DOI:10. 20079/j. issn. 1001-893x. 220821002

# 基于 GRU 和 LSTM 组合模型的车联网信道分配方法*

王 磊,王永华,何一汕,伍文韬

(广东工业大学 自动化学院,广州 510006)

**摘 要**:针对车联网中高通信需求和高移动性造成的车对车链路(Vehicle to Vehicle,V2V)间的信道冲突及网络效用低下的问题,提出了一种基于并联门控循环单元(Gated Recurrent Unit,GRU)和长短期记忆网络(Long Short-Term Memory,LSTM)的组合模型的车联网信道分配算法。算法以降低 V2V 链路信道碰撞率和空闲率为目标,将信道分配问题建模为分布式深度强化学习问题,使每条 V2V 链路作为单个智能体,并通过最大化每回合平均奖励的方式进行集中训练、分布式执行。在训练过程中借助 GRU 训练周期短和 LSTM 拟合精度高的组合优势去拟合深度双重 Q 学习中 Q 函数,使 V2V 链路能快速地学习优化信道分配策略,合理地复用车对基础设施(Vehicle to Infrastructure,V2I)链路的信道资源,实现网络效用最大化。仿真结果表明,与单纯使用 GRU 或者 LSTM 网络模型的分配算法相比,该算法在收敛速度方面加快了 5 个训练回合,V2V 链路间的信道碰撞率和空闲率降低了约 27%,平均成功率提升了约 10%。

**关键词**:车联网(IoV);信道分配;深度双重 Q 学习;GRU-LSTM 组合模型

**开放科学(资源服务)标识码(OSID)**:



**中图分类号**:TN929. 5 **文献标志码**:A **文章编号**:1001-893X(2024)02-0273-08

# A Channel Allocation Method for Internet of Vehicles Based on GRU and LSTM Hybrid Model

WANG Lei, WANG Yonghua, HE Yishan, WU Wentao

(School of Automation, Guangdong University of Technology, Guangzhou 510006, China)

**Abstract**: For the problems of channel conflict between Vehicle to Vehicle (V2V) links and low network utility caused by high communication requirements and high mobility in the Internet of Vehicles (IoV), a new channel allocation algorithm for the IoV based on hybrid model of parallel Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) is proposed. This algorithm aims to reduce the V2V links channel collision rate and idle rate, models the channel allocation problem as a distributed deep reinforcement learning problem, makes each V2V link as a single agent, and performs centralized training and distributed execution by maximizing the average reward per episode. In the training process, the hybrid advantages of the short training period of GRU and the high fitting accuracy of LSTM are used to fit the Q function in deep double Q-learning, so that the V2V links can quickly learn and optimize the channel allocation strategy to reuse the Vehicle to Infrastructure (V2I) links channel resources reasonably and maximize network utility. Simulation results show that compared with the allocation algorithm that simply uses the GRU or LSTM network model, the proposed algorithm accelerates the convergence rate by 5 training episodes, reduces the channel collision rate and idle rate between V2V links by about 27%, and increases the average success rate by about 10%.

**Key words**: Internet of Vehicles (IoV); channel allocation; GRU-LSTM hybrid model; double Q-learning

* **收稿日期**:2022-08-21;**修回日期**:2022-10-07
**基金项目**:国家自然科学基金资助项目(61971147);广东省研究生教育创新计划项目(2020JGXM040)
**通信作者**:王永华 Email:wangyonghua@ gdut. edu. cn

## 0 引 言

随着 5G 通信技术的发展,车联网(Internet of Vehicles,IoV)受到了越来越多的关注。车联网中存在着不同类型的连接,分为车对基础设施(Vehicle to Infrastructure,V2I)和车对车(Vehicle to Vehicle,V2V)链路。在 5G 蜂窝 V2X 网络中,需要同时满足高速率的海量数据传输以供娱乐,另一方面更需要可靠的信道资源以供必要的通信,因此,信道资源是实现车辆间的相互通信关键条件。为满足这种不同场景下的通信需求,文献[1]对 5G 网络中异构网络应用场景以及未来的研究趋势进行了讨论。然而信道资源的稀缺,显然已经不能满足当前车联网中的高通信需求。因此需要设计更加智能的信道分配方案,降低通信时信道冲突,最大化车联网的网络效用,提升信道资源利用率。

为应对这个挑战,文献[2]为基于设备到设备的车载网络开发了一种启发式空间频谱复用方案,减轻了对完整信道状态信息(Channel State Information,CSI)的要求;文献[3]指出的最大化 V2I 链路吞吐量的 V2X 资源分配方案能适应缓慢变化的大规模信道衰落,从而减少网络信令开销;文献[4]利用网络切片技术联合优化频谱资源块分配和车辆信号发射功率控制,最大化信息娱乐服务切片的平均和吞吐量。然而,这些算法大多假设车联网环境背景信息已知,但在实际情况下大多无法满足。深度强化学习由于在处理大状态和动作空间时能够提供目标值(称为 Q 值)的良好近似值而备受关注。文献[5]针对车联网可分配频谱资源数目未知的情况,提出了一种基于深度 Q 网络(Deep Q-Network,DQN)的联合缓存和计算资源方案。为进一步解决高移动性和多数目车辆环境中的频谱资源难以集中式管理问题,文献[6]提出了一种用于 V2V 和 V2I 通信的混合式频谱复用和功率分配方案,并设计基于卷积神经网络(Convolutional Neural Networks,CNN)的实时决策方法实现频谱复用和功率分配。

虽然使用深度强化学习算法能够实现车辆自主探索未知空间,智能地解决信道分配问题,但在实际车联网中由于传输需求不同,网络拓扑结构的变化十分迅速,从而使得传统的深度神经网络(Deep Neural Network,DNN)对这种在时间序列上变化快速的数据进行建模,运用到深度强化学习中时也很难让智能体学习到有效的信道分配策略。针对这个问题,目前的研究大多只是将长短期记忆(Long Short-Term Memory, LSTM)或者门控循环单元(Gated Recurrent Unit,GRU)去替代 DNN 在深度强化学习中的拟合 Q 函数的作用。虽然 LSTM 和 GRU 都能够处理前后连续的历史序列,但 LSTM 本身由于其结构内部参数较多,如果时间跨度很大,在网络比较深的情况下会使得计算量变大,很耗时,且有过拟合的风险[7]。同样,虽然 GRU 的简单结构,让其在训练时拥有比 LSTM 更低的计算复杂度,但在拟合精度上却比不上 LSTM。这种由于网络结构上的缺陷导致的算法性能上的不足,会使车联网中的信道分配问题难以寻找到最优解,导致算力上的浪费。

将 GRU 训练周期短与 LSTM 拟合精度和稳定性高的两个优点结合起来,能使算法更加高效和稳定[8-10]。本文以此为出发点,考虑将 GRU-LSTM 组合网络模型结合到分布式强化学习中,并围绕如何降低车联网中 V2V 链路的信道冲突并最大化网络效用的问题进行研究。

## 1 系统模型及问题陈述

### 1.1 系统模型

图 1 所示为由单个基站(Base Station,BS)以及 $M$ 条 V2I 链路和 $N$ 条 V2V 链路构成的十字路口车联网无线通信模型[11],$M$ 条 V2I 链路将车辆与 BS 进行连接,承载着娱乐以及交通管理数据(非安全数据)的传输,$N$ 条 V2V 链路主要承载安全数据的传输。为保证高质量 V2I 链路通信,假设每条 V2I 链路已被预先分配了不同的正交频谱子载波以消除网络中 V2I 链路之间的干扰,同时假设 V2V 链路对 V2I 链路的干扰也在理想状态内。V2I 链路作为授权用户,拥有独立的信道,V2V 链路可提供相邻车辆之间的直接通信。为了提高频谱利用率,V2V 链路作为感知用户需要利用与环境交互获得的部分可知信息,动态地感知 V2I 链路的信道条件,复用 V2I 链路的上行链路频谱进行信息交换,即在不影响 V2I 链路的正常通信的情况下以下垫式接入到其信道中来完成各自的传输任务。

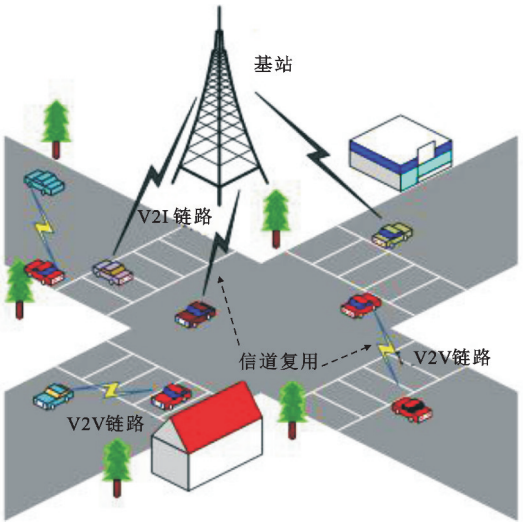


**图 1　车联网系统模型**[13]

Fig. 1 Internet of Vehicles system model[13]

因此如何设计一种快速稳定的算法完成这种信道资源稀少的场景下的信道分配问题,且能最大程度上降低信道冲突,提高 V2V 链路复用 V2I 链路信道资源的利用率是研究的重中之重。假设 V2I 链路被分配的正交信道数集合为 $C^*=\{1,2,3,\cdots,C\}$,而 V2V 链路的数量集合表示为 $N^*=\{1,2,3,\cdots,N\}$,当复用上行链路资源时,在每个时隙 V2V 链路都可以任意选择 V2I 链路的信道,且可以动态的选择继续留在该信道还是切换信道发送信息。因此,为实现 V2V 链路在共享 V2I 链路过程中最大化网络效用,尽可能降低信道冲突,就必须考虑各 V2V 链路之间的信道碰撞率,以及信道空闲率。

### 1.2　信道碰撞率

定义 $k$ 为时隙 $t$ 下第 $c$ 条 V2I 链路中选择复用此信道传输信息的 V2V 链路的数量,规定仅仅只能存在单条 V2V 链路选择复用第 $c$ 条 V2I 的信道时信息才能够发送成功,当有两条及两条以上的 V2V 链路共同选择复用同一条 V2I 链路时,就定义为产生了信道的碰撞,信息必定传输失败,此时的碰撞次数就为 1,如式(1)所示:

$$\beta_c=\begin{cases}1, k>1\\0, k\leqslant 1\end{cases} \tag{1}$$

因此,将 $i$ 次信息传输过程中 $C$ 条 V2I 链路信道中产生的碰撞总次数与这 $i$ 次传输中的总信道数的比值,定义为这 $i$ 次传输中的信道碰撞概率 $\mu$,如式(2)所示:

$$\mu=\frac{\sum_{c=1}^{C}\beta_c\times i}{C\times i} \tag{2}$$

### 1.3　信道空闲率

定义 $\phi$ 为信道空闲率来间接表示 V2I 链路信道的利用情况。当 $n$ 条 V2V 链路都进行了信道的共享策略后,第 $c$ 个信道中的剩余容量 $\gamma_c$ 如式(3)所示。规定当第 $c$ 条 V2I 信道被占用且 V2V 链路成功发送了信息,那么该信道的剩余容量 $\gamma_c$ 就为 0;如果该条信道上,发生了多条 V2V 链路的竞争,造成了通信失败,此信道就没有被利用,其剩余容量 $\gamma_c$ 为 1;当然,如果某条信道没有被 V2V 用户选择共享,其信道剩余容量 $\gamma_c$ 自然也为 1。

$$\gamma_c=\begin{cases}1, k>1 \text{ or } k=0\\0, k=1 \text{ 且成功发送了信息}\end{cases} \tag{3}$$

规定将 $i$ 次信息传输过程中 $C$ 条 V2I 链路信道的剩余容量 $\gamma_c$ 之和与这 $i$ 次传输过程中的总信道数的比值,表示该回合信道空闲率,如式(4)所示:

$$\phi=\frac{\sum_{c=1}^{C}\gamma_c\times i}{C\times i} \tag{4}$$

可见,信道空闲率与碰撞率呈正相关关系,信道空闲率的降低,也间接表明了碰撞率的降低和信道利用率的提升。因此,本文提出的算法将围绕这两个优化指标来进行设计和实现。

## 2　本文提出的算法

### 2.1　深度强化学习算法框架

本文的车联网信道分配场景中,由于真实环境信息是未知的、高维复杂的,因此,将信道资源分配问题建模为深度强化学习问题,提出一种基于 GRU-LSTM 组合网络模型的深度双重 Q 学习算法框架(Hybrid GRU-LSTM DDQN,HG-LDDQN),算法结构如图 2 所示。

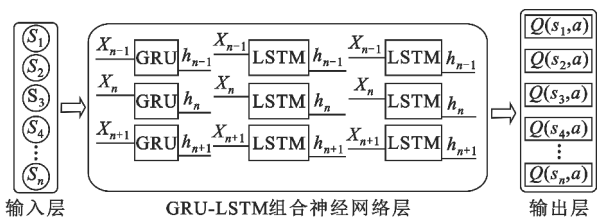


**图 2　HG-LDDQN 算法结构框图**

Fig. 2 HG-LDDQN algorithm structure block

HG-LDDQN 算法与环境交互模型如图 3 所示。算法模型采用集中训练、分布式执行的方式,将每条 V2V 链路作为智能体与环境进行交互,接收环境观察结果 $O(t)$,以得到环境中在 $t$ 时隙下的状态信息

$S(t)$;将 $t$ 时隙下的状态 $S(t)$ 送入 GRU-LSTM 组合神经网络模型中进行训练,得到 Q 函数的值 $Q(s,a)$。然后,依据 Q 值智能体得到下一步所要进行的动作 $A(t)$,并且在同一种奖励评判机制下,每条 V2V 链路单独获得回报 $R_n(t)$,继而反复探索训练,更新 GRU-LSTM 组合网络。最后,通过迭代学习最大化每回合的平均奖励,来改善信道分配策略。

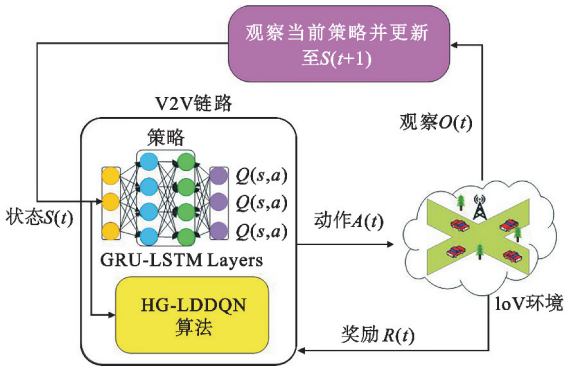


**图 3 HG-LDDQN 算法与环境交互模型**

Fig. 3 HG-LDDQN algorithm and environment interaction model

下面对 HG-LDDQN 算法与环境交互模型中的几个深度强化学习要素分别进行阐述。

1) 状态空间

在算法模型中,$t$ 时隙下的状态空间 $S(t)$ 是通过 V2V 链路对环境进行观察 $O(t)$ 后得到的,其包含三部分,即 V2V 链路作为智能体的动作 $a(t)$、当前每个信道的剩余容量 $\delta(t)$ 以及确认字符信号(Acknowledge character,ACK)的返回结果 $\eta(t)$。

如果 V2V 链路用户已经在 $t$ 时隙选择了第 $c$ 条信道($1\leqslant c\leqslant C$)进行数据传输,那么将该条信道状态 $a_c(t)$ 设置成 1,剩余的信道状态设置成 0。$a(t)$ 如式(5)所示:

$$a(t)=\{a_1(t),a_2(t),\cdots,a_c(t)\} \tag{5}$$

此外,在时隙 $t$ 对于当前 $C$ 个信道中的第 $c$ 个信道按式(3)中定义的单条 V2I 信道的剩余容量 $\gamma_c$ 的计算方法,计算此刻所有 V2I 链路信道的剩余容量 $\delta(t)$,如式(6)所示:

$$\delta(t)=\{\gamma_1,\gamma_2,\cdots,\gamma_c\} \tag{6}$$

假设在时隙 $t$ 完成信道共享后,V2V 链路间发送数据包的同时也会给对方发送一条 ACK 信号,如果数据传输成功就返回一个数值为 1 的 ACK 信号,传输失败,则返回的 ACK 信号为 0。ACK 信号返回结果 $\eta(t)$ 如式(7)所示:

$$\eta(t)=\begin{cases}1,\text{数据传输成功}\\0,\text{数据传输失败}\end{cases} \tag{7}$$

由此,构成了在时隙 $t$ 下的状态空间 $S(t)$,如式(8)所示:

$$S(t)=\{a(t),\delta(t),\eta(t)\} \tag{8}$$

2)动作空间

根据可选信道 $c$,$n$ 条 V2V 链路在 $t$ 时隙的可选动作空间 $A(t)$ 由式(9)定义为

$$A(t)\in\{0,1,2,3,\cdots,c\} \tag{9}$$

即每条 V2V 链路都可以选择此时刻网络空间中的任一 V2I 链路的信道。当 $t$ 时刻下第 $n$ 条 V2V 链路的动作值 $a_n(t)=0$ 时,代表该条 V2V 链路在 $t$ 时刻下选择不接入 V2I 的信道。

3)奖励值设定

在 $t$ 时隙下,第 $n$ 条 V2V 链路成功发送信息后,根据 V2V 的接收方返回的 ACK 信号状态,对该次动作 $a_n(t)$ 的选择给予一个奖励值 $R_n(t)$。如果返回 ACK 信号为 1,说明数据信息发送成功,即表明 V2V 链路合理地复用了 V2I 的信道,同时避免了信道的冲突,给予该次动作 $a_n(t)$ 数值为 1 的正向奖励;反之,不给予奖励。因此,将 $t$ 时隙下第 $n$ 条 V2V 链路的动作 $a_n(t)$ 的奖励值 $R_n(t)$ 定义为

$$R_n(t)=\begin{cases}1,\eta(t)=1\\0,\text{otherwise}\end{cases} \tag{10}$$

## 2.2 基于 GRU-LSTM 组合网络模型的深度双重 Q 学习算法

根据前述的强化学习的基本要素,对本文提出的算法结构进行分块阐述。

### 2.2.1 输入层

在本算法中,每条 V2V 链路都被看作是一个智能体,智能体观察并采集 $t$ 时刻下的每个 V2V 链路的状态值 $S_t\in\{S_1,S_2,S_3,\cdots,S_m\}$ 作为 GRU-LSTM 组合网络的输入。当 V2V 链路在状态 $S_t$ 执行动作 $a(t)$,根据环境返回的 $\eta(t)$ 获得一个奖励 $R(t)$ 后,就转移至下一个状态 $S_{t+1}$。

### 2.2.2 GRU-LSTM 组合神经网络层

由于车联网的高移动性和网络拓扑的快速变化,经典的 DNN 无法学习到前后联系的历史序列,同时循环神经网络(Recurrent Neural Network,RNN)存在梯度消失和梯度爆炸以及可能过拟合的缺陷,因此,本算法使用 GRU-LSTM 组合神经网络模型。该组合神经网络模型的网络结构有 3 层。第一层采用 GRU,它将 LSTM 中的遗忘门和输入门合并为一个“更新门”,减小了矩阵乘法,更容易使算法收敛,可以减少训练时间[12]。但 GRU 的拟合精度不如多参数的 LSTM,并且双层 LSTM 的精度要优于单层 LSTM[13]。因此,模型的第二层和第三层结构均采用 LSTM。下面对该组合层进行分层介绍。

第一层神经网络由多个 GRU 单元组成。对于每个 GRU 单元,如图 4 所示,$\boldsymbol{Z}_t$ 为当前时刻的输入,$\boldsymbol{Y}_{t-1}$ 为上一个时刻的输出,$\boldsymbol{Y}_t$ 为当前时刻的输出。

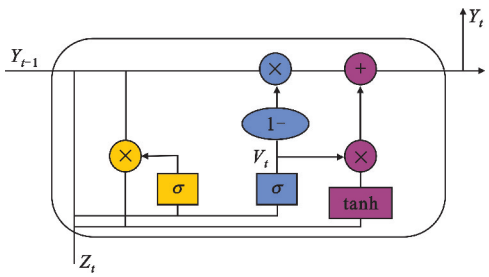


**图 4　GRU 单元结构图**[10]

Fig. 4 GRU network structure[10]

GRU 有两个门,第一个门为更新门 $v_t$,决定了有多少历史信息可以继续传递给未来。更新门 $v_t$ 的计算方法如公式(11)所示[8]:

$$v_t=\sigma(\boldsymbol{W}_v\cdot[\boldsymbol{Y}_{t-1},\boldsymbol{Z}_t]+\boldsymbol{b}_v) \tag{11}$$

式中:$\boldsymbol{W}_v$ 为更新门的权重矩阵;$\boldsymbol{b}_v$ 为偏差向量;$\sigma$ 表示激活函数 sigmoid。

第二个门为重置门 $r_t$,主要功能是确定有多少历史信息不能传递到下一个状态。重置门 $r_t$ 的计算方法如公式(12)所示[8]:

$$r_t=\sigma(\boldsymbol{W}_r\cdot[\boldsymbol{Y}_{t-1},\boldsymbol{Z}_t]+\boldsymbol{b}_r) \tag{12}$$

式中:$\boldsymbol{W}_r$ 为重置门的权重矩阵;$\boldsymbol{b}_r$ 为偏差向量。

计算出更新门 $v_t$ 和重置门 $r_t$ 后,GRU 将会计算候选隐藏状态 $h_t$。候选隐藏状态$\boldsymbol{h}_t$ 的计算方法如公式(13)所示[8]:

$$\boldsymbol{h}_t=\tanh(\boldsymbol{W}_h\cdot[r_t\cdot\boldsymbol{Y}_{t-1},\boldsymbol{Z}_t]+\boldsymbol{b}_h) \tag{13}$$

式中:$\boldsymbol{W}_h$ 为对应的权重参数;$\boldsymbol{b}_h$ 为对应的偏差参数;tanh 代表双曲正切函数。

最后 $t$ 时刻 GRU 的输出 $\boldsymbol{Y}_t$ 的计算方法如公式(14)所示[8]:

$$\boldsymbol{Y}_t=(1-v_t)\cdot\boldsymbol{Y}_{t-1}+v_t\cdot\boldsymbol{h}_t \tag{14}$$

在 GRU 网络层输出后第二层和第三层是 LSTM 网络层,对比于 RNN 和 GRU,LSTM 模型的拟合精度总体更高,如图 5 所示。

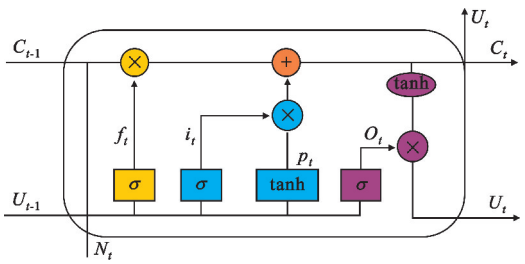


**图 5　LSTM 单元结构**[10]

Fig. 5 LSTM network structure[10]

LSTM 有 3 个门,如图 5 所示,$\boldsymbol{C}_{t-1}$ 为前一时刻神经元的状态,$\boldsymbol{U}_{t-1}$ 为前一时刻神经元的输出,$\boldsymbol{N}_t$ 为当前时刻的输入,$\boldsymbol{C}_t$ 为当前时刻神经元的状态,$\boldsymbol{U}_t$ 为当前时刻神经元的输出。以下是每个 LSTM 单元的前向传播公式:

$$f_t=\sigma(\boldsymbol{W}_f\cdot[\boldsymbol{U}_{t-1},\boldsymbol{N}_t]+\boldsymbol{b}_f) \tag{15}$$

式中:$W_f$ 是遗忘门的权重矩阵;$\boldsymbol{b}_f$ 是偏差向量;$f_t$ 表示最后一层神经元被遗忘的概率[8]。

$$i_t=\sigma(\boldsymbol{W}_i\cdot[\boldsymbol{U}_{t-1},\boldsymbol{N}_t]+\boldsymbol{b}_i) \tag{16}$$

式中:$\boldsymbol{W}_i$ 是输入门的权重矩阵;$\boldsymbol{b}_i$ 是偏差向量;$i_t$ 表示当前需要保留的负载信息的比例[8]。

$$p_t=\tanh(\boldsymbol{W}_c\cdot[\boldsymbol{U}_{t-1},\boldsymbol{N}_t]+\boldsymbol{b}_c) \tag{17}$$

式中:$\boldsymbol{W}_c$ 是输入门的权重矩阵;$\boldsymbol{b}_c$ 是偏差向量;$p_t$ 是当前需要保留的负载信息的比例[8]。

$$\boldsymbol{C}_t=f_t\cdot\boldsymbol{C}_{t-1}+i_t\cdot p_t \tag{18}$$

$$o_t=\sigma(\boldsymbol{W}_o\cdot[\boldsymbol{U}_{t-1},\boldsymbol{N}_t]+\boldsymbol{b}_o) \tag{19}$$

式(19)中:$\boldsymbol{W}_o$ 为输出门的权重矩阵;$\boldsymbol{b}_o$ 为偏差向量;$o_t$ 为输出门[8]。

$$\boldsymbol{U}_t=o_t\cdot\tanh(\boldsymbol{C}_t) \tag{20}$$

此处,LSTM 层的输入就是 GRU 网络层的输出 $\boldsymbol{Y}_t$。显然,此组合网络的数据更新过程比单纯的 LSTM 更简洁,也比单纯的 GRU 网络拟合 Q 值过程更具有精确性和稳定性。

在组合神经网络中,使用 Huber 损失函数来计算算法训练时的目标值 $Y$ 以及估计值 $f(x)$ 之间的差值。Huber 损失是平方损失和绝对损失的综合,它克服了平方损失和绝对损失的缺点,不仅使损失函数具有连续的导数,而且利用均方误差(Mean Square Error,MSE)梯度随误差减小的特性,可取得更精确的最小值,也对异常点更加鲁棒,可以提高算法的稳定性[14]。Huber 损失计算方法如式(21)所示[14]:

$$L=\begin{cases}\dfrac{1}{2}(Y-f(x))^2, & |Y-f(x)|\leqslant\delta\\[2ex] \dfrac{1}{2}(Y-f(x))^2, & |Y-f(x)|>\delta\end{cases} \tag{21}$$

式中:$\delta$ 为选择超参数,作为选择 MSE 与 MAE 时的评判值,由反复实验确定。

### 2.2.3　输出层

为解决算法训练中的过度估计问题,使用 DDQN 来解耦目标 Q 值动作的选择和目标 Q 值的计算[15]。具体而言,使用两个深度组合模型 Q 网络,$Q_1$ 网络用于选择动作 $a_n(t)$,$Q_2$ 网络用于估计与所选动作相关联的 Q 值。DDQN 中的 Q 值的近

似估算公式如式(22)所示[15]：

$$Q(a_n(t)) \approx R_n(t+1) + \widetilde{Q}_2(\underset{a}{\operatorname{argmax}}(\widetilde{Q}_1(a))) \tag{22}$$

将提出的 HG-LDDQN 算法为所有 V2V 链路进行训练，训练步骤如下：

1 初始化：迭代轮数 $T$，V2I 链路条数 C，V2V 链路条数 $N$，步长 $\alpha$，衰减因子 $\gamma$，探索率 $\varepsilon$，经验回放池 $D$，当前 GRU-LSTM net1 的参数 $\omega$，目标 GRU-LSTM net2 的参数 $\omega'=\omega$，所有状态和动作对应的价值 Q

2 For iteration $i=1,\cdots,I$ do

3 For episode $m=1,\cdots,M$ do

4 For time-slot $t=1,\cdots,T$ do

5 For V2V links $n=1,\cdots,N$ do

6 从环境中观察得到状态值 $X_n(t)$，输入到 GRU-LSTM net1，产生对应所有可选的动作 $a\in\{0,1,2,\cdots C\}$ 的估计 Q 值 $Q(a)$

7 用 $\epsilon$-贪婪法在当前 Q 值输出中选择动作 $a_n\in\{0,1,2,\cdots C\}$，并且返回一个奖励值 $R_n(t+1)$

8 观察环境得到下一个状态值 $X_n(t+1)$，并且将其输入 GRU-LSTM net1 与 GRU-LSTM net2 中，产生对于所有动作 $a\in\{0,1,2,\cdots,C\}$ 的估计 Q 值 $\widetilde{Q}_1(a)$ 及 $\widetilde{Q}_2(a)$

9 在经验回放池中存储<$s,a,r,s'$>

10 从经验回放池中随机抽取批量样本训练组合神经网络

11 计算当前的目标 Q 值：

$Q(a_n(t))\leftarrow R_n(t+1)+\widetilde{Q}_2(\underset{a}{\operatorname{argmax}}(\widetilde{Q}_1(a)))$

12 计算目标 Q 值与估计 Q 值的 Huber loss 与网络权重 $\omega$

13 End for

14 End for

15 End for

16 使用状态输入 $X_n(t)$ 和输出 Qs 训练 GRU-LSTM net1

17 每一个 iteration 使 $Q_2\leftarrow Q_1$

18 End for

## 3 实验与结果分析

仿真场景为位于十字路口道路的双向和单向车道区域，其宽为 300 m，长为400 m。场景中车辆起始位置和行驶方向在区域范围内随机初始化，在该范围内规定有 2 条 V2I 链路、3 条 V2V 链路以及 1 个基站。在该场景模型中，使用 HG-LDDQN 算法实现 3 条 V2V 链路共享 V2I 链路的 2 个信道条件的尝试，分别在信道碰撞率、信道空闲率以及平均奖励和平均成功率 4 个评价指标上与其他信道分配算法对比，以验证 HG-LDDQN 算法的性能。

实验中构建图 2 中的 GRU-LSTM 组合神经网络，GRU 层和两层 LSTM 均设置 128 个神经元。Huber 损失函数的超参数 $\delta$ 经过大量实验设置为 1.35。实验每次输入 $t-5$ 个时刻的状态序列，使用 Adam 算法优化网络权重 $\omega$，经验池 $D$ 的容量设置为 1 000，探索率 $\varepsilon$ 设置为 0.02，探索率的衰减率设置为 0.000 1，学习率设置为 0.01，奖励折扣设置为 0.9，干扰设置成 0.1，模拟退火常数设置为 1。

### 3.1 信道碰撞率对比

图 6 表示在 55 000 次的迭代中，3 条 V2V 链路在动态共享 2 条 V2I 链路的信道时的碰撞率的变化情况，每 5 000 次作为一个回合，对数据结果进行一次记录。从图中可见，没有历史序列前后记忆功能的 DQN 算法在处理这种历史序列的学习任务时几乎没有学习能力，碰撞率很大，而对于单一循环网络算法而言，GRU+DDQN 算法由于具有比 LSTM+DDQN 更为简单的结构，其学习迭代的更快。但这两种算法最后的收敛表现差不多，在第 10 个训练回合时收敛到 0.27 左右。相较而言，HG-LDDQN 算法由于使用了 GRU-LSTM 混合网络模型，兼具 GRU 和 LSTM 网络单元的双重性能，能将 GRU 网络单元结构简单、训练快速的优势运用到 V2V 链路的训练中，当训练达到第 4 个回合时碰撞率就以最大的下降速度降低，使 V2V 链路之间的碰撞次数迅速减少，同时又因为 LSTM 网络单元中的多参数能带来更加精确的拟合精度，使得 HG-LDDQN 算法不仅提前 5 个训练回合完成收敛，又能够将碰撞率维持在比其他算法训练结果更低的 0.006 附近。

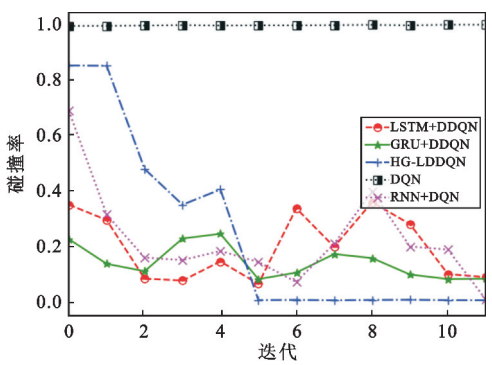


**图 6 3 条 V2V 链路共享 2 条 V2I 链路信道时的碰撞率**

Fig. 6 Collision probability when 3 V2V links share the channel of 2 V2I links

### 3.2 平均奖励对比

图 7 为 3 条 V2V 链路共享 2 条 V2I 链路信道

时的平均奖励的对比,可见 HG-LDDQN 算法凭借 GRU-LSTM 组合网络中 GRU 网络单元的简单结构,使 V2V 链路能够在第 4 个回合以后快速学习获得奖励,又可以凭借组合网络中 LSTM 网络单元的多参数拟合精确的特点,使 V2V 链路在第 5 个回合后几乎每次都能成功共享 V2I 链路的 2 条信道,完成信息成功发送,学习到了比其他算法更优的信道分配策略。本文算法比 RNN+DQN 算法提前约 6 个训练回合收敛,而 GRU+DDQN 和 LSTM+ DDQN 算法由于单一的网络结构无法在整体性能上表现出组合优势,导致在整体的算法性能上不如 HG-LDDQN 算法高效和稳定,最终的平均奖励值只能收敛到 1.8 附近,甚至不如传统的 RNN+DQN 算法。DQN 算法还是因为使用 DNN 的原因,处于一种无法学习的状态,几乎不能获得奖励。

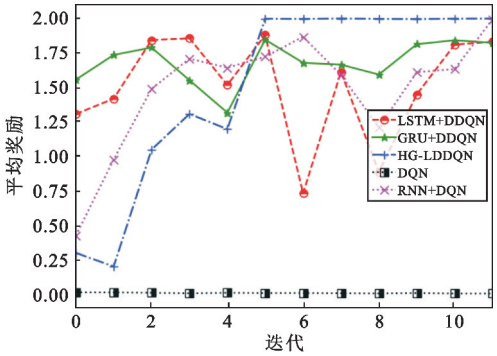


**图 7　3 条 V2V 链路共享 2 条 V2I 链路信道时的平均奖励**
Fig. 7 Average reward when 3 V2V links share the channel of 2 V2I links

### 3.3　信道空闲率对比

图 8 为 3 条 V2V 链路共享 2 条 V2I 链路的信道时的空闲率的对比。由于建模时允许某些 V2V 链路可以选择不发送信息,即不选择信道接入,因此该图与碰撞率的图有些许的差别。显而易见 HG-LDDQN 算法由于组合网络模型结构带来的双重优势,在收敛速度上比 LSTM+DDQN 或者 GRU+DDQN 算法快 5 个训练回合,比 RNN+DQN 快 6 个训练回合。在收敛后的空闲率上,随着迭代次数的增加,HG-LDDQN 算法能使信道空闲率稳定在较低的水准,使 V2I 的 2 条信道基本都有 V2V 链路成功的共享,相较于单一网络结构的 LSTM + DDQN 或者 GRU+DDQN 算法下降了约 27%。DQN 算法同样由于网络结构的原因,不具备学习历史序列数据的能力。RNN+DQN 算法下,信道的空闲率呈现出上下振荡的不稳定性,以及收敛速度慢的情况。

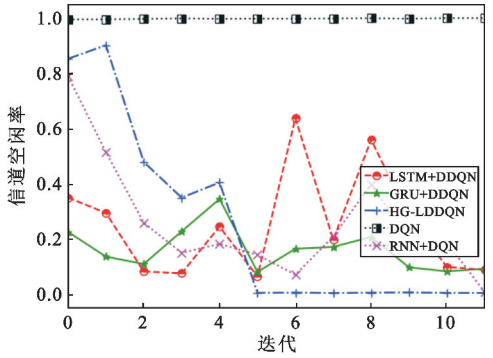


**图 8　3 条 V2V 链路共享 2 条 V2I 链路信道时的信道空闲率**
Fig. 8 Channel idle probability when 3 V2V links share the channel of 2 V2I links

### 3.4　平均成功率的对比

图 9 表示 3 条 V2V 链路尝试共享 2 条 V2I 链路的信道的过程中的平均成功率情况。由于奖励函数的设计是每次对于 V2V 链路成功共享到 V2I 链路信道,并完成信息传输的动作选择就设置奖励值就加 1,发生碰撞信道共享失败,奖励值就为 0。因此,每一个回合内的累计的成功共享次数与该回合内的累计奖励值是一致的,可以看到平均化后的成功率折线图是和奖励图的趋势是一致的。从图中仍然可以发现,HG-LDDQN 算法具有明显优势,能够快速完成收敛,使平均成功率达到了接近 1 的效果,比 GRU+DDQN 和 LSTM+DDQN 算法下的平均成功率提高了约 10%,能够保证在之后的每个时隙中 V2I 的 2 个信道中都有 V2V 链路成功进行了共享且完成了信息传输。

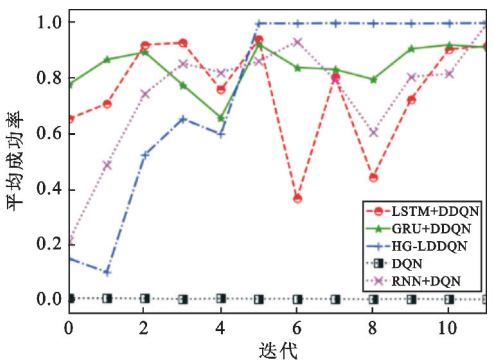


**图 9　3 条 V2V 链路共享 2 条 V2I 链路信道时的平均成功率**
Fig. 9 Average success rate when 3 V2V links share the channel of 2 V2I links

## 4　结束语

本文研究了针对车联网中 V2V 链路复用 V2I 链路信道时的信道冲突以及网络效用低下的问题,

提出了一种基于 GRU 和 LSTM 组合模型的动态信道分配算法。该算法以最大化每回合平均奖励为目标训练 V2V 链路,不需要在线协调,可实现多个 V2V 链路通过实时探知环境状态,选择 V2I 链路未使用的空闲频谱以完成 V2V 链路自身信息的传输任务,同时解决了大状态空间下 V2V 链路用户随着车联网节点拓扑结构变化带来的训练困难、训练周期长的问题。仿真实验结果表明,该算法能使 V2V 链路作为智能体在与环境不断交互过程中学习到合理的信道共享策略,有效地解决了快速变化的车联网环境中的信道分配问题,同时减少了 V2V 链路用户的信道碰撞率以及空闲率,间接最大化了 V2V 链路复用 V2I 链路信道资源的利用率。

后续将会在本文的基础上对 V2I 以及 V2V 链路的频谱资源分配进行信道及功率的联合优化研究。

## 参考文献:


[1] XU Y,GUI G,GACANIN H,et al. A survey on resource allocation for 5G heterogeneous networks: current research, future trends and challenges [J]. IEEE Communications Surveys and Tutorials, 2021, 23(2): 668-695.

[2] BOTSOV M, M K, KELLERER W, et al. Location dependent resource allocation for mobile device-to-device communications [C]//Proceedings of 2014 IEEE Wireless Communications and Networking Conference. Istanbul:IEEE,2014:1679-1684.

[3] SUN W, STROM E, BRANNSTROM F, et al. Radio resource management for D2D-based V2V communication [J]. IEEE Transactions on Vehicular Technology,2015,65(8):6636-6650.

[4] 李悦,任春莉,章国安. 车联网中网络切片资源分配方案[J]. 电讯技术,2023,63(1):85-92.

[5] HE Y,ZHAO N,YIN H. Integrated networking,caching, and computing for connected vehicles: a deep reinforcement learning approach[J]. IEEE Transactions on Vehicular Technology,2017,67(1):44-55.

[6] CHEN M, CHEN J, CHEN X, et al. A deep learning based resource allocation scheme in vehicular communication systems[C]//Proceedings of 2019 IEEE Wireless Communications and Networking Conference. Marrakesh:IEEE,2019:1-6.

[7] YANG S,YU X,ZHOU Y. LSTM and GRU neural network performance comparison study:taking yelp review dataset as an example [C]//Proceedings of 2020 International Workshop on Electronic Communication and Artificial Intelligence. Shanghai:IEEE,2020:98-101.

[8] 贺小伟,徐靖杰,王宾,等. 基于 GRU-LSTM 组合模型的云计算资源负载预测研究[J]. 计算机工程,2022,48(5):11-17.

[9] NI R, CAO H. Sentiment analysis based on GloVe and LSTM-GRU [C]//Proceedings of 2020 39th Chinese Control Conference. Shenyang:IEEE,2020:7492-7497.

[10] UWAISU A, YAHAYA A S, KAMAL S M, et al. A hybrid deep stacked LSTM and GRU for water price prediction [C]//Proceedings of the 2nd International Conference on Computer and Information Sciences. Sakaka:IEEE,2020:1-6.

[11] LIANG L,YE H,LI G Y. Spectrum sharing in vehicular networks based on multi-agent reinforcement learning [J]. IEEE Journal on Selected Areas in Communications,2019,37(10):2282-2292.

[12] ELSAYED N, MAIDA A S, BAYOUMI M. Gated recurrent neural networks empirical utilization for time series classification [C]//Proceedings of 2019 International Conference on Internet of Things and IEEE Green Computing and Communications and IEEE Cyber, Physical and Social Computing and IEEE Smart Data. Atlanta:IEEE,2019:1207-1210.

[13] YU Y,SI X,HU C ,et al. Areview of recurrent neural networks: LSTM cells and network architectures [J]. Neural Computation,2019,31(7):1235-1270.

[14] YOU M,LU A. A robust TDOA based solution for source location using mixed Huber loss[J]. Journal of Systems Engineering and Electronics,2021,32(6):1375-1380.

[15] HASSELT H V, GUEZ A, SILVER D. Deep reinforcement learning with double Q-learning [C]//Proceedings of 2016 National Conference on Artificial Intelligence. Phoenix:AAAI,2016:2094-2100.


## 作者简介:


**王 磊** 男,1993 年生于陕西汉中,2017 年获学士学位,现为硕士研究生,主要研究方向为认知无线网络和深度强化学习。

**王永华** 男,1979 年生于河北石家庄,2009 年获博士学位,现为副教授,主要研究方向为认知无线网络、机器学习。

**何一汕** 男,1998 年生于湖南郴州,2020 年获学士学位,现为硕士研究生,主要研究方向为认知无线网络。

**伍文韬** 男,1997 年生于广东佛山,2020 年获学士学位,现为硕士研究生,主要研究方向为认知无线电。