Multi-Agent Learning Empowered Collaborative Decision for Autonomous Driving Vehicles

Jiayu Cao, Supeng Leng and Ke Zhang
School of Information and Communication Engineering
University of Electronic Science and Technology of China
liuda945@163.com, {spleng,zhangke}@uestc.edu.cn

***Abstract*—Autonomous vehicles play an important role 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 empowered collaborative decision provides a potential approach to address the problem. This paper proposes a multi-dimensional information fusion mechanism, which improves the utilities of vehicular information processing and autonomous driving. Moreover, we design an intelligent distributed decision algorithm for autonomous driving applications, which optimizes road traffic flow under vehicular resource constraints. Numerical results demonstrated that our proposed scheme significantly increases the system revenue.**

***Index Terms*—Autonomous driving, Multi-agent, Information fusion, Mobile edge service**

I. Introduction

Intelligent Transportation Systems (ITS) has emerged as an appealing paradigm to solve traffic problems, such as road congestion, environmental pollution, and energy consumption, with Connected and Autonomous Vehicle (CAV) technology as its important part [1] [2]. In the operation of CAV, vehicle connected technology and autonomous driving technology, which realize vehicular network communication and driving status detection, constitute the technical basis of traffic management and scheduling [3].

Along with the proliferation of smart vehicles and powerful traffic sensors, massive data set such as location information and map navigation information will be detected and analyzed. These information normally has spatial-temporal characteristics, that is, it is only valid whitin a certain time or area. Since the independent vehicle's autonomous driving control decision is closely related to the driving state of other vehicles, the autonomous driving application requires accurate real-time driving state information to further process the control strategy, whose state collection task must be completed before the data is invalid. However, the constrained environmental perception range of an individual vehicle seriously undermine the effectiveness of the state collection.

Mobile edge services have emerged as a potential approach to address the above problem, which cache the traffic information on edge nodes in proximity to the vehicles for supporting low-latency data acquisition. In [4], the authors proposed a short-term traffic prediction model based on the mobile edge technology. In [5], the authors introduced mobile edge service in the optimization problem of platoon resource allocation. Although mobile edge technology brings an efficient way to make up for the constrained perception range of smart vehicles, various information fusion modes and limited edge resources in mobile edge services still restrict the effectiveness of data acquisition tasks.

In autonomous driving applications, most existing works usually adopt reinforcement learning to train autonomous driving models. In [6], the authors proposed a vehicle following model based on reinforcement learning to improve the real-time driving efficiency at signalized intersections. In [7], the authors developed a highway merging method based on multi-strategy decision and reinforcement learning. However, the real autonomous driving system is not a simple single-vehicle system or centralized decision-making system, but actually a distributed multi-agent system. Therefore, it is still a critical challenge to make distributed collaborative decision for autonomous vehicles while achieving efficient data acquisition.

Although previous studies have provided some insights about autonomous driving services [8] [9], mobile edge services empowered distributed autonomous driving technology has not been well investigated. Moreover, autonomous driving services in previous works usually assume that data collection tasks have been completed before the data is invalid, and inflexible information fusion mode can not match CAV application scenarios with highly dynamic topology and multiple transmission modes. To fill this gap, we propose a multi-dimensional information fusion mechanism and design an intelligent distributed decision algorithm for autonomous driving applications, which jointly optimizes data acquisition and road traffic flow.

The main contribution of this work is listed as follows.

- We propose a multi-dimensional information fusion mechanism, which combines mobile edge services with the spatial-temporal characteristics of vehicle travel information.
- We investigate the impact of diverse information fusion modes on autonomous driving tasks and derive the relational model.
- We design an intelligent distributed decision algorithm for autonomous driving applications, which optimizes

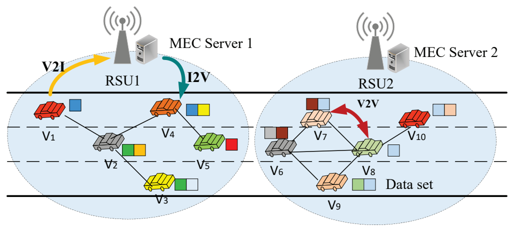

Fig. 1. Collaborative decision model for autonomous driving tasks

information fusion gain and traffic flow under vehicular resource constraints.

The rest of the paper is organized as follows. System model is presented in Section II. In Section III, we propose a multi-dimensional information fusion mechanism. In Section IV, we design an intelligent distributed decision algorithm for autonomous driving applications. We present numerical results in Section V and conclude the paper in Section VI.

II. SYSTEM MODEL

Fig. 1 shows the framework of collaborative decision for autonomous driving tasks. There are I smart vehicles running on the roads. Each vehicle is equipped with on-board sensors that can detect vehicle traffic information. M RSUs are randomly distributed along the roads. Each RSU is equipped with an MEC server, which provides the computing and caching services for the vehicles. Each vehicle has computing and caching capabilities and can independently collect data as an agent to implement collaborative decision. Both vehicles and RSUs have communication capabilities, and can distribute and collect vehicle travel information through V2V and V2I technologies.

In the multi-dimensional information fusion model, the travel information generated by CAV normally has spatial-temporal characteristics, which only reflects the vehicle state in a specific time or space. For example, vehicles with road planning demands require to collect geographic information of the current area, which is only applicable to the area where the current vehicle is located. Once the vehicle enters other areas, the previously collected geographic information will be invalid. Therefore, the state collection task must be completed before the data is invalid. In this paper, we assume that vehicle driving information is valid only within the same RSU range, and vehicle must complete the data acquisition task before leaving the RSU coverage area.

The vehicle interconnection relationship within the range of RSU m can be presented as a vehicular communication network topology diagram $G_m(V_m, E_m)$, where V_m is the set of vehicles covered by RSU m, and each node represents a vehicle. (i, j) presents the interconnection between vehicle i and j. Vehicles can share their data sets with the other vehicles when they are in close proximity.

In vehicular edge network, there are mainly two multi-dimensional information fusion modes. One is that vehicle leverages the edge server to cache the travel information on RSU, and the other vehicles leverage V2I technology to complete the data acquisition; The other is that smart vehicle acts as the edge node itself to store its own travel information, and completes information fusion through V2V technology.

When vehicle completes information fusion with leveraging edge server, the bandwidth of V2I technology is set as B, which is divided into N channels. The transmission rate of vehicle i to RSU m through n channel is

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

where $p_{i,m,n}$ and $\alpha_{i,m,n}$ are the transmission power and channel gain of vehicle i to RSU m on n channel, respectively. σ^2 is the noise power. $\beta_{i,m,n}$ is an indicator variable, when $\beta_{i,m,n} = 1$, channel n is assigned to vehicle i to transmit information to RSU m. $y_{i',m}$ is the probability that vehicle i' chooses to complete information fusion with RSU m.

The data size of autonomous driving information of vehicle i is s_i. The unit price of the V2I channel per unit time is h_m. The cost of vehicle i uploading the information to RSU m through channel n is

$$o_{i,m,n} = h_m\frac{s_i}{l_{i,m,n}}, \tag{2}$$

The cost for other vehicle i' to obtain travel information of vehicle i from RSU m is

$$o_{i',m,i}^{down} = h_m\frac{s_i}{l_m^{down}}, \tag{3}$$

$l_m^{down}$is the offloading transmission rate of RSU m.

When the vehicle selects itself as edge node to store its travel information and completes the information fusion through V2V technology, the communication cost generated by V2V can be ignored in this mode because each vehicle needs each other's information to perform their automatic driving service. Transmission speed of vehicle V2V communication is l_{v2v}.

In the autopilot application model, there are 3 lanes in the scene in Fig. 1. The speed limit of 3 lanes is V_1, V_2 and V_3 respectively. The acceleration of vehicle i is a_i, the braking speed is b_i, the current driving speed is v_i, position information is x_i and the current lane of the vehicle is z_i. The distance between vehicle i and vehicle ahead $i+1$ is

$$s_i = x_{i+1} - x_i + (v_{i+1} - v_i)\tau, \tag{4}$$

τ is the duration of the decision. l_i is the length of vehicle. The road traffic flow (the number of vehicles passing through per unit time) is

$$f = \sum_{i=1}^{I} v_i / \sum_{i=1}^{I}(s_i + l_i). \tag{5}$$

III. MULTI-DIMENSIONAL INFORMATION FUSION MECHANISM FOR AUTOPILOT APPLICATION

Due to various information fusion modes have different delay and cost performance, the information fusion strategy will affect the effectiveness of data collection for autonomous driving services [10].

On the one hand, when vehicle is at the edge of communication network topology (such as vehicle V_1), compared with the V2V fusion method, the vehicle tends to choose information fusion mode with edge server, which can bring faster data diffusion and shorter propagation hops. For example, if vehicle V_1 selects the V2V diffusion mode, vehicle V_5 may have left the RSU communication range when data is propagated to vehicle V_5 by hopping.

On the other hand, when the remaining travel time of other vehicles within RSU coverage is relatively loose (such as vehicle V_{10}), V2V diffusion mode is obviously more suitable for this situation than V2I mode.

The gain of vehicle i selects V2I diffusion mode is

$$g_{i,m} = \frac{\sum\limits_{j=i,j\neq i}^{V_m} D_{i,j} s_i}{V_m l_{v2v}}, \tag{6}$$

where, $D_{i,j}$ is the shortest path length between vehicle i and j in the communication topology diagram. The gain of vehicle i selects V2V diffusion mode is

$$g_{i,v} = \sum_{j=i,j\neq i}^{V_m} \frac{d_{j,m}}{v_j}, \tag{7}$$

where, $d_{j,m}$ is the remaining distance of vehicle j within the coverage area of RSU m.

The optimization problem of simultaneously maximizing information fusion gain and traffic flow on the road network under vehicular resources constraints can be formulated as

$$\begin{aligned}
&\max_{\{v_i,x_{i,v},y_{i,m},\beta_{i,m,n}\}} \quad f \sum_{m=1}^{M}\sum_{i=1}^{V_m}\left(x_{i,v}g_{i,v}+y_{i,m}g_{i,m}\right)\\
&C1 \quad \beta_{i,m,n}\in\{0,1\}, i\in I, m\in M, n\in N\\
&C2 \quad \sum_{n=1}^{N}\beta_{i,m,n}=1, i\in I \ , m\in M;\\
&C3 \quad x_{i,v}+y_{i,m}=1, i\in I \ ;\\
&C4 \quad v_i\le V_1, when z_i=1; v_i\le V_2, when z_i=2;\\
&\qquad v_i\le V_3, when z_i=3;, i\in I\\
&C5 \quad \sum_{j=1,j\neq i}^{V_m} y_{j,m}o_{i,m,j}^{down}+y_{i,m}\sum_{n=1}^{N}\beta_{i,m,n}o_{i,m,n}\le c_i, i\in I\\
&C6 \quad y_{j,m}\left(\sum_{n=1}^{N}\beta_{j,m,n}\tfrac{s_j}{l_{j,m,n}}+\tfrac{s_j}{l_m^{down}}\right)+\\
&\qquad x_{j,v}\tfrac{s_j}{l_{v2v}}D_{i,j}\le\tfrac{d_{i,m}}{v_i}, i,j\in V_m\\
&C7 \quad 0\le x_{i,v}, y_{i,m}\le 1, i\in I\\
&C8 \quad s_i\ge s_0+Tv_i+\tfrac{v_i\Delta v_i}{2\sqrt{a_i b_i}}, i\in I
\end{aligned} \tag{8}$$

$x_{i,v}$ is the probability that vehicle chooses V2V diffusion mode, $y_{i,m}$ is the probability that vehicle chooses V2I diffusion mode. c_i is the maximum expected cost of vehicle i. s_0 is the hard distance that must be maintained between vehicles (the minimum distance when vehicle speed is 0, such as waiting for a red light, etc.), Δv_i is the speed difference between vehicle i and its preceding car, and T is the reaction time of the driver. Constraint C1 gives the domain of $\beta_{i,m,n}$. Constraint C2 indicates that vehicle can only choose one channel to transmit with RSU; Constraint C3 indicates that vehicle must choose a data diffusion method. Constraint C4 indicates the speed constraint. Constraint C5 indicates the cost constraint. Constraint C6 indicates that vehicle i needs to obtain automatic driving information of vehicle j before leaving the RSU communication range. Constraint C7 gives the domain of $x_{i,v}$, $y_{i,m}$. Constraint C8 indicates that the vehicle-vehicle distance needs to be greater than the minimum safety distance.

IV. INTELLIGENT DISTRIBUTED DECISION ALGORITHM

Each vehicle is an agent, and the local action space of vehicle i at time slot l is $A_i^l=\{\Delta v_i^l, \Delta x_{i,v}^l, \Delta y_{i,m}^l, \beta_{i,m,n}^l\}$. The local state space of vehicle i at time l is $S_i^l = \{v_i^l, x_{i,v}^l, y_{i,m}^l, \beta_{i,m,n}^l\}$. At time l, when vehicle i performs action A_i^l in state S_i^l, its local reward function is defined as

$$u_i^l = \sum_{i=1}^{V_m} v_i^l \Big/ \sum_{i=1}^{V_m}\left(s_i^l+l_i\right)\left(x_{i,v}^l g_{i,v}+y_{i,m}^l g_{i,m}\right) \tag{9}$$

The main steps of the proposed intelligent distributed decision algorithm are shown in Algorithm 1.

Algorithm 1 Intelligent distributed decision algorithm for autonomous driving applications

1: **for** episode = 1 to K **do**
2: Initialize the environment information to obtain the initial state set.
3: **for** t = 1 to max-episode-length **do**
4: For each agent i, get action set A_i^l according to the Actor-Critic model, calculate reward u_i^l and new state set S_i^{l+1} according to (9). Store $\left(s_i^l, a_i^l, u_i^l, s_i^{l+1}\right)$ in the experience cache.
5: $s_i^l = s_i^{l+1}$
6: For each agent i, update the Actor and Critic network based on the actions of other agents.

$$\begin{aligned}
&\min L(\theta_i) = E_{s,a,u,s'}\left[\left(y-Q_i^{\pi}\left(s_i,a_i,...,a_{V_m}\right)\right)^2\right]\\
&y = u_i+\gamma Q_i^{\pi'}\left({s_i}',{a_i}',...,{a_{V_m}}'\right)|_{{a'}_i-{\pi'}_i(o_i)}\\
&\nabla_{\theta_i}J(\mu_i) = E[\nabla_{\theta_i}\pi_i\left(a_i|o_i\right)\\
&\nabla_{a_i}Q_i^{\pi}\left(s_i,a_1,...,a_{V_m}\right)|_{a_i=\pi_i(o_i)}]
\end{aligned} \tag{10}$$

7: Update the policy network parameters of each agent.

$$\theta'_i = \lambda\theta_i + (1-\lambda)\theta'_i \tag{11}$$

8: **end for**
9: **end for**

θ_n indicates the parameters of n agent strategy, and o_i indicates the observed value of agent i. Q_i^{π} indicates the state-action function of agent i.

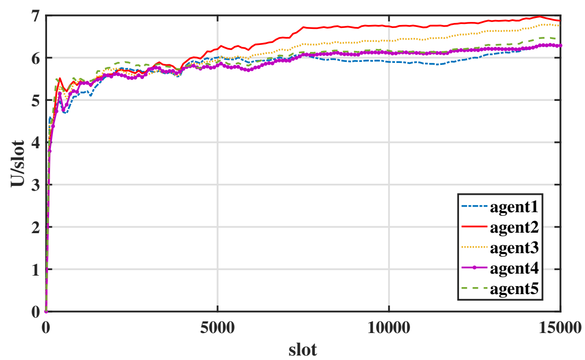

Fig. 2. Distributed collaborative decision convergence diagram

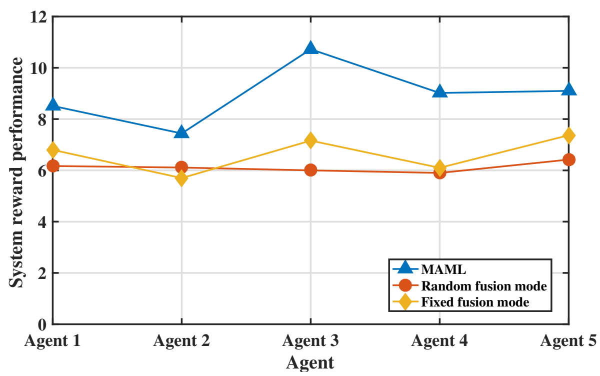

Fig. 3. Comparison of reward performance with different schemes

V. NUMERICAL RESULTS

We consider five intelligent vehicles with the intelligent vehicle acting as its own agent, and the expected maximum cost is randomly taken from (100,200) units. The data size of travel information of vehicle is randomly taken from (500,600) units. The maximum acceleration of vehicle is 0.73 m/s^2, and the ideal braking deceleration of vehicle is 1.67 m/s^2. The driver's reaction time is 1.6 s. The speed limit of 3 three lanes is 30 km/h, 45 km/h and 60 km/h, respectively. There are five channels available for uploading data.

Fig. 2 shows the convergence of multi-agent learning empowered collaborative decision algorithm for autonomous driving applications. The convergence curves of different agents are given respectively. It can be seen that the historical accumulative revenue of each Agent tends to converge as the training slot increases. The convergence state can be reached when the training slot is around 10000. It is proved that the MADDPG empowered collaborative decision for automatic driving applications proposed in this paper is convergent.

Fig. 3 compares system reward performance of the proposed scheme with random fusion mode scheme and fixed fusion mode scheme. MADDPG empowered collaborative decision scheme yields the highest rewards, as it combines multi-dimensional information fusion mode selection and MADDPG algorithm, while adaptively optimizing automatic driving collaborative decision strategies according to various expected costs of different vehicles. In random fusion mode scheme, information fusion mode and automatic driving collaborative decision strategies are selected randomly. In fixed information fusion mode scheme, the influence of multiple information fusion mode on the revenue of automatic driving task is ignored. In each Agent, the system revenue of MADDPG empowered algorithm is higher than the system revenue of random selection scheme and fixed information fusion mode scheme . Therefore, MADDPG empowered collaborative decision scheme for autonomous driving tasks proposed in this paper can greatly improve the system revenue of autonomous driving.

VI. CONCLUSION

In this paper, we introduced mobile edge services into the Internet of vehicles, and proposed a multi-dimensional information fusion mechanism which fully considered their spatial-temporal characteristics. An intelligent distributed collaborative decision scheme for autonomous driving was designed to maximize information fusion gain and traffic flow under the resource constraints. Numerical results demonstrated that the system revenue of our proposed collaborative decision scheme is 37% and 25% higher than that of random scheme and fixed fusion mode scheme, respectively.

REFERENCES

[1] P. Ängskog, P. Näsman and L. Mattsson, "Resilience to Intentional Electromagnetic Interference Is Required for Connected Autonomous Vehicles," IEEE Transactions on Electromagnetic Compatibility, vol. 61, no. 5, pp. 1552-1559, Sep. 2019.

[2] Y. Chen, C. Lu and W. Chu, "A Cooperative Driving Strategy Based on Velocity Prediction for Connected Vehicles With Robust Path-Following Control," IEEE Internet of Things Journal, vol. 7, no. 5, pp. 3822-3832, Jan. 2020.

[3] Y. Xing, C. Lv and D. Cao., "Personalized Vehicle Trajectory Prediction Based on Joint Time-Series Modeling for Connected Vehicles," IEEE Transactions on Vehicular Technology, vol. 69, no. 2, pp. 1341-1352, Dec. 2019.

[4] S. Yang, Y. Su, Y. Chang et al., "Short-Term Traffic Prediction for Edge Computing-Enhanced Autonomous and Connected Cars," IEEE Transactions on Vehicular Technology, vol. 68, no. 4, pp. 3140-3153, Feb. 2019.

[5] K. Li, W. Ni, E. Tovar and M. Guizani, "Optimal Rate-Adaptive Data Dissemination in Vehicular Pla-toons," IEEE Transactions on Intelligent Transportation Systems, accepted, Sep. 2019.

[6] M. Zhou, Y. Yu and X. Qu, "Development of an Efficient Driving Strategy for Connected and Automated Vehicles at Signalized Intersections: A Reinforcement Learning Approach," IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 1, pp. 433-443, Sep. 2019.

[7] T. Nishi, P. Doshi and D. Prokhorov, "Merging in Congested Freeway Traffic Using Multipolicy Deci-sion Making and Passive Actor-Critic Learning," IEEE Transactions on Intelligent Vehicles, vol. 4, no. 2, pp. 287-297, Mar. 2019.

[8] M. D. Vaio, P. Falcone, R. Hult et al., "Design and Experimental Validation of a Distributed Interaction Protocol for Connected Autonomous Vehicles at a Road Intersec-tion," IEEE Transactions on Vehicular Technology, vol. 68, no. 10, pp. 9451-9465, Aug. 2019.

[9] Y. Chen, J. Zha and J. Wang, "An Autonomous T-Intersection Driving Strategy Considering On-coming Vehicles Based on Connected Vehicle Technology," IEEE/ASME Transactions on Mechatronics, vol. 24, no. 6, pp. 2779-2790, Sepr. 2019.

[10] L. Yang , L. Zhang, Z. He etal., "Efficient Hybrid Data Dissemination for Edge-Assisted Automated Driving," IEEE INTERNET OF THINGS JOURNAL, vol. 7, no. 1, pp. 148-159, Oct. 2019.