

面向自动驾驶应用的车联多智能体信息融合协同决策机制研究

曹佳钰, 冷甦鹏, 张科

(电子科技大学信息与通信工程学院, 四川 成都 611731)

摘要: 自动驾驶车辆是智能交通系统的重要组成部分, 自动驾驶应用需要在采集大量交通数据的基础上, 实施信息处理与控制决策。由于交通信息的时空属性, 即信息只在某一时间段内或某一区域内有效, 独立车辆有限的信息感知范围严重制约了自动驾驶应用数据采集的有效性。车联多智能体协同决策为上述问题提供了可行的解决思路, 提出了一种多维度感知信息融合机制, 提升车载信息融合对自动驾驶任务的增益。在此基础上, 设计了面向自动驾驶应用的智能分布式的决策算法, 在最大化信息融合对自动驾驶任务增益的同时, 最大化路网交通车流量, 并满足自动驾驶车辆的成本和资源约束。仿真结果验证了所提算法的收敛性和实用性。

关键词: 自动驾驶; 多智能体; 信息融合; 移动边缘服务

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.2096-3750.2020.00179

Multi-agent driven collaborative decision mechanism of information fusion for autonomous driving vehicles

CAO Jiayu, LENG Supeng, ZHANG Ke

School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

Abstract: Autonomous vehicles play an important part in intelligent transportation systems. In these vehicles, driving control decision is obtained based on the collection of massive traffic states and intensive information processing. However, the spatial-temporal characteristics of the traffic states and the constrained environmental perception range of an individual vehicle seriously undermine the effectiveness of the state collection. Multi-agent driven collaborative decision provides a potential approach to address the problem. A multi-dimensional information fusion mechanism was proposed, which improved the gain of vehicular information fusion for autonomous driving tasks. Moreover, an intelligent distributed decision algorithm was designed for autonomous driving applications, which maximized the road traffic flow while maximizing the gain of information fusion on the autonomous driving mission under vehicular cost and resource constraints. Numerical results demonstrate the convergence and practicability of the proposed algorithm.

Key words: autonomous driving, multi-agent, information fusion, mobile edge service

1 引言

近年来, 科技的进步极大地推动了智能交通的发展, 智能交通系统在解决交通拥堵、环境污染、能源消耗等问题上受到了越来越多的关注。在此背景下, 网联自动驾驶汽车 (CAV, connected and au-

tonomous vehicle) 技术应运而生^[1-3]。CAV 技术可以分解为车辆互联技术和自动驾驶技术, 车辆互联技术主要包括两种类型的通信链路, 即车与车 (V2V, vehicle to vehicle) 通信和车与路边基站 (V2I, vehicle to infrastructure) 通信。V2V 通信允许车载用户进行短距离和中距离通信, 提供低成本部署, 并以低时

收稿日期: 2020-06-17; 修回日期: 2020-07-28

通信作者: 冷甦鹏, spleng@uestc.edu.cn

基金项目: 四川省科技计划重点研发项目 (No.2019YFG0520)

Foundation Item: The Key R&D Project of Sichuan Science and Technology Plan (No.2019YFG0520)

延支持短消息传递。V2I 通信使车辆能够连接互联网,通过路边基站进行信息传播和提供信息娱乐服务。自动驾驶技术可以利用车载传感器检测车辆的行驶信息,如车辆的速度、位置、加速度等,并通过对这些信息的处理,提高整个网络的安全性和通行效率。由此可见,CAV 技术是智能交通系统的核心基础,是城市现代化的重要组成部分。

随着智能汽车和交通传感器的普及以及传感器性能的提升,道路上行驶的车辆将产生大量行驶信息,这些信息(如路径规划的位置信息、地图导航信息等)通常具有时空属性,即只在某一段时间内或某一区域内有效。在 CAV 技术中,独立车辆的自动驾驶控制决策与其他车辆的行驶信息紧密相关。因此,自动驾驶应用需要准确地了解实时驾驶环境,即在数据失效之前完成数据采集任务,才能进一步处理得到自动驾驶的控制策略。然而,在实际的交通环境中,因噪声、干扰和物理遮挡的存在,独立车辆的信息感知范围有限,这严重制约了自动驾驶应用数据采集的有效性。

为了解决上述问题,移动边缘服务将车辆的行驶信息缓存在道路附近的边缘节点(车辆或备有边缘服务器的路边基站)成为了一个可行的解决方案^[4]。当车辆采集自身行驶信息后,可以直接将行驶信息缓存在附近的边缘节点,并通过边缘节点将信息传输至其他车辆,弥补了车辆感知能力有限的不足,大大降低了车辆数据采集的时间消耗。然而,边缘服务器和车辆的缓存资源有限,而且不同的信息融合模式在时延和成本方面具有不同的表现。因此,在具有多种传输模式和动态拓扑结构的车辆网络中,高效的自动驾驶信息采集仍是一个需要智能解决方案的挑战。

在自动驾驶应用中,车辆的控制策略可以是对驾驶动作(如速度、加速度、制动减速度等)的控制^[5],也可以是对更高层的驾驶策略(如超车、路径规划等)的控制^[6]。在解决自动驾驶问题时,常使用强化学习训练模型,给出单个车辆在动态变化场景下最优的驾驶控制策略。然而,现实中的自动驾驶系统不是简单的单车系统或集中式决策系统,而是一个分布式的多智能体系统。因此,如何在完成高效数据采集的同时给出分布式的自动驾驶控制决策仍然是一个待探索的问题。

CAV 技术为未来的交通系统带来了巨大的潜

力,已经引起了研究者的广泛关注。文献[7]提出了一种用于 CAV 通过路口的全分布式交互协议,利用拉萨尔不变性原理设计了一种有效的调整工具,确保车辆在到达十字路口之前达到所需的队形。文献[8]针对联网自动驾驶中的运动预测问题提出了一种考虑不同驾驶风格的前车轨迹预测联合时序的建模方法,利用长短时记忆递归神经网络技术,实现了对前车精确和个性化的轨迹预测。文献[9]研究了多车协同自主停车问题,将自主停车轨迹规划问题转化为最优控制问题,通过高斯伪谱法实现了停车时间最小化。

随着智能交通系统的兴起和自动驾驶应用需求的不断涌现,近年来的研究工作主要集中在基于移动边缘技术的自动驾驶服务。文献[10]研究了车速预测问题,提出了一种基于移动边缘技术的融合交通灯模型和车速模型的短期交通预测模型,有效地捕捉交通状况的实时变化,成功解决了短期流量预测问题。文献[11]针对车辆的编队管理问题设计了一种新的动态规划算法解决编队的资源分配优化问题,在保证误码率的前提下,最小化编队的总传播时延。文献[12]将 Follow Me edge-cloud (FMec) 概念引入自动驾驶领域,利用移动边缘服务架构和软件定义网络,在满足自动驾驶时延需求的同时,最小化全局成本。文献[13]讨论了自动驾驶的定位问题,提出了一种基于车辆区块链和移动边缘服务的 GPS 定位误差演化共享框架,保证了合作伙伴和数据的安全性和可信度,提高了车辆定位精度。

在自动驾驶应用中,大多数研究通常采用强化学习训练自动驾驶模型。文献[14]提出了一种基于强化学习的车辆跟随模型,利用信息共享的优势,提高了在信号交叉口的实时行驶效率和安全性,降低了燃油消耗。文献[15]研究了高速公路上的自动驾驶问题,提出了一种基于多策略决策和强化学习的高速公路合并方法,实现了车辆的安全合并。

自动驾驶系统是一种典型的多智能体系统,随着多智能体技术的兴起,一些研究也将关注点放在分布式多智能体技术与自动驾驶领域的结合上。文献[16]讨论了车辆编队的分散控制问题,结合多智能体理论提出了一个分散控制律补偿通信链路存在的非均匀时变时延效应,使得车辆可以达到指定的队列,同时根据编队队首的速度前进。文献[17]针对自动驾驶环境下的交叉路口通车状况开发了一种联合

优化交通信号灯和车辆轨迹的控制方法，并利用多智能体技术进行仿真，提高车辆通行效率并减少停车次数。文献[18]将自动驾驶汽车的频谱共享检测违规问题定义为一个具有基数的多代理规划问题，利用众包信息确定最佳基数，用尽可能低的成本实现所需的检测精度和定位精度。

尽管现有研究提供了一些关于自动驾驶服务和多智能体技术结合的见解，但是基于移动边缘服务的分布式自动驾驶技术还没有被研究。此外，现有的研究大多数默认自动驾驶服务在数据失效之前完成了数据采集任务，没有考虑不同信息融合技术对自动驾驶服务的影响。与现有研究不同的是，本文研究了具有多样信息融合模式的自动驾驶网络，并在分布式多智能体系统的基础上提出了最优的自动驾驶决策。

为了解决自动驾驶任务数据采集的有效性问题的，本文提出了一种基于移动边缘服务的多维度感知信息融合机制，设计了智能分布式的自动驾驶控制策略，联合优化数据采集和自动驾驶服务。本文的主要贡献如下。

- 1) 提出了一种多维度感知信息融合机制，该机制在融合了移动边缘技术的同时，充分考虑了车辆行驶信息的时空属性。
- 2) 研究了不同信息融合模式对自动驾驶任务的影响，构建了关系模型。
- 3) 设计了智能分布式的协同决策策略，使得信息融合增益和路网交通车流量最大化，并满足自动驾驶应用的成本和资源约束。

2 系统模型

面向自动驾驶任务的车联协同决策模型如图1所示。 I 辆智能车辆在道路上行驶，每个车辆都装有车载传感器，可以探测车辆自身的行驶信息（如速度、位置、加速度等）。道路两边分布有 M 个路边基站（RSU, road site unit），每个RSU都备有移动边缘计算（MEC, mobile edge computing）服务器，可以为车辆提供移动边缘缓存和计算服务。记所有智能车辆的集合为 VA ，所有RSU的集合为 RA 。

其中每个车辆都具有计算能力和缓存能力，可以独立地作为智能体（agent）采集数据，实施信息处理和决策。车辆和RSU都具有通信能力，可以通过V2V通信和V2I通信实现车辆行驶信息的采集和扩散。

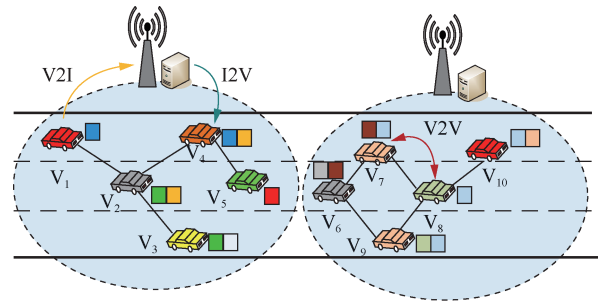


图1 面向自动驾驶任务的车联协同决策模型

2.1 多维信息融合模式

在互联自动驾驶场景中，多维信息融合是数据采集和数据扩散的不同说法。在后文中，信息融合过程将等同于数据采集和数据扩散。

每个车辆都备有传感器，可以将探测到的自身行驶信息缓存在车载单元或路边MEC服务器上，通过V2V通信或V2I通信扩散至其他车辆。大多数车辆产生的行驶信息都具有时空属性，即这些数据只在某一段时间段或某一区域内有效，如有路径规划需求的车辆需要收集当前区域的地理信息，这些地理信息仅适用于当前车辆所处的区域，一旦车辆进入了其他区域，此前收集的地理信息将不适用于新区域的路径规划任务。因此，为了使互联车辆可以很好地完成自身的自动驾驶任务，车辆必须在行驶信息失效之前完成数据采集工作。本文设定车辆的行驶信息在同一RSU通信范围内是有效的，即车辆必须在离开RSU通信范围之前完成其自动驾驶任务的数据采集工作。

为了更好地表示互联车辆之间的通信网络关系，RSU m 通信范围内的车辆互联关系可以表示为车辆通信网络拓扑图 $G_m(VR_m, E_m)$ 。其中， VR_m 为RSU m 通信范围内车辆的集合，共有 q 个节点，每个节点代表一个车辆。边 $E_m(i, j)$ 表示 VR_m 内车辆 i 和车辆 j 之间的互联关系的集合。在车辆通信网络中，当车辆彼此靠近时，车辆可以互相传输自身的数据集。

在车辆边缘网络中，车辆的自动驾驶服务主要有两种多维信息融合模式：1) 车辆借助边缘服务器，将车辆行驶信息缓存在RSU上，其他车辆利用V2I通信完成数据采集工作；2) 智能车辆自身作为边缘节点，存储自身行驶信息，通过V2V通信完成信息融合工作。

当车辆借助边缘服务器完成信息融合工作时，设V2I通信的带宽大小为 B ，被分为 N 个信道。可

以得到车辆 i 采用 V2I 通信通过信道 n 发送信息至 RSU m 的传输速率为

$$l_{i,m,n} = \beta_{i,m,n} \frac{B}{N} \log_2 \left(1 + \frac{P_{i,m,n} \alpha_{i,m,n}}{\sigma^2 + \sum_{i'=1, i' \neq i}^{V_m} y_{i',m} \beta_{i',m,n} P_{i',m,n} \alpha_{i',m,n}} \right) \quad (1)$$

其中, $p_{i,m,n}$ 和 $\alpha_{i,m,n}$ 分别是信道 n 上车辆 i 至 RSU m 的发送功率和信道增益。 σ^2 是噪声功率。 $\beta_{i,m,n}$ 是一个指示变量, 当 $\beta_{i,m,n}=1$ 时, 信道 n 被分配给车辆 i 传输行驶信息至 RSU m ; 当 $\beta_{i,m,n}=0$ 时, 没有分配给车辆 i 和 RSU m 的通信。 $y_{i',m}$ 为车辆 i' 选择借助 RSU m 完成信息融合的概率。

车辆将行驶信息上传至 RSU 会产生通信开销, 设车辆 i 行驶信息的数据大小为 s_i , 单位时间占用 V2I 通信信道的开销为 h_m , 则可以得到车辆 i 通过信道 n 将信息上传至 RSU m 的开销为

$$o_{i,m,n} = h_m \frac{s_i}{l_{i,m,n}} \quad (2)$$

其余车辆 i' 从 RSU m 处获得车辆 i 行驶信息的开销为

$$o_{i',m,i}^{\text{down}} = h_m \frac{s_i}{l_m^{\text{down}}} \quad (3)$$

其中, l_m^{down} 为 RSU m 的下行传输速率。

当车辆选择自身作为边缘节点存储自身行驶信息、通过 V2V 通信完成信息融合工作时, 由于车辆彼此都需要对方的信息执行自动驾驶服务, 此模式下的车辆属于互惠互利的状态。可以不考虑 V2V 通信产生的开销。可以看出, 在条件允许的情况下, 车辆更倾向于选择 V2V 通信的方式传播信息, 从而充分利用车载边缘资源, 使得通信成本最小化, 减轻 RSU 的负担。在该模式下, 车辆 V2V 通信的传输速度为 l_{v2v} 。

2.2 互联车辆自动驾驶应用模型

如图 1 所示, 场景中有 3 个车道, 3 个车道的限速分别为 V_1 、 V_2 和 V_3 。车辆 i 的加速度为 a_i , 制动速度为 b_i , 当前的行驶速度为 v_i , 车辆当前所在车道为 z_i , 为了尽可能提高道路的通行效率, 假设该场景中的车辆均保持最合理的安全距离行驶。根据车辆 i 的行驶信息可以计算出当前时刻车辆 i 与前车的车距为^[19]

$$s_i = s_0 + Tv_i + \frac{v_i \Delta v_i}{2\sqrt{a_i b_i}} \quad (4)$$

其中, s_0 是车与车之间必须保持的硬性车间距离 (即车速为 0 时的最小车间距离, 如等红灯等场景), Δv_i 是车辆 i 与前车的速度差, T 是司机的反应时间。道路车流量 (单位时间通过的车辆数量) 为

$$f = \rho_e v_e \frac{\sum_{i=1}^I v_i}{\sum_{i=1}^I (s_i + l_i)} \quad (5)$$

其中, ρ_e 是车辆的平均密度, v_e 是车辆的平均速度, l_i 是车辆 i 的车身长度。

在互联自动驾驶场景中, 车辆可以通过对自身速度、加速度、车道等的控制最大化道路车流量。

3 面向自动驾驶的多维感知信息融合机制

3.1 多维感知信息融合模式对自动驾驶任务的增益

在互联自动驾驶网络环境中, 如前文所述有两种信息融合策略。由于不同的信息融合模式有不同的时延和开销表现, 信息融合策略会影响自动驾驶服务数据采集工作的效果^[20]。

1) 当车辆处于通信网络拓扑的边缘位置时 (如图 1 中车辆 V_1), 相较于 V2V 通信, 车辆更倾向于选择速度更快、跳数更少的借助边缘服务器的信息融合模式把自身数据扩散至其他车辆。因为本文假设车辆的行驶信息仅在同一 RSU 的通信范围内是有效的, 车辆必须在离开 RSU 通信范围之前完成其自动驾驶任务的数据采集工作。在这种情况下, 若车辆 V_1 选择 V2V 通信的扩散模式, 当数据逐跳扩散至车辆 V_5 时, 车辆 V_5 可能已经离开了 RSU 的通信范围。

2) 当其他车辆留在 RSU 通信范围内的时间较为宽松时 (如图 1 中车辆 V_{10}), 相比于开销更大的 V2I 通信, V2V 通信明显更适合此场景。

由此可见, 不同的信息融合模式对自动驾驶任务有不同的增益。增益越大表示此种信息融合模式越适合当前的场景, 采用此种信息融合模式可以使更多的车辆用较低的开销在离开 RSU 通信范围之前完成其自动驾驶任务的数据采集工作。

本文引入车辆通信网络拓扑图 $G_m(VR_m, E_m)$ 中车辆 i 节点至其余车辆的平均跳数, 计算车辆 i 采

用 V2I 通信完成数据扩散的增益，增益公式为

$$g_{i,m} = \frac{\sum_{j \in VR_m, j \neq i} D_{i,j} S_i}{V_m l_{v2v}} \quad (6)$$

其中， $D_{i,j}$ 为通信网络拓扑图中车辆 i 和车辆 j 之间的最短路径长度， $g_{i,m}$ 代表车辆 i 借助 RSU m 完成数据扩散的增益，即车辆 i 的信息经 V2V 通信逐跳扩散至其余车辆消耗的平均时间。若 $g_{i,m}$ 过大，则表示车辆 i 处于网络边缘，由前文分析可知，此时 V2I 通信更适合车辆 i 。

本文利用除车辆 i 的其余车辆在 RSU m 通信范围内的剩余时间表示车辆 i 通过 V2V 通信完成数据扩散的增益，增益公式为

$$g_{i,v} = \sum_{j \in VR_m, j \neq i} \frac{d_{j,m}}{v_i} \quad (7)$$

其中， $d_{j,m}$ 为车辆 j 在 RSU m 通信范围内的剩余距离， $g_{i,v}$ 表示车辆 i 通过 V2V 通信完成数据扩散的增益。由前文分析可知， $g_{i,v}$ 越大则 V2V 通信越适合车辆 i 。

3.2 自动驾驶问题的马尔可夫过程框架

由于不同的信息融合模式在自动驾驶服务中有不同的增益效果，且车辆对自身驾驶行为的控制决策直接影响自动驾驶应用的实施，在最大化信息融合增益的基础上，路网交通车流量的最大化问题如式(8)所示。

$$\begin{aligned} & \max_{\{i,v,x_{i,v},y_{i,m},\beta_{i,m,n}\}} f \sum_{m=1}^M \sum_{i \in VR_m} (x_{i,v} g_{i,v} + y_{i,m} g_{i,m}) \\ \text{C1: } & \beta_{i,m,n} \in \{0,1\}, i \in VA, m \in RA, n = 1, 2, \dots, N \\ \text{C2: } & \sum_{n=1}^N \beta_{i,m,n} = 1, i \in VA, m \in RA \\ \text{C3: } & x_{i,v} + y_{i,m} = 1, i \in VA \\ \text{C4: } & \text{当 } z_i = 1 \text{ 时, } v_i \leq V_1; \text{ 当 } z_i = 2 \text{ 时, } v_i \leq V_2; \\ & \text{当 } z_i = 3 \text{ 时, } v_i \leq V_3 \\ \text{C5: } & \sum_{j \in VR_m, j \neq i} y_{j,m} o_{i,m,j}^{\text{down}} + y_{i,m} \sum_{n=1}^N \beta_{i,m,n} o_{i,m,n} \leq c_i, i \in VA \\ \text{C6: } & y_{j,m} \left(\sum_{n=1}^N \beta_{j,m,n} \frac{s_j}{l_{j,m,n}} + \frac{s_j}{l_m^{\text{down}}} \right) + \\ & x_{j,v} \frac{s_j}{l_{v2v}} D_{i,j} \leq \frac{d_{i,m}}{v_i}, i, j \in VR_m \\ \text{C7: } & 0 \leq x_{i,v}, y_{i,m} \leq 1 \end{aligned} \quad (8)$$

其中， $x_{i,v}$ 是车辆选择 V2V 通信的概率， $y_{i,m}$ 是车辆选择 V2I 通信的概率， c_i 是车辆 i 预期的最大开销。C1 表示指示变量 $\beta_{i,m,n}$ 的定义域；C2 表示车辆只能选择一个信道实现与 RSU 的传输；C3 表示车辆必须选择一种数据扩散方式，将自身的行驶信息传输至其他车辆；C4 表示车辆速度应不大于所处车道的最高限速；C5 表示车辆 i 的通信成本应小于车辆 i 预期的最大开销；C6 表示车辆 i 需要在离开 RSU 通信范围之前，获得车辆 j 的行驶信息；C7 表示决策变量的定义域。

4 面向自动驾驶的智能分布式决策算法

4.1 分布式马尔可夫决策过程框架

自动驾驶系统实际上是一个分布式的多智能体系统，不是简单的单车系统或集中式决策系统。因此，相比于集中式的马尔可夫决策过程 (MDP, Markov decision process) 框架，分布式的多车马尔可夫决策过程 (MVMDP, multi-vehicle MDP) 框架明显更符合实际需求^[20]。

MVMDP 框架可以定义为 $\langle S, A, p, u \rangle$ ， S 是一个有限的状态集合， A 是一个有限的动作集合， p 是执行动作 $a (a \in A)$ 后，状态 $s (s \in S)$ 到 $s' (s' \in S)$ 的转移概率， u 是执行动作 a 后获得的即时奖励， π 表示一个从状态映射到动作的策略。提出 MVMDP 框架的目的旨在找到最优策略 π^* ，使得奖励函数最大化。

在 MVMDP 框架中，每个智能车辆被视为一个 agent，与未知环境进行交互，获取经验，然后迭代学习，最终得到最优策略。所提出的框架中的详细元组定义如下。

1) 状态集合：每个智能车辆在 l 时刻的状态集合记为 S_l^i ，车辆只能通过自身对环境的观察获得局部的状态集合。本文定义车辆的局部状态空间包括互联车辆当前的驾驶行为信息（如加速度、速度等）和车辆当前的信息融合策略（如信道选择和通信方式选择等）。

2) 动作集合：在互联自动驾驶场景中的动作应该是车辆对自身驾驶行为的调整控制和对信息融合模式策略的调整控制。因为车辆驾驶行为和信息融合模式选择概率的变量空间是连续的，所以本文选择多智能体深度确定性策略梯度 (MADDPG, multi-agent deep deterministic policy gradient) 算法

解决 MVMDP 问题。

3) 奖励函数: 在本文中, 自动驾驶应用的目的是最大化道路车流量、优化信息融合模式。因此, 奖励函数应该考虑两方面: 一是不同信息融合模式对自动驾驶服务的增益; 二是如式(5)所示的单位时间通过道路的车辆数量。

由于传统的基于奖励值的算法, 如差分法 (TD, temporal difference), 在实际应用中存在一些不足, 不能有效地处理连续动作空间中的任务, 且最终的解可能不是全局最优的, 本文选择了一种基于深度确定性策略梯度 (DDPG, deep deterministic policy gradient) 的算法与多智能体系统结合, 求解 MVMDP 问题。

4.2 基于多智能体深度确定性策略梯度的车联协同决策算法

DDPG 架构如图 2 所示, DDPG 是一种在学习过程中同时学习策略函数和价值函数的策略梯度深度强化学习算法。DDPG 的 agent 通过策略梯度法估计策略权值, 直接从未处理的观察空间中进行学习, 同时使用 Actor-Critic 模型学习价值函数, 更新 Actor 模型。DDPG 采用随机行为策略进行策略探索, 且是对确定性的目标策略进行估计, 大大降低了学习的复杂性。此外, DDPG 选择的操作来自连续空间, 这个空间适用于式(8)中的自动驾驶策略值。作为一种强大的学习方法, DDPG 已被广泛应用于频谱管理、无线能量采集、交通信号控制等领域。在车载网络环境下, DDPG 被用于引导无人驾驶车辆的航向跟踪、调整车辆悬架系统、控制自动驾驶行为等领域。

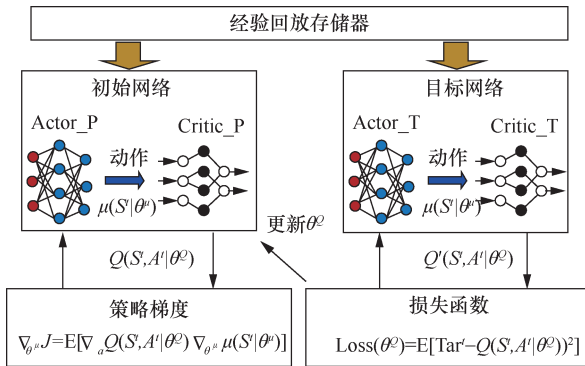


图 2 DDPG 架构

在 MADDPG 算法中, 单个 agent 的训练同单个 DDPG 算法的训练过程类似, 不同之处主要体现在 Critic 的输入上: 在 DDPG 算法中, Critic 的输

入是一个状态—动作对信息; 在 MADDPG 中, 每个 agent 的 Critic 输入除了自身的状态—动作对信息, 还有额外的信息, 如其他 agent 的动作信息。

将原始问题(8)构建为 MADDPG 过程, 每个车辆作为智能体, l 时刻车辆 i 的局部动作空间为

$$A_i^l = \{ \Delta v_i^l, \Delta x_{i,v}^l, \Delta y_{i,m}^l, \beta_{i,m,n}^l \} \quad (9)$$

l 时刻车辆 i 的局部状态空间为

$$S_i^l = \{ v_i^l, x_{i,v}^l, y_{i,m}^l, \beta_{i,m,n}^l \} \quad (10)$$

在 l 时刻, 当车辆 i 在状态 S_i^l 下执行了动作 A_i^l 时, 其局部奖励函数定义为

$$u_i^l = \frac{\sum_{i \in VR_m} v_i^l}{\sum_{i \in VR_m} (s_i^l + l_i)} (x_{i,v}^l g_{i,v} + y_{i,m}^l g_{i,m}) \quad (11)$$

由此可以构建一个状态空间为 $S = \{S_{v_0}^l, \dots, S_i^l\}$, 动作空间为 $A = \{A_{v_0}^l, \dots, A_i^l\}$ 的多智能体 DDPG 决策过程。基于 MADDPG 和多维信息融合的车联协同决策算法的流程如算法 1 所示。

算法 1 面向自动驾驶应用的多智能体车联协同决策算法

for episode = 1 to K do

 初始化环境信息, 得到初始状态集合 s^0

 for $t = 1$ to max-episode-length do

 对于每个 Agent i , 根据 Actor-Critic 模型得到动作集合 a_i^t , 执行动作 a_i^t , 根据式(11)计算得到奖励值 u_i^t 和新的状态集合 s_i^{t+1} , 将 $(s_i^t, a_i^t, u_i^t, s_i^{t+1})$ 存储在经验缓存区中 $s_i^t = s_i^{t+1}$

 for Agent $i = 1$ to I do

 利用式(12)更新 Critic 网络

$$\min L(\theta_i) = E_{s,a,u,s'} \left[\left(y - Q_i^\pi(s_i, a_1, \dots, a_q) \right)^2 \right] \quad (12)$$

$$y = u_i + \gamma Q_i^\pi(s_i', a_1', \dots, a_q') |_{a_i' = \pi_i'(o_i)}$$

 利用梯度公式(13), 更新 Actor

$$\begin{aligned} \nabla_{\theta_i} J(\mu_i) &= E[\nabla_{\theta_i} \pi_i(a_i | o_i) \cdot \\ \nabla_{a_i} Q_i^\pi(s_i, a_1, \dots, a_{V_m}) &|_{a_i = \pi_i(o_i)}] \end{aligned} \quad (13)$$

 end for

 更新每个智能体的策略网络参数

$$\theta_i' = \tau \theta_i + (1 - \tau) \theta_i^t \quad (14)$$

end for

end for

其中, θ_n 表示 n 个智能体策略的参数, o_i 表示智能体 i 的观测值, Q_i^r 表示智能体 i 的状态—动作函数。

5 仿真结果与性能分析

本节对所提出的面向自动驾驶应用的多智能体车联协同决策算法的性能进行评估。本文采用 MADDPG 算法解决式(8)的最优化问题, 令智能车辆自身作其代理人训练数据集, 得到联合信息融合和路网交通车流量优化的协同决策, 最终车辆得到的奖励函数值反映了最大化信息融合增益的最大路网交通车流量。在多智能体算法的执行过程中, 智能体之间交互的信息参数数据量很小, 可以忽略不计, 在本文中不考虑智能体之间的信息传输性能对多智能体算法的影响。

本文考虑 5 辆智能车辆, 其预期的最大开销从 (100, 200) 中随机取值, 车辆自动驾驶行驶信息的数据大小从 (500, 600) MB 中随机取值, 车辆的最大加速度为 0.73 m/s^2 , 车辆理想的制动减速度为 1.67 m/s^2 , 司机的反应时间为 1.6 s, 3 个车道的限速分别为 30 km/h、45 km/h、60 km/h。考虑一个 RSU 分布在路旁, 设其有 5 个信道可以用来上传数据, 信道

带宽为 0.3 MHz, 噪声功率 σ 为 -114 dBm , 每个车辆的发送功率为 24 dBm ^[21]。

面向自动驾驶应用的多智能体车联协同决策算法的收敛性如图 3 所示。横轴为系统的训练次数, 纵轴为 agent 的奖励除以训练次数的值。图 3 分别给出了不同 agent 在不同预期开销下的收敛曲线。由图 3 可以看出, 随着训练次数的增加, 每个 agent 的奖励除以训练次数的值在不同的环境下收敛。在训练次数为 10 000 左右即可达到收敛状态, 证明了本文提出的面向自动驾驶应用的基于 MADDPG 的车联协同决策算法在不同环境下都具有收敛性。

不同算法的系统收益对比如图 4 所示, 比较了不同 agent 采用本文提出的基于 MADDPG 的算法、随机选择算法和固定信息融合模式算法的系统收益。本文所提算法结合了智能的多维信息融合模式选择和 MADDPG 算法, 使得每个 agent 可以自适应地优化自动驾驶策略。在随机选择算法中, 随机选择信息融合模式。在固定信息融合模式算法中, 忽略了信息融合模式对自动驾驶任务收益的影响。由于在每个 agent 中, 随机和固定信息融合模式算法都比基于 MADDPG 算法的系统收益低, 因此本文提出的面向自动驾驶任务的基于 MADDPG 和信

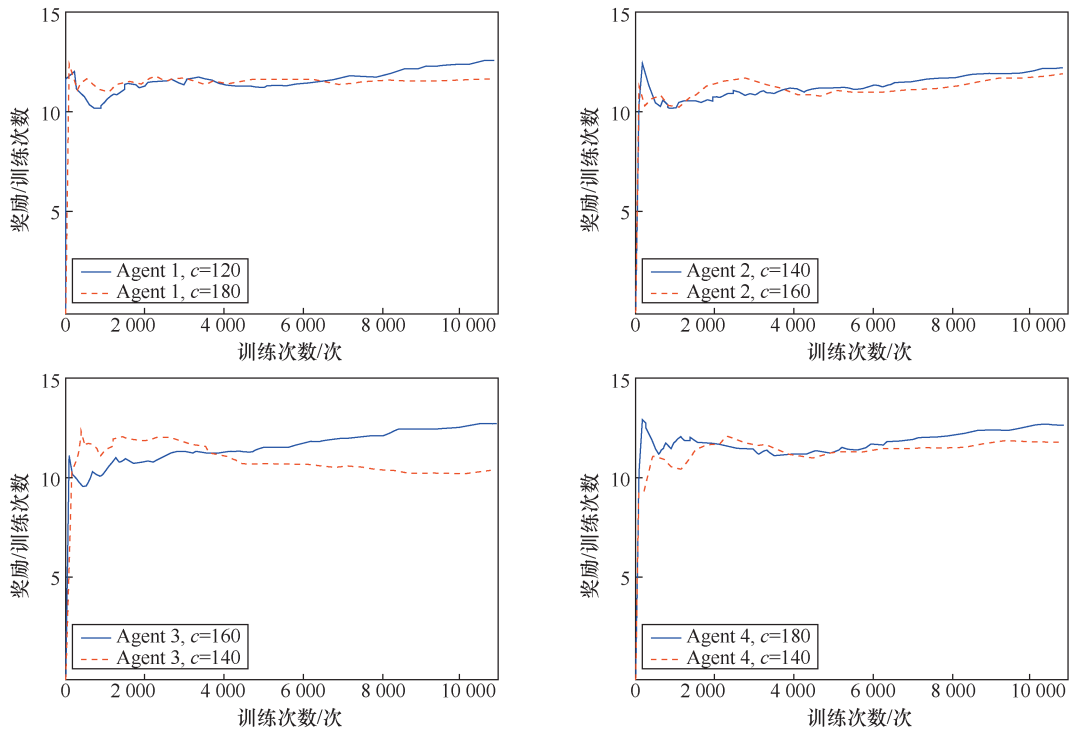


图3 面向自动驾驶应用的多智能体车联协同决策算法的收敛性

息融合的车联协同决策算法可以大大提升自动驾驶的系统收益。

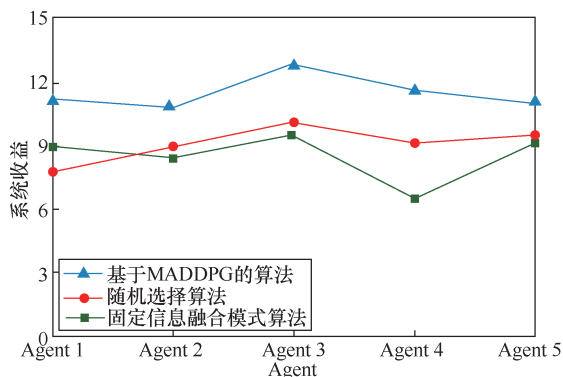


图4 不同算法的系统收益对比

智能车辆在不同预期开销下选择 V2V 通信的概率如图 5 所示。可以看出，随着预期开销的增大，agent 选择 V2V 通信的概率降低。这是因为与 V2V 通信相比，V2I 通信的时延更小，车辆采用 V2I 通信更容易在行驶数据失效前完成数据采集工作。但是 V2I 通信具有较高的成本开销，若车辆的预期开销有限，车辆更倾向于选择 V2V 通信。可以看出，随着车辆预期开销的增大，不同 agent 都呈现出倾向选择 V2I 通信的趋势。

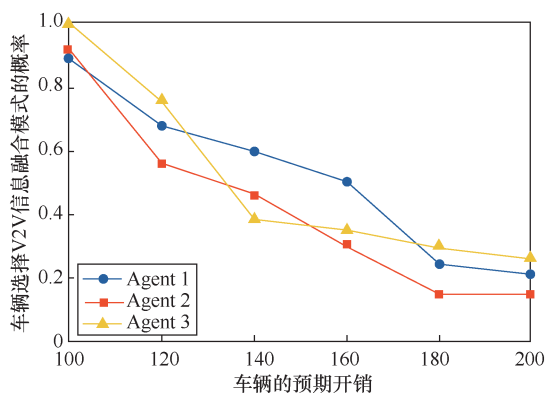


图5 智能车辆在不同预期开销下选择 V2V 通信的概率

6 结束语

本文将移动边缘技术融入了车联网，提出了一种多维度感知信息融合机制，并充分考虑了车辆行驶信息的时空属性。本文设计了面向自动驾驶的智能分布式协同决策方案，使得信息融合增益和路网交通车流量最大化，并满足网联车辆的成本和资源约束。仿真结果表明，与传统算法相比，本文提出的方案显著降低了学习成本。

参考文献:

- [1] ANGSKOG P, NASMAN P, MATSSON L G. Resilience to intentional electromagnetic interference is required for connected autonomous vehicles[J]. IEEE Transactions on Electromagnetic Compatibility, 2019, 61(5): 1552-1559.
- [2] CHEN Y M, LU C, CHU W B. A cooperative driving strategy based on velocity prediction for connected vehicles with robust path-following control[J]. IEEE Internet of Things Journal, 2020, 7(5): 3822-3832.
- [3] XU H L, FENG S, ZHANG Y, et al. A grouping-based cooperative driving strategy for CAVs merging problems[J]. IEEE Transactions on Vehicular Technology, 2019, 68(6): 6125-6136.
- [4] DU J B, YU F R, CHU X L, et al. Computation offloading and resource allocation in vehicular networks based on dual-side cost minimization[J]. IEEE Transactions on Vehicular Technology, 2019, 68(2): 1079-1092.
- [5] ZHANG Y M, ZHANG G H, FIERRO R, et al. Force-driven traffic simulation for a future connected autonomous vehicle-enabled smart transportation system[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(7): 2221-2233.
- [6] CHEN Y M, ZHA J Q, WANG J M. An autonomous T-intersection driving strategy considering oncoming vehicles based on connected vehicle technology[J]. ASME Transactions on Mechatronics, 2019, 24(6): 2779-2790.
- [7] VAIO M D, FALCONE P, HULT R, et al. Design and experimental validation of a distributed interaction protocol for connected autonomous vehicles at a road intersection[J]. IEEE Transactions on Vehicular Technology, 2019, 68(10): 9451-9465.
- [8] XING Y, LV C, CAO D P. Personalized vehicle trajectory prediction based on joint time-series modeling for connected vehicles[J]. IEEE Transactions on Vehicular Technology, 2020, 69(2): 1341-1352.
- [9] WU B, QIAN L J, LU M L, et al. Optimal control problem of multi-vehicle cooperative autonomous parking trajectory planning in a connected vehicle environment[J]. IET Intelligent Transport Systems, 2019, 13(11): 1677-1685.
- [10] YANG S R, SU Y J, CHANG Y Y, et al. Short-term traffic prediction for edge computing-enhanced autonomous and connected cars[J]. IEEE Transactions on Vehicular Technology, 2019, 68(4): 3140-3153.
- [11] LI K, NI W, TOVAR E, et al. Optimal rate-adaptive data dissemination in vehicular platoons[J]. IEEE Transactions on Intelligent Transportation Systems, 2019: 1-11.
- [12] AISSIOUI A, KSENTINI A, GUEROUI A M, et al. On enabling 5G automotive systems using follow me edge-cloud concept[J]. IEEE Transactions on Vehicular Technology, 2018, 67(6): 5302-5316.
- [13] LI C L, FU Y C, YU F R, et al. Vehicle position correction: a vehicular blockchain networks-based GPS error sharing framework[J]. IEEE Transactions on Intelligent Transportation Systems, 2020: 1-15.
- [14] ZHOU M F, YU Y, QU X B. Development of an efficient driving strategy for connected and automated vehicles at signalized intersections: a reinforcement learning approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(1): 433-443.
- [15] NISHI T, DOSHI P, PROKHOROV D. Merging in congested freeway

- traffic using multipolicy decision making and passive actor-critic learning[J]. IEEE Transactions on Intelligent Vehicles, 2019, 4(2): 287-297.
- [16] SOUZA F O D, TORRES L A B, MOZELLI L A, et al. Stability and formation error of homogeneous vehicular platoons with communication time delays[J]. IEEE Transactions on Intelligent Transportation Systems, 2019: 1-12.
- [17] 戴荣健, 丁川, 鹿应荣, 等. 自动驾驶环境下车辆轨迹及交通信号协同控制[J]. 汽车安全与节能学报, 2019, 10(4): 531-539.
- DAI R J, DING C, LU Y R, et al. Cooperated control of signal and vehicle trajectory under the autonomous vehicle environment[J]. Automotive Safety and Energy, 2019, 10(4): 531-539.
- [18] CAREEM M A A, DUTTA A, WANG W F. Spectrum enforcement and localization using autonomous agents with cardinality[J]. IEEE Transactions on Cognitive Communications and Networking, 2019, 5(3): 702-715.
- [19] RO J W, ROOP P S, MALIK A, et al. A formal approach for modeling and simulation of human car-following behavior[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(2): 639-648.
- [20] YANG L, ZHANG L L, HE Z J, et al. Efficient hybrid data dissemination for edge-assisted automated driving[J]. IEEE Internet of Things Journal, 2020, 7(1): 148-159.
- [21] NING A L, DONG P R, KONG X J, et al. A cooperative partial computation offloading scheme for mobile edge computing enabled Internet of things[J]. IEEE Internet of Things Journal, 2019, 6(3): 4804-4814.

[作者简介]



曹佳钰（1995- ），女，山西大同人，电子科技大学硕士生，主要研究方向为车联网和移动边缘计算。



冷颢鹏（1973- ），男，四川资中人，电子科技大学教授、博士生导师，主要研究方向为物联网、车联网、新一代宽带无线网络、无线自组织网、智能交通信息系统的资源管理、介质访问控制、路由、组网与互联、智能算法理论及技术应用等。



张科（1978- ）男，重庆人，博士，电子科技大学讲师，主要研究方向为物联网、车联网边缘计算与存储。