.2010.5657770.
- [4]FAWAZ K, GHANDOUR A, OLLEIK M, et al. Improving reliability of safety applications in vehicle ad hoc networks through the implementation of a cognitive network[C]// 2010 17th International Conference on Telecommunications, Piscataway: IEEE, 2010: 798-805.
- [5]MOHSENI M, ADVE R S. Inter-operator spectrum sharing model and economic analysis[C]//2021 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN). Piscataway: IEEE, 2021:185-192.
- [6]吴桐, 袁培燕. 一种基于指数移动平均的物联网边缘设备选择机制 [J]. 应用科学学报, 2025,43(1):183-194.
- [7]温雯, 胡智鑫, 郝志峰. 面向序列推荐的深度化指数移动平均学习方法 [J]. 计算机科学与探索, 2025,19(3):774-786.
- [8]黄煜梵, 彭诺蕻, 林艳, 等. 基于 SAC 强化学习的车联网频谱资源动态分配 [J]. 计算机工程, 2021,47(9):34-43.
- [9]蓝思明. 强化学习算法的泛化性研究 [D]. 合肥: 中国科学技术大学, 2024.
- [10]LI M F, LIU H B, XIE F, et al.Adaptive distributed control for leader-follower formation based on a recurrent SAC algorithm[J].Electronics,2024,13(17):3513.

【作者简介】

陈一凡 (2001—), 男, 安徽亳州人, 硕士研究生, 研究方向: 认知车联网频谱共享。

陈玲玲 (1980—), 通信作者 (email:cll807900@163.com), 女, 吉林长春人, 博士, 教授, 研究方向: 认知无线电。

王一帆 (1999—), 女, 陕西渭南人, 硕士研究生, 研究方向: 无人机通信安全。

(收稿日期: 2025-07-22 修回日期: 2026-02-12)