

基于深度 Q 学习的蜂窝车联网边路资源分配算法

许辉

(中兴通讯股份有限公司终端事业部, 四川 成都 610041)

摘要: 针对蜂窝车联网系统中不同优先级业务的边路资源自主选择分配, 分析了基于参考信号能量的自主选择算法流程, 设计了能量门限方程; 针对能量方程参数估计问题, 将基于能量的自主选择算法与深度 Q 学习算法结合, 通过有限次算法迭代得到能量门限方程的最优参数值。仿真结果表明, 基于深度 Q 学习的边路资源分配算法可以满足不同优先级车联网业务的边路资源需求, 同时提高系统分组接收率。

关键词: 蜂窝车联网; 边路; 自主资源选择; 机器学习; 分组接收率

中图分类号: TN929.5

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202309

Sidelink resource allocation algorithm of C-V2X based on deep Q learning

XU Hui

Terminal Business Division, ZTE Corporation, Chengdu 610041, China

Abstract: For the sidelink resource autonomous selection scheme of different priority services in cellular-vehicle to everything (C-V2X) system, the procedure of autonomous selection algorithm based on reference signal energy was analyzed, and the energy threshold equation was designed. For the energy equation parameter estimation problem, energy-based autonomous selection algorithm was combined with deep Q learning algorithm, and the optimal parameter value of the energy threshold equation was obtained by iteration of the finite-degree algorithm. Simulation results showed that the side resource allocation algorithm based on deep Q learning could ensure the side resource requirements of V2X services with different priorities, and improve the packet reception ratio performance of the system.

Key words: C-V2X, sidelink, autonomous resource selection, machine learning, PRR

0 引言

车联网技术是智能交通系统发展的主要支撑技术^[1]。车联网技术以网联车辆为移动感知对象, 在通信网络中实现多系统交互, 感知道路交通实时状况, 使智能交通平台对道路和车辆进行全程实时控制, 从而提高交通效率和交通安全性。蜂窝车联网 (cellular-vehicle to everything, C-V2X) 是融合蜂窝通信与设备到设备 (device to device, D2D) 直通信的车联网通信技术^[2]。

在基于移动通信系统的 C-V2X 中, 为了区别传统的用户设备与基站之间的上行链路和下行链路, 车辆终端与其他节点 (除基站) 之间的通信链路被称为边路 (sidelink)。为了满足不同车联网业务的需求, 要对边路无线资源进行合理分配^[3]。

针对车联网边路资源分配问题, 研究人员提出了各自的观点。参考文献[4]提出了资源分配与基于博弈论的任务卸载决策联合优化策略, 提高了系统总增益。参考文献[5]中每个车辆根据信道感知结果独立选择边路传输资源, 并保留所选资源供今后使

收稿日期: 2023-01-23; 修回日期: 2023-02-24

通信作者: 许辉, xu.hui6@zte.com.cn

基金项目: 国家重点研发计划资助项目 (No.2018YFB1308600)

Foundation Item: The National Key Research and Development Program of China (No.2018YFB1308600)

用,提高了资源分组接收率(packet reception radio, PRR)。为了提高车联网通信可靠性,参考文献[6]提出了一种调整车辆广播速率的资源控制模型,以提高信道拥塞环境下车联网通信的可靠性。参考文献[7]结合感知型半持续调度和强化学习等理论,研究了分布式无线资源调度,并结合发射功率控制和非合作博弈等理论研究了分布式拥塞控制。参考文献[8]通过联合优化任务执行位置,对基于精英选择策略的遗传算法做了部分改进,设计了边路资源分配算法,降低了系统成本。参考文献[9]提出了一种位置感知任务加载配置策略,降低了边路资源分配处理时间及网络时延。在多频段协作环境中,参考文献[10]采用深度Q网络学习算法降低5G基站部署的复杂性。参考文献[11]提出了一种基于滚动时域强化学习(receding horizon reinforcement learning, RHRL)的侧向控制方法。车辆的侧向控制量由前馈和反馈两部分构成,前馈控制量由参考路径的曲率及动力学模型直接计算得出;反馈控制量通过采用滚动时域强化学习算法求解最优跟踪控制问题得到。该文献将无限时域最优控制问题转化为若干有限时域控制问题进行求解。参考文献[12]针对连续动作空间下的自动驾驶换道决策,采用改进的深度强化学习算法双延迟深度确定性策略梯度算法(twin delayed deep deterministic policy gradient algorithm, TD3)构建换道模型控制自动驾驶汽车的换道驾驶行为。参考文献[13]提出了基于改进TD3算法的车道保持端到端决策新方案,构建多数据融合TD3算法框架,设计了指导型奖励函数,采用分类与高价值优先级经验回放方法,提高样本利用率,加快算法收敛速度。

上述参考文献根据不同优化目标,针对边路资源分配及车辆控制问题,分别提出了不同的解决方案。但现有研究内容未考虑车联网通信环境中实际的业务部署情况,且边路资源分配决策是个多目标优化问题,需要综合考虑不同优先级业务的边路资源需求、系统分组接收率性能等因素。对于一个车辆发送的数据包,PRR由 X/Y 计算,其中 Y 是该车辆(w, l)范围内的车辆数量, X 是 Y 中成功接收到数据包的车辆数量。针对边路资源分配中提高分组接收率、不同优先级业务资源分配区分处理等要求,现有资源分配方案效果不佳,还需继续优化。

本文以C-V2X通信环境中用户设备(user equipment, UE)自主选择边路资源的通信模型^[14]

为基础,设计了用于边路资源自主选择的能量门限方程,在门限方程中同时考虑占用资源业务优先级和待发送业务优先级,利用深度Q学习(deep Q learning)算法确定能量门限方程的最优参数,设计了以提高车联网中系统PRR为目标的回报函数,以有效利用车联网系统边路资源,保障不同优先级车联网服务。

本文首先介绍UE自主资源选择算法的原理和流程,然后提出深度Q学习和自主选择边路资源结合的算法实现方案,最后给出了仿真场景和仿真结果。

1 自主资源选择

本文采用的基于D2D技术的C-V2X架构如图1所示。

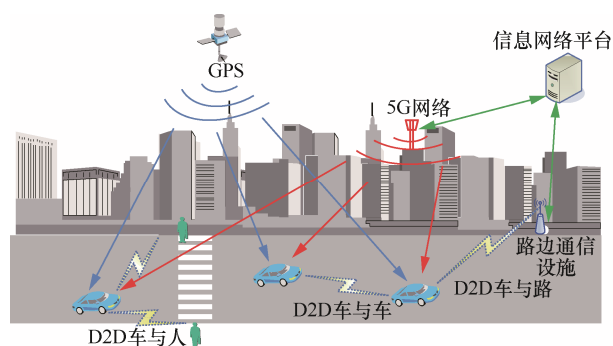


图1 本文采用的基于D2D技术的C-V2X架构

用户设备自主资源选择方案可应用于无网络覆盖或基站不支持C-V2X的环境,适用于所有的车联网场景^[15]。用户设备进行边路自主资源选择时,会基于一定的资源选择规则自主地从资源池中选择资源,以传输边路车联网控制消息和数据。

1.1 自主选择资源算法原理

为了避免不同UE自主选择的资源产生冲突,在V2X UE选择确定使用的资源之前,UE需要在资源检测窗口内对边路可用资源进行检测,判断出其中已被其他UE占用的资源以及可能空闲的资源。然后,在满足边路数据发送需求的资源选择窗口中选择合适的资源进行发送。边路业务具有一定的周期性,其资源调度分配(scheduling assignment, SA)信息中包含对后续资源使用的预留指示^[16];发送UE能够通过检测窗口内检测到的边路信号及信息来预测后续资源选择窗口中的资源使用情况,并基于此选择资源,达到尽可能避免资源选择冲突的目的。

每个需要传输的边路数据都有一个对应的 SA 边路资源控制信息，SA 信息包含边路数据的资源配置信息、优先级信息等。SA 和相应的车联网数据在相同的子帧中传输。

根据 SA 信息中携带的业务优先级信息，当其他 UE 检测到该资源控制信息时，根据彼此待传业务之间的优先级高低来选择是否避让该资源。

1.2 自主资源选择算法流程

UE 可以基于以下 3 种方法自主选择边路资源：能量感知 (sensing)；随机选择；地理位置信息。

其中，随机选择指 UE 在发送资源池中随机选择所需资源，不考虑该资源是否可用；地理位置信息指根据车辆位置和与地理位置相关的资源数据库选择边路资源。

本文采用的能量感知算法的流程如图 2 所示，可以分为三大步骤^[17]：步骤 1，获取所有可用资源；步骤 2，根据 SA 信息和能量感知进行资源排除；步骤 3，对剩余资源进行排序，从排序结果中随机选择一个资源。

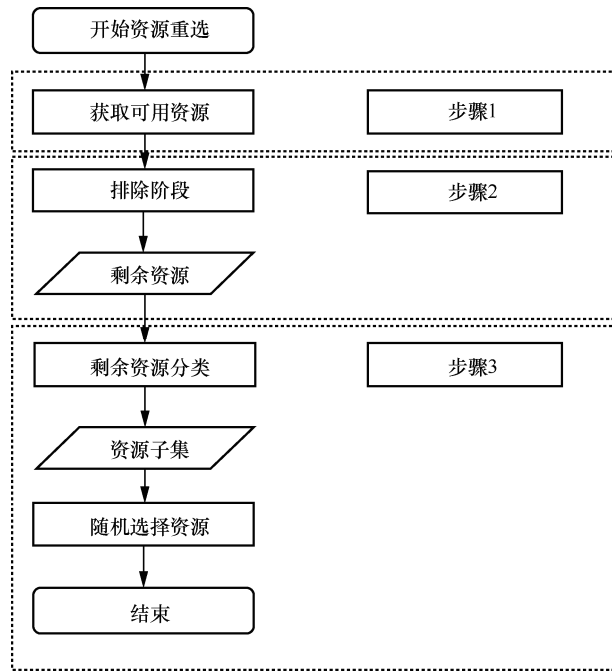


图 2 能量感知算法流程

2 基于 Q 学习的资源分配算法

本节内容针对能量感知算法中的步骤 2，即资源排除阶段进行阐述。在自主选择资源的能量感知方案中对可选择边路资源进行评估，为了排除其中

已被占用的资源，需要设计一个能量门限，UE 测量的资源能量高出该门限的资源被排除。具体的门限值由门限方程计算给出。下面给出能量门限方程，并采用深度 Q 学习算法选取最优的能量门限方程参数，从而在保障高优先级 V2X 业务资源需求的同时，最大化 PRR。

2.1 能量门限方程

3GPP 标准未给出门限方程的具体形式，只给出了设计原则：门限方程是两个优先级的函数，两个优先级指占用资源的 V2X 业务优先级和待发送 V2X 业务优先级。需设计一个满足标准规则的门限方程，从而在资源选择的时候区别对待不同优先级业务。假设优先级数值越高，代表优先级越高，共有 8 个优先级， $p=8$ 为最高优先级， $p=1$ 为最低优先级。

根据标准中的原则，本文设计如式 (1) 所示的能量门限方程：

$$T = b + a \times 10 \times \ln \left(\frac{1 - e^{-p_1}}{1 - e^{-p_2}} \right) \quad (1)$$

式 (1) 是待发送车联网业务优先级 p_1 的增函数，即当待发送车联网业务优先级越高时（如 $p_1 > p_2$ ），门限值 T 越高，被排除的资源越少，可供待发送的较高优先级业务选择的资源越多；式 (1) 是占用资源的车联网业务优先级 p_2 的减函数，即当占用资源的车联网业务优先级越高时（如 $p_2 > p_1$ ），门限值 T 越低，被排除的资源越多，可供待发送的较低优先级业务选择的资源越少。

式 (1) 中 b 为系统中接收能量的均值，因为门限 T 要和接收能量进行比较，所以这里选择系统平均接收能量作为门限的基础值； a 为一个大于 0 的变量， a 值的大小影响不同优先级业务抢占边路资源的比例和分组接收率的大小。通过设置不同的 a 值，得到不同的门限值 T ，根据门限值 T 可仿真得到相应的分组接收率。下面采用深度机器学习算法获取最优的参数 a 取值，即通过选取合适的 a ，实现对不同优先级业务资源的合理分配，同时获取较高的分组接收率。

2.2 深度 Q 学习

在门限方程中， a 的取值影响不同优先级业务的资源选择，同时要在保证一定的优先级区分度的情况下最大化系统的 PRR 值，这属于多目标优化问题。仿真中计算 PRR 时， $w=0$ ， $l=$ 高速公路 320 m

基线, 城市 150 m 基线。

假设每次边路资源分配与当前车联网服务需求相关, 具有无记忆性, 属于马尔可夫过程, 可利用深度机器学习算法迭代得到 a 的最优取值。

机器学习中的 Q 学习算法使用时间差分 (time difference, TD) 法进行离线学习, 可以对马尔可夫决策过程 (Markov decision process, MDP) [18] 求解最优策略。在 Q 学习中, 用行动价值函数 $Q(s, a)$ 表示每一对状态-动作 (state-action) 的效用。

Q 学习算法的核心是根据旧的 Q 值和新的 Q 值估计进行权重平均的一个值迭代更新, Q 就是在某一时刻 s 状态下 ($s \in S$), 采取动作 a_t ($a_t \in A$) 能够获得收益的期望, 环境会根据动作反馈相应的回报 r_t , 根据回报来选取能够获得最大收益的动作。当将深度学习作为估计 Q 值函数的函数逼近方法时, 称之为深度 Q 学习或深度 Q 网络 (deep Q network, DQN)。目前强化学习发展迅速, 如离线策略 (off-policy) 算法 SAC (soft actor-critic), 在无模型的强化学习算法中, SAC 是一个非常高效的算法, SAC 算法可解决离散动作空间和连续动作空间的强化学习问题。TD3 和 SAC 的网络结构类似, 也是 off-policy 的强化学习算法, TD3 是一种确定性策略强化学习算法, 适合高维连续动作空间。考虑到算法实现的复杂度和车辆运行的可靠性要求, 本文的能量方程参数估计并未采用 SAC/TD3 等算法, 而是采用了实现更为简单和成熟的深度 Q 学习算法。分析表明, 通过变换策略函数的形式, 本文也可以采用 SAC/TD3 等离线策略算法与边路资源自主选择相结合的方法。

2.3 基于深度 Q 学习的边路资源分配算法

Q 学习的主要因素包括: 动作集合、状态集合、折扣因子、学习因子、 Q 表以及回报函数。本方案中算法部分的主要内容就是合理地选择智能体, 设计动作集合、状态集合、折扣因子、学习因子以及回报函数, 达到让智能体可以快速收敛到最优结果的目的。

智能体是 Q 学习中的行为主体。本研究中选择 C-V2X 网络系统为智能体, 目标是为系统选出最优的门限方程参数 a , 在 Q 表中选取 PRR 值作为 Q 值, 即选取最优参数 a , 使得 PRR 最大化。动作集合和状态集合是智能体可能选取的动作和选择动作之后所处的状态的集合。当智能体选择一个动作 a_t 时, 根据系统的反馈, 智能体会转换到一个状态

s , 此时经过回报函数计算就会得到一个回报值, 用这个回报值去更新 Q 表。 Q 表以动作和状态对的形式存在, 是每个动作和状态组合的评估值, 也是选择动作的依据。

深度 Q 网络的时间差分方程如式 (2) 所示:

$$Q(\underset{\text{状态 动作}}{s_t, a_t}) \leftarrow (1 - \alpha) \times \underset{\text{当前值}}{Q(s_t, a_t)} + \underset{\text{学习值}}{\alpha \times \left(\underset{\text{回报}}{r_t} + \underset{\text{折扣}}{\gamma} \times \underset{\text{最优未来值估计}}{\max_a Q(s_{t+1}, a)} \right)} \quad (2)$$

式 (2) 就是 Q 学习算法迭代更新的计算式, 根据下一个状态 s_{t+1} 选取最优的 Q 值乘以衰变 γ 加上回报值作为 Q 学习值。机器学习开始阶段, 智能体会随机选择动作 a_t , 然后用每一次动作 a_t 的反馈回报 r_t 值去更新 Q 表, 式 (2) 为 Q 表的更新函数, 智能体根据最大 Q 表值的规则选择动作 a_t 。若干次学习后 Q 表中的 PRR 会达到收敛, 此时动作选择也会达到收敛, 并得到最优动作, 根据最优动作 a_t 和辅助参数 c 即可得到最优参数 a , 辅助参数 c 如式 (3) 所示, 状态集合 s 如式 (4) 所示。

参数 a : $a = 25 + (a_t - 5) \times c$,

$$c = \frac{a_{\max} - a_{\min}}{a_{t\max} - a_{t\min}} \quad (3)$$

状态集合:

$$s = \begin{cases} s0(\text{var} \leq \text{var}0) \\ s1(\text{var} > \text{var}0) \end{cases} \quad (4)$$

其中, var 为选择当前动作 a_t 得到的资源抢占比例的方差, $\text{var}0$ 是要满足的门限; 不同优先级的区分度用各个优先级业务资源抢占比例的方差 var 表示。方差 var 大于规定门限 $\text{var}0$, 表示当前动作具有较高的优先级区分度, 状态设置为 $s1$ 。不同优先级业务抢占资源的区分度越大, 不同优先级业务资源越可能得到保障。优先级的区分度用各个优先级被抢占比例的方差表示, 方差大于一个规定门限值表示可以实现区分不同优先级的目的。

门限 $\text{var}0$ 是本文深度 Q 学习算法的约束条件。不同的 $\text{var}0$ 可得到不同的最优参数 a , 以及不同的算法收敛的 PRR 结果。算法的目标是获取最大可收敛的 PRR 值。

为了提高边路分组接收率性能, 设计回报函数 r_t , 如式 (5) 所示:

$$r_t = \begin{cases} 1 - e^{\frac{-(u_1 - u_0)}{u_{\max} - u_0} \times 10} & (s = s_1) \\ -1 & (s = s_0) \end{cases} \quad (5)$$

其中, u_1 是当前动作得到的 PRR 值, u_0 是平均 PRR 值, $u_{\max}$ 是最大 PRR 值。回报函数的目标是在不同优先级业务抢占边路资源时, 有利于高优先级业务资源选择的动作 a_t 得到的 PRR 较高, 回报 r_t 较高, Q 值较大, 实现了在满足不同优先级的占有资源业务和待发送业务的资源分配条件下, 提高边路传输的分组接收率的目的。如果当前动作的 PRR 较高, 且不同优先级业务抢占资源的区分度较大, 回报 r_t 就较高; 如果当前动作得到的 PRR 较低, 则回报 r_t 较低。

本文 DQN 算法的流程如图 3 所示。

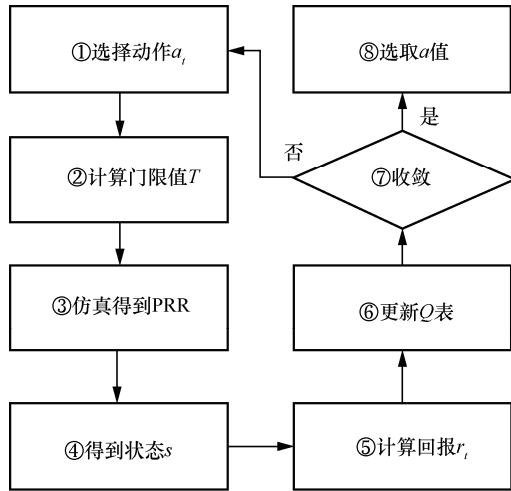


图3 本文 DQN 算法的流程

深度 Q 学习算法的目标是: 不断更新动作、状态, 为系统选出最优的能量门限方程的参数。

本文 DQN 算法流程的步骤说明如下。

- 步骤②根据步骤①选择的动作 a_t , 利用式(3)得到 a , 分别设置业务优先级 p_1 和 p_2 , 利用式(1)计算得到相应的能量门限值 T 。
- 步骤③根据步骤②计算的门限值 T , 对不同场景仿真, 得到相应的分组接收率 PRR 和不同优先级业务抢占资源的方差 var 。
- 步骤④根据步骤③得到的方差 var , 利用式(4)得到当前的状态 s ; 其中不同优先级业务资源的区分度越大, 状态越好。
- 步骤⑤根据步骤③和④得到的 PRR 和 s , 利用式(5)计算得到回报 r_t 。其中 PRR 越大, 状态越好, 得到的回报越大, 从而实

现不同优先级业务的资源需求目标, 提高系统 PRR。

- 步骤⑥根据步骤①、④、⑤的 a_t 、 s 、 r_t , 利用式(2)更新 Q 表记录。
- 步骤⑦判断 Q 表是否收敛, 收敛是指 Q 表(即 (a, s) 值)取得稳定值。如果是, 则确定当前的动作 a_t 和参数 a 为最优值; 否则, 转向步骤①, 继续选择新的动作 a_t , 循环迭代上述过程。

算法伪码如下:

Input: action set $a_t = \{0, 1, 2, \dots, 10\}$, priority p_1 , $p_2 = \{1, 2, \dots, 8\}$, study factor $\alpha = 0.9$, discount $\gamma = 0.5$, resource variance $\text{var}_0 = 0.015$, constant $b = -87$;

1 Random initialization PRR, a ;

2 repeat

3 for $i = 1, \dots, N$,

do

4 choose sample a_t from action set, random choose priority p_1, p_2 ;

5 calculate PRR value and status s ;

6 calculate return function r_t ;

7 update PRR value.

8 end

9 until PRR is maximum and convergent; output: PRR, a .

3 算法评估

本节对深度 Q 学习算法进行仿真, 主要指标为系统 PRR 性能评估算法的有效性和对于不同优先级 C-V2X 业务的资源区分能力。

3.1 仿真场景

仿真中车辆部署场景有两种, 分别是城区场景和高速公路场景。图 4(a) 所示是城区场景道路配置, 图 4(b) 所示是高速路场景道路配置, 车辆位置每间隔 100 ms 更新一次^[19], 模拟车辆移动。本文中车辆自主选择边路资源场景, 即不需要路边单元 (road side unit, RSU) 参与资源分配, 因此仿真中没有设置 RSU。

仿真参数选择如下。假设城区场景范围最小为 $[1299 \text{ m} \times 750 \text{ m}]$, 水平、垂直各有双向 12 车道; 高速路长度 $\geq 2 \text{ km}$, 双向 6 车道; 假设相同车道内的平均车间距 $= 2.5 \text{ s} \times$ 车辆绝对速度, 仿真中假设所有车道具有相同的密度/速度^[13], 即仿真中车辆匀速

运动, 每个车辆周围的车辆数量可根据车辆速度和长、宽范围计算得到, 根据周围车辆数量和信息的资源需求, 即可得到相应的PRR值。 $\text{var}0$ 设为0.015。考虑到车辆以实际速度运行将导致实验的复杂性和资源分配算法难度提高, 为了简化仿真实现, 在仿真实验中, 假设所有的车辆匀速运动。

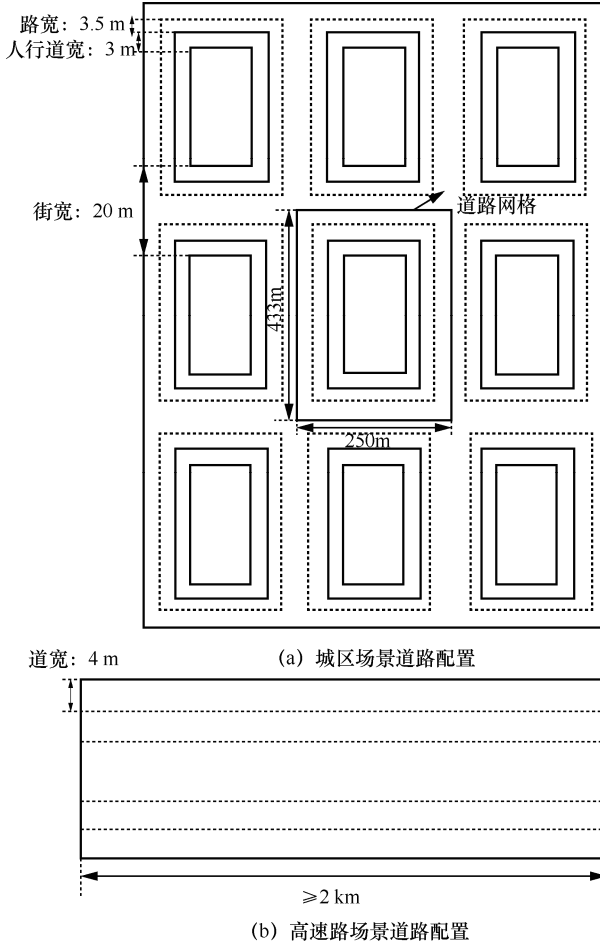


图4 不同场景下的道路配置

仿真中的参数设置见表1。假设边路SA在一个子帧中使用两个物理资源块 (physical resource block, PRB), SA在相应数据之前发送两次, SA和数据在相同的子帧上发送。数据包为190 B和300 B, 分别使用2个和3个连续子帧传输。

3.2 算法性能评估

图5(a)和图5(b)分别给出了城区环境中60 km/h和15 km/h的2种自主选择资源的分组接收

率累积分布函数 (cumulative distribution function, CDF) 仿真结果, 车辆间距离 $[0, 150]$ m, 平均车间距分别为: $2.5 \text{ s} \times 60 \text{ km/h} \approx 42 \text{ m}$, $2.5 \text{ s} \times 15 \text{ km/h} \approx 10 \text{ m}$ 。

仿真中PRR的CDF是 $w=0$, l =高速公路320 m基线, 城市150 m基线。其中PRR的CDF是指: 所有横坐标小于或等于 x 的PRR值出现概率的和, 即 $\text{CDF}_x = P(\text{PRR} \leq x)$, 例如在图5(a)中, 能量感知算法中PRR小于或等于0.7的比例是0.3, 即有30%的车辆PRR小于或等于0.7。

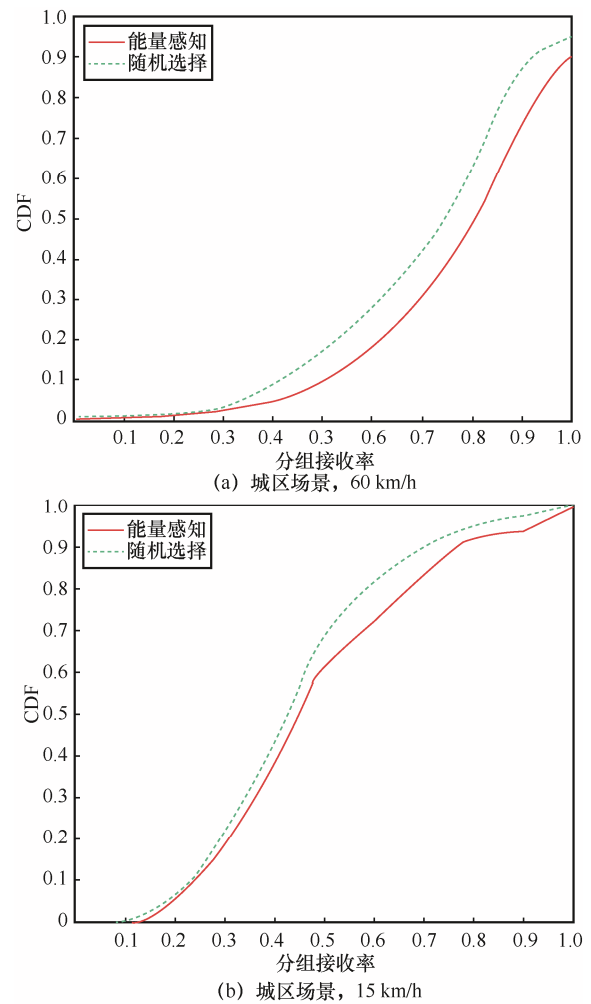


图5 城区场景仿真结果

两种场景下, 能量感知方案性能均优于随机选择方案, 因为能量感知降低了资源冲突概率, 从而提高了分组接收成功率。车辆密度较低时 (如车速为60 km/h), 本身碰撞概率较低, 整体分组接收

表1

仿真参数

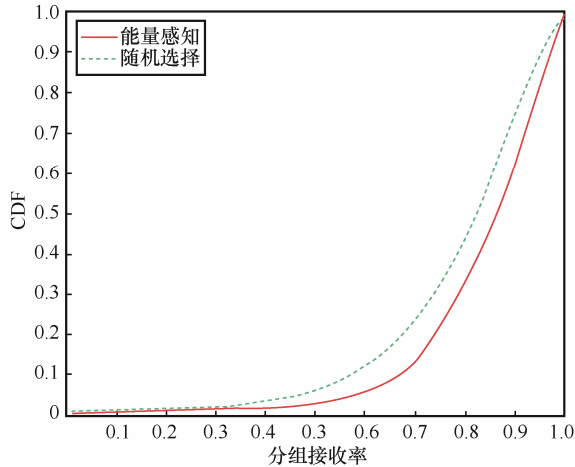
参数	优先级 $p1, p2$	动作集合 a_i	b	折扣 γ	学习因子 α	资源分配信息 SA	子信道	数据包
值	{1,2,3,4,5,6,7,8}	{0,1,2,3,...,10}	-87	0.5	0.9	30 bit (发两次)	10 PRB	190/300 B

率性能较好；车辆密度较高时（如车速为 15 km/h），碰撞概率进一步增大，分组接收率性能较差。低密度时，考虑 PRR 平均 CDF 性能，如式（6）所示：

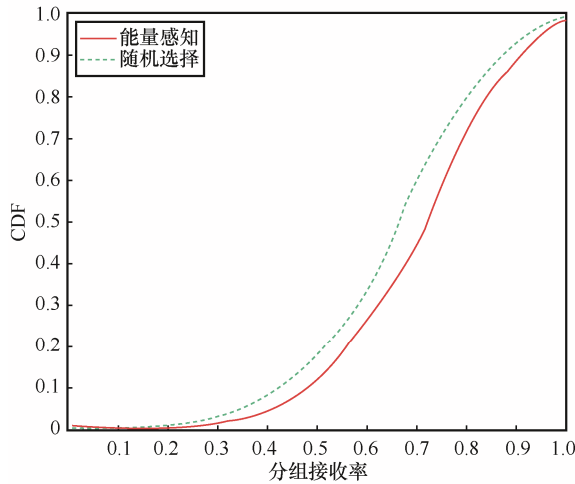
$$\delta = \frac{\text{average}(F_{\text{ran}}) - \text{average}(F_{\text{sen}})}{\text{average}(F_{\text{ran}})} \quad (6)$$

式（6）中 F_{ran} 表示随机选择算法 PRR 的 CDF， F_{sen} 表示能量感知算法 PRR 的 CDF。采用能量感知方案，平均分组接收率性能提高了 25%；高密度时，如 15 km/h，平均分组接收率提高了 8%。

图 6（a）和图 6（b）分别给出了高速路环境中 140 km/h 和 70 km/h 的 2 种自主选择资源的分组接收率累积分布函数仿真结果，车辆间距离 [0, 320] m，平均车间距分别为：2.5 s × 140 km/h ≈ 97 m，2.5 s × 70 km/h ≈ 48 m。



(a) 高速路场景，140 km/h



(b) 高速路场景，70 km/h

图 6 高速路场景仿真结果

高速路场景下车辆密度较低，分组接收率性能有所提高，且车速越高，车辆密度越低，分组接收

率性能越好。对比结果与城区场景类似，能量感知方案优于随机选择方案，能量感知降低了资源冲突概率，从而提高了分组接收成功率。低密度时，如 140 km/h，采用能量感知方案平均分组接收率性能提高了 22%；高密度时，如 70 km/h，平均分组接收率性能提高了 14%。

单独统计市区场景中车速为 60 km/h，车辆间距离 [0, 150] m 时，优先级 p_2 为 1 和 8 的 V2X 数据 PRR 与系统平均 PRR 累积分布的对比，如图 7 所示。高优先级车联网业务的平均分组接收率提高了 24%。

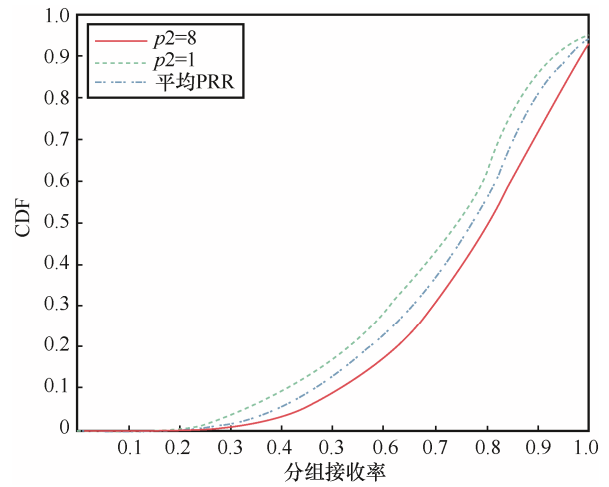


图 7 不同优先级业务的分组接收率

图 8（a）和图 8（b）所示为 a 取不同值时，不同优先级业务抢占资源的比例。

图 8 中横坐标为业务优先级，纵坐标为该优先级业务抢占资源的比例，不同优先级之间的纵坐标差值即低优先级资源被抢占的比例，差值越大，说明资源被抢占比例越高。可以看出业务优先级越高，其抢占资源的比例越高，优先级为 8 的业务抢占资源的比例接近 1。分析其原因：方程是关于待发送车联网业务优先级 p_1 的增函数， p_1 越大，门限值越高，资源排除可能性越小，即可抢占资源越多，资源抢占概率越高。 a 的值影响各优先级的比例分布，图 8（a）中， $a=22$ 时，优先级相差超过 4 的业务间的抢占比例较高，如优先级 5 比优先级 1 提高了 90% 以上，但相邻优先级业务抢占比例相近，不易区分，如优先级 2 比优先级 1 提高了 20%；图 8（b）中，相邻优先级业务间的抢占比例较高，如优先级 2 比优先级 1 提高了 55%，优先级差值超过 1 的抢占比例更高。 a 的大小会影响不同优先级业务资源的抢占比例，且随着 a 的增大，不同优先级业务的区分度增加。

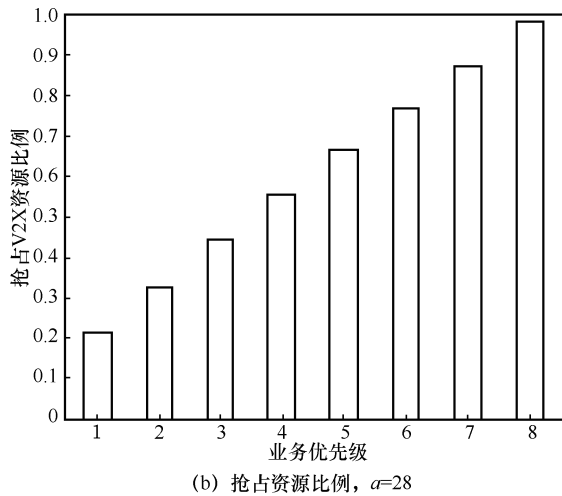
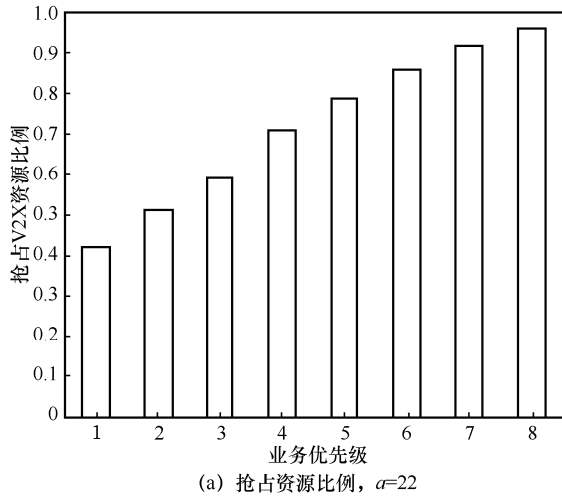
图8 不同 α 值时的抢占资源比例

图9所示是市区场景, 抢占比例方差设置为0.01时的机器学习仿真结果, 横轴为Q学习迭代次数, 纵轴为分组接收率。从仿真结果图中可以看出, 学习1400次之后动作和PRR值基本达到收敛, 并且可以证明系统环境的改变以及要满足的方差阈值(代表着不同优先级的区分度)的改变也会影响系统的收敛结果, 从而导致系统收敛到另外的值。

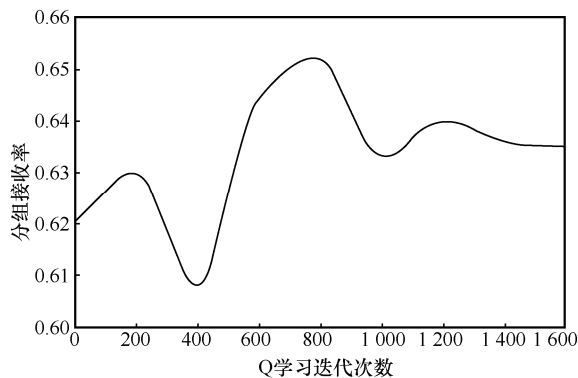


图9 机器学习迭代结果

图10的实线与虚线分别是深度Q网络机器学习和对比数据的PRR累积分布函数图。其中深度Q网络学习得到的最优 $\alpha=28$, 对比数据为参数 $\alpha=[10, 20, 30, 40]$ 时的算术平均PRR性能。仿真中 α 取值范围为 $[10, 40]$ 。

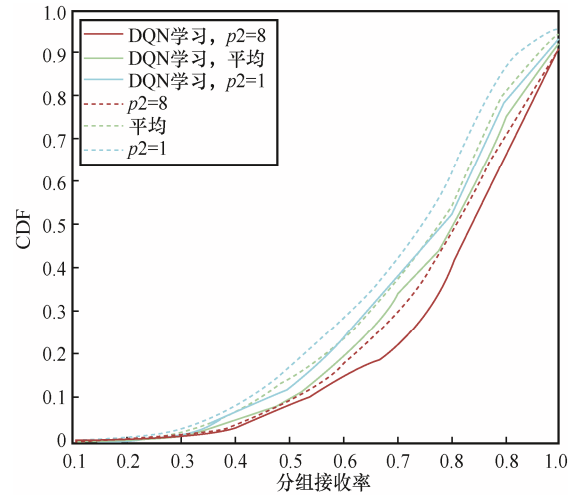


图10 不同优先级下的DQN算法性能

通过图10可以看出, 对于高优先级V2X业务, 采用DQN的平均分组接收率有11%的提升; 对于低优先级V2X业务, 采用DQN算法的平均分组接收率有12%的提升。仿真中发现, 在不同优先级区分度不小于 $\text{var0}=0.015$ 的条件下, 深度Q学习在 $\alpha=28$ 时达到收敛, 改变约束条件 var0 的值, 可得到不同的最优 α 和PRR值。这说明基于DQN的资源分配算法可获取最优门限参数值, 满足不同优先级业务的资源需求, 同时提高了分组接收率。

4 结束语

本文针对C-V2X环境的边路资源自主选择, 设计了基于不同车联网业务优先级的能量门限方程, 为了得到最优的门限方程参数, 提出基于深度Q学习的资源分配算法, 采用时间差分法, 利用后继状态的 Q 值估计当前 Q 值函数, 设计了基于分组接收率的回报方程。

仿真结果表明, 基于深度Q学习的边路资源分配算法可以保障不同优先级车联网业务的资源需求, 同时提高系统PRR性能, 并且在有限次的迭代之后取得收敛。下一步工作将通过设置不同形式的不同优先级联合门限方程, 同时考虑不同约束条件的影响, 采用机器学习仿真分析最优参数设计, 并将仿真结果与基站调度、基于地理位置的资源分配

方案进行对比,从而给出针对不同场景下的车联网边路资源优选策略。

参考文献:

- [1] 陈山枝, 时岩, 胡金玲. 蜂窝车联网(C-V2X)综述[J]. 中国科学基金, 2020, 34(2): 179-185.
CHEN S Z, SHI Y, HU J L. Cellular vehicle to everything (C-V2X): a review[J]. Bulletin of National Natural Science Foundation of China, 2020, 34(2): 179-185.
- [2] 葛雨明. 我国LTE-V2X标准化及测试验证进展[J]. 移动通信, 2019, 43(11): 36-39.
GE Y M. Progress on LTE-V2X standardization and testing in China[J]. Mobile Communications, 2019, 43(11): 36-39.
- [3] 杨波, 张莹, 蔡之骏, 等. 基于LTE-V的车载V2X系统研究[J]. 移动通信, 2019, 43(11): 75-80.
YANG B, ZHANG Y, CAI Z J, et al. Research on the vehicle V2X system based on LTE-V[J]. Mobile Communications, 2019, 43(11): 75-80.
- [4] 刘继军, 邹山花, 卢先领. MEC中资源分配与卸载决策联合优化策略[J]. 计算机科学与探索, 2021, 15(5): 848-858.
LIU J J, ZOU S H, LU X L. Joint optimization strategy of resource allocation and unloading decision in MEC[J]. Computer Science and Exploration, 2021, 15(5): 848-858.
- [5] 余翔, 陈晓东, 王政, 等. 基于LTE-V2X的联网资源分配算法[J]. 计算机工程, 2021, 47(2): 188-192.
YU X, CHEN X D, WANG Z, et al. Network resource allocation algorithm based on LTE-V2X[J]. Computer Engineering, 2021, 47(2): 188-192.
- [6] NGUYEN L H, NGUYEN V L, KUO J J. Efficient reinforcement learning-based transmission control for mitigating channel congestion in 5G V2X sidelink[J]. IEEE Access, 2022, 10: 62268-62281.
- [7] 陆岩. C-V2X通信分布式无线资源调度关键技术研究[D]. 上海: 东华大学, 2022.
LU Y. Research on key technologies of distributed radio resource scheduling in C-V2X communication[D]. Shanghai: Donghua University, 2022.
- [8] 杨天, 杨军. 移动边缘计算中的卸载决策与资源分配策略[J]. 计算机工程, 2021, 47(2): 19-25.
YANG T, YANG J. Unloading decision and resource allocation strategy in moving edge computing[J]. Computer Engineering, 2021, 47(2): 19-25.
- [9] HAMZAH H, LE D T, KIM M, et al. Location-aware task offloading for MEC-based high mobility service[C]//Proceedings of 2021 International Conference on Information Networking. Piscataway: IEEE Press, 2021: 708-711.
- [10] CHEN H M, WANG S F, WANG P, et al. Deep Q-learning for intelligent band coordination in 5G heterogeneous network supporting V2X communication[J]. Wireless Communications and Mobile Computing, 2022: 1-12.
- [11] 张兴龙, 陆阳, 李文璋, 等. 基于滚动时域强化学习的智能车辆侧向控制算法[J]. 自动化学报, 2022, 45: 1-12.
ZHANG X L, LU Y, LI W Z, et al. Lateral control algorithm for intelligent vehicles based on rolling time domain reinforcement learning[J]. Acta Automatica Sinica, 2022, 45: 1-12.
- [12] 裴晓飞, 莫烁杰, 陈祯福, 等. 基于TD3算法的人机混驾交通环境自动驾驶汽车换道研究[J]. 中国公路学报, 2021, 34(11): 246-254.
PEI X F, MO S J, CHEN Z F, et al. Lane changing of autonomous vehicle based on TD3 algorithm in human-machine hybrid driving environment[J]. China Journal of Highway and Transport, 2021, 34(11): 246-254.
- [13] 田康, 于楠, 李擎, 等. 基于改进TD3的自动驾驶车道保持决策方法[J]. 北京交通大学学报, 2022, 46(5): 84-94.
TIAN K, YU D, LI Q, et al. Lane keeping decision-making method of autonomous vehicles based on improved TD3[J]. Journal of Beijing Jiaotong University, 2022, 46(5): 84-94.
- [14] 罗薇, 汪梦珍, 许玲. 车联网高层协议关键技术[J]. 中兴通讯技术, 2020, 26(1): 35-39.
LUO W, WANG M Z, XU L. Main technologies adopted in high layer protocol for Internet of vehicles[J]. ZTE Technology Journal, 2020, 26(1): 35-39.
- [15] SANSANUM H. Technology and device for autonomous resource selection for internet of vehicles (V2X) transmission[P]. 2020.
- [16] 3GPP. Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); overall description: TS 36.300[S]. 2018: 289-302.
- [17] 3GPP. Evolved universal terrestrial radio access (E-UTRA); physical layer procedures: TS 36.213[S]. 2018: 422-447.
- [18] 刘伯阳, 马杰, 白静, 等. 基于MDP的协作认知边缘计算网络资源分配方案[J]. 西安邮电大学学报, 2022, 27(1): 23-28.
LIU B Y, MA J, BAI J, et al. MDP based resource allocation scheme for cooperative cognitive mobile edge computing networks[J]. Journal of Xi'an University of Posts and Telecommunications, 2022, 27(1): 23-28.
- [19] 3GPP. Study on LTE-based V2X services: TR 36.885[S]. 2016: 60-72.

[作者简介]



许辉(1972-), 男, 博士, 中兴通讯股份有限公司终端事业部资深标准研究员, 主要研究方向为5G/6G中的物联网、车联网和广播多播服务等标准研究。