DOI: 10.23977/autml.2026.070202
ISSN 2516-5003 Vol. 7 Num. 2

A Multi-Agent Reinforcement Learning-Based Collaborative Decision-Making Method for Intelligent Connected Vehicle Clusters in High-Density Traffic Scenarios

Quan Yu[1], Zhuo Bao[1,a,*]

[1]*School of Electrical and Control Engineering, North China University of Technology, Beijing, 100144, China*
[a]*1067919244@qq.com*
**Corresponding author*

Keywords: Intelligent connected vehicles; High-density traffic; Multi-agent reinforcement learning; Collaborative decision-making; MAPPO; Neighborhood interaction

Abstract: To address the limitations of local observations, insufficient modeling of inter-vehicle interactions, and the difficulty of balancing traffic efficiency and operational stability in collaborative decision-making for intelligent connected vehicle clusters under high-density traffic conditions, this paper proposes a collaborative decision-making method based on multi-agent reinforcement learning. First, according to the operational characteristics of vehicle groups in high-density traffic scenarios, the collaborative decision-making requirements are clarified with the objectives of improving road resource utilization, average travel speed, and traffic flow stability. Second, the vehicle-cluster decision-making problem in high-density traffic scenarios is formulated as a decentralized partially observable Markov decision process. An explicit neighborhood interaction mechanism is introduced to characterize the dynamic local relationships among vehicles, and a structured reward function integrating traffic efficiency, operational stability, safety constraints, and rule compliance is designed. Furthermore, under the centralized training and decentralized execution framework, collaborative policy learning for vehicle clusters is achieved based on MAPPO. Finally, a high-density traffic simulation scenario is constructed on the highway-env platform, and comparative experiments are conducted against Rule-based, IPPO, and NI-MAPPO methods. The experimental results show that the proposed method effectively improves the collaborative traffic performance of vehicle clusters under high-density traffic conditions. In the test scenario, the average speed reaches 25.041 m/s, the speed standard deviation is 1.393 m/s, and the collision rate is 0, indicating favorable traffic efficiency, operational stability, and safety.

1. Introduction

With the rapid development of the Internet of Vehicles, autonomous driving, and swarm intelligence technologies, collaborative decision-making for intelligent connected vehicle clusters has gradually become an important research direction for improving road traffic efficiency and the operating quality of transportation systems[1].

High-density traffic scenarios constitute one of the typical problems in the study of collaborative decision-making for intelligent connected vehicles[2]. To address the above issues, this paper proposes a multi-agent reinforcement learning-based collaborative decision-making method for intelligent connected vehicle clusters in high-density traffic scenarios[3].

2. Theoretical Basis

2.1 Behavioral Characteristics of Vehicle Clusters in High-Density Traffic Scenarios

The core characteristics of high-density traffic scenarios lie in limited road resources, a large number of vehicles, and frequent group interactions[4]. Compared with low-density traffic environments, vehicle groups in such scenarios are more susceptible to local congestion, speed disturbances, and lane-load imbalance, thereby leading to reduced traffic efficiency and weakened traffic flow stability[5].

2.2 Fundamentals of Multi-Agent Reinforcement Learning

In a multi-agent system, the interaction relationships among agents are time-varying[6]. The action of a single agent changes the environment state and further affects the observations and decision choices of other agents. Therefore, the overall system behavior is not a simple superposition of independent agent behaviors, but rather emerges from continuous interactions[7].

2.3 MAPPO-Related Methods

MAPPO is a multi-agent proximal policy optimization method that essentially extends PPO to cooperative multi-agent tasks[8]. This method generally adopts the centralized training and decentralized execution mechanism[9]. During training, more comprehensive global information is exploited to improve value estimation quality, whereas during execution each agent independently makes decisions according to its own local observations[10].

3. Model Formulation

3.1 Vehicle Kinematic Model

To describe the basic vehicle motion behavior in high-density traffic scenarios, this paper adopts a simplified vehicle kinematic model[11]. Let the vehicle position at time t be (x_t, y_t), the heading angle be ψ_t, the longitudinal speed be v_t, the longitudinal acceleration be a_t, the front-wheel steering angle be δ_t, the wheelbase be L, and the discrete time step be Δt. Then, the vehicle state update can be expressed as

$$x_{t+1} = x_t + v_t \cos(\psi_t) \Delta t \tag{1}$$

$$y_{t+1} = y_t + v_t \sin(\psi_t) \Delta t \tag{2}$$

$$\psi_{t+1} = \psi_t + \frac{v_t}{L} \tan(\delta_t) \Delta t \tag{3}$$

$$v_{t+1} = v_t + a_t \Delta t \tag{4}$$

3.2 Dec-POMDP Problem Formulation

Considering a collaborative decision-making system composed of N intelligent connected vehicles, the vehicle-cluster decision-making problem in a high-density traffic scenario is formulated as a decentralized partially observable Markov decision process (Dec-POMDP)[12], which is represented as

$$G = \langle N, S, \{A_i\}_{i=1}^{N}, \{O_i\}_{i=1}^{N}, P, R, \Omega, \gamma, \rho_0 \rangle \tag{5}$$

At time t, the joint action of all vehicles is expressed as

$$a_t = \{a_1^t, a_2^t, \cdots, a_N^t\} \tag{6}$$

3.3 Explicit Neighborhood Interaction Modeling

In high-density traffic scenarios, the effectiveness of vehicle decision-making strongly depends on the behavioral variations of surrounding neighboring vehicles[13]. To enhance the perception capability for local traffic structures, this paper constructs an explicit neighborhood interaction mechanism[14]. Let the position vector of vehicle i at time t be p_i^t. Then, its perceived neighborhood can be defined as

$$N_i^p(t) = \{j \mid \| p_j^t - p_i^t \| \leq r_p,\ j \neq i\} \tag{7}$$

Furthermore, considering communication constraints, the interaction neighborhood can be expressed as

$$N_i^c(t) = \{j \mid \| p_j^t - p_i^t \| \leq r_c,\ j \neq i\} \tag{8}$$

3.4 Reward Function Design

To enable vehicle clusters to achieve efficient, safe, and stable traffic operation in high-density scenarios, this paper designs a structured reward function[15]. The overall reward consists of a traffic-efficiency reward, a stability reward, a safety-constraint reward, and a rule-compliance reward, that is

$$r_t = w_1 r_t^{\mathrm{eff}} + w_2 r_t^{\mathrm{sta}} + w_3 r_t^{\mathrm{safe}} + w_4 r_t^{\mathrm{rule}} \tag{9}$$

3.5 MAPPO-Based Collaborative Learning under CTDE

This paper adopts the MAPPO method under the centralized training and decentralized execution framework to achieve multi-agent collaborative policy learning[16]. The policy network outputs a continuous action distribution. The policy distribution of the ith agent at time t is defined as

$$\pi_\theta(a_i^t \mid z_i^t) = \mathcal{N}(\mu_\theta(z_i^t), \sigma_\theta(z_i^t)) \tag{10}$$

3.6 Evaluation Metrics

For the high-density traffic scenario, this paper selects average speed, speed standard deviation, and collision rate as the core evaluation metrics.

The metric calculation formulas are as follows.

(1) Average speed

The formula is as follows

$$\bar{v} = \frac{1}{N}\sum_{i=1}^{N} v_i \tag{11}$$

(2) Speed standard deviation
The formula is as follows

$$\sigma_v = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(v_i - \bar{v})^2} \tag{12}$$

(3) Collision rate
The formula is as follows

$$R_{\text{col}} = \frac{N_{\text{col}}}{N_{\text{all}}} \times 100\% \tag{13}$$

4. Experimental Analysis

4.1 Experimental Scenario and Parameter Settings

The specific configuration of the simulation platform is shown in Table 1.

Table 1: Experimental hardware and software configuration.

Environment setting	Configuration information
CPU model	Intel Core i7-10750H
GPU model	NVIDIA RTX 2060(8GB)
Python version	Python 3.9.18
Operating system	Windows 11
Simulation platform	highway-env

The high-density traffic scenario is mainly used to investigate the collaborative traffic capability of different methods under congested traffic flow conditions. The scene diagram is shown in Figure 1.

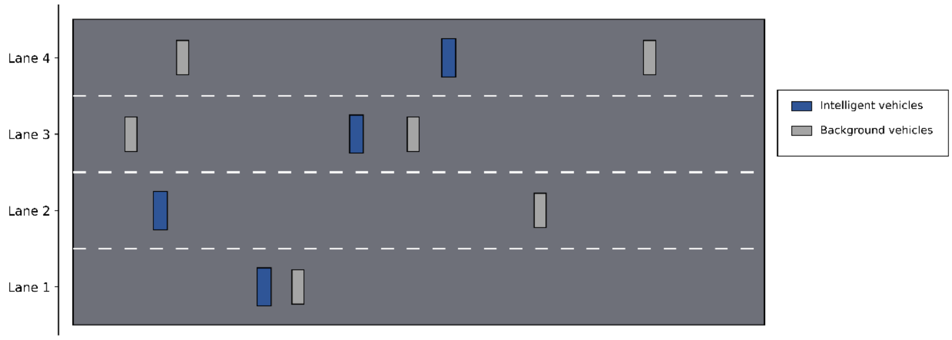

Figure 1: Schematic of the high-density traffic scenario.

4.2 Baseline Models

To verify the effectiveness of the proposed multi-agent collaborative decision-making method in complex traffic scenarios, four methods were selected for comparison: a Rule-based method, an Independent Proximal Policy Optimization method (IPPO), a Non-Interactive Multi-Agent Proximal Policy Optimization method (NI-MAPPO), and the centralized training and decentralized execution multi-agent method with explicit interaction modeling and structured reward design proposed in this paper, namely CTDE-MAPPO .

4.3 Decision Performance Analysis in the High-Density Traffic Scenario

In the high-density traffic scenario, vehicle clusters are required to realize efficient, safe, and stable collaborative operation under the constraint of limited road resources. The graph of average speed changes is shown in Figure 2.

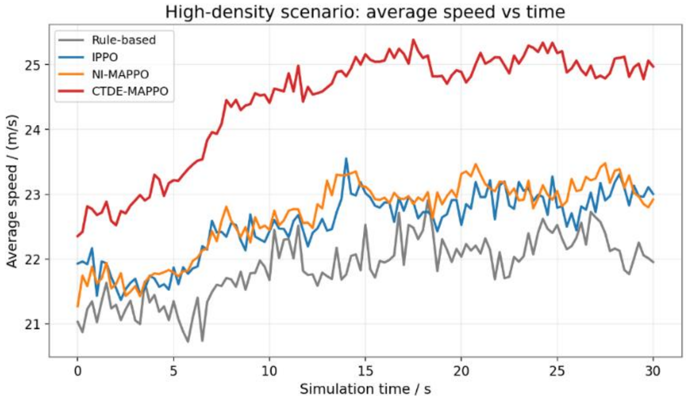

Figure 2: Average-speed-versus-time curves under the high-density traffic scenario.

The results show that the average speeds of different methods in the high-density traffic scenario all undergo an evolution process from initial adjustment to gradual stabilization, but their variation trends differ significantly. The schematic diagrams of the results are shown in Figures 3 to 5.

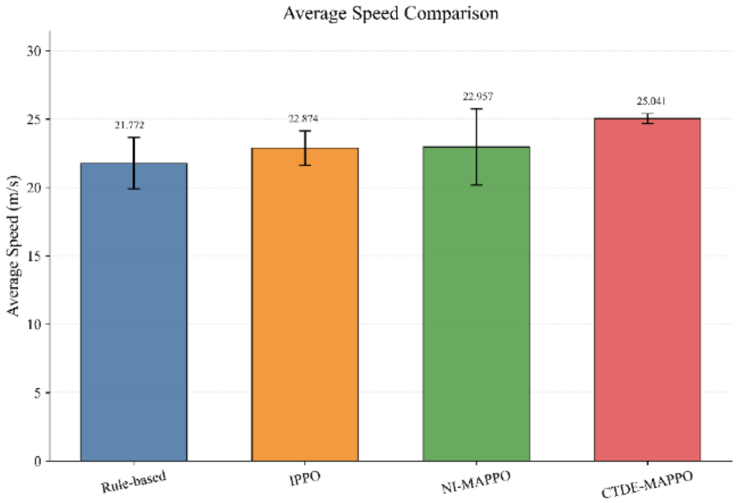

Figure 3: Comparison of average speed in the high-density traffic scenario.

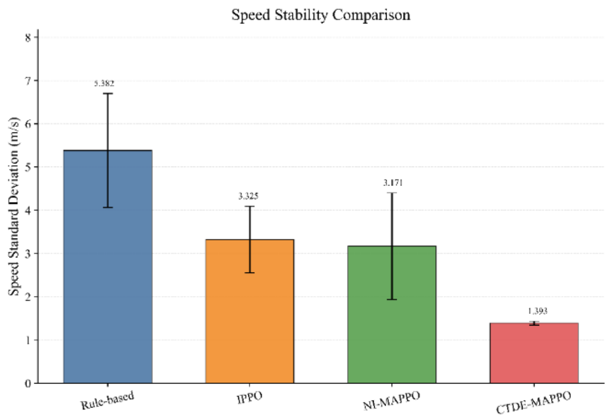

Figure 4: Comparison of speed standard deviation in the high-density traffic scenario.

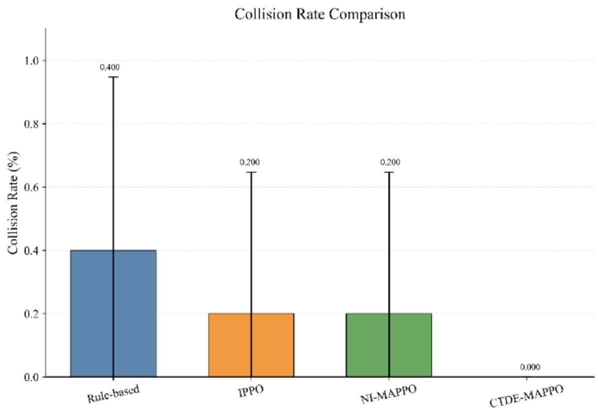

Figure 5: Comparison of collision rate in the high-density traffic scenario.

To further compare the traffic efficiency and operational stability of each method from the perspective of overall results, this paper calculates the average speed, speed standard deviation, and collision rate in the high-density traffic scenario. The quantitative results are shown in Tables 2 and 3.

Table 2: Performance comparison of different methods in the high-density traffic scenario.

Method	Average speed/(m • s-1)	Speed standard deviation/(m • s-1)	Collision rate/%
Rule-based	21.772	5.382	0.400
IPPO	22.874	3.325	0.200
NI-MAPPO	22.957	3.171	0.200
CTDE-MAPPO	25.041	1.393	0

Table 3: Relative performance improvement of the proposed method over other methods.

Method	Change in average speed /%	Reduction in speed standard deviation/%	Reduction in collision rate/%
Rule-based	15.02	74.11	100
IPPO	9.47	58.11	100
NI-MAPPO	8.08	56.07	100

From the current experimental results, the proposed method demonstrates favorable comprehensive performance in the high-density traffic scenario.

5. Conclusion

The experimental results show that the proposed method outperforms Rule-based, IPPO, and non-interactive MAPPO methods in terms of average speed, speed standard deviation, and collision rate, indicating that it can effectively improve the collaborative traffic efficiency and operational stability of intelligent connected vehicle clusters under high-density traffic conditions while satisfying safety constraints. Future research may further extend and validate the method under more complex road structures and practical communication constraints.

References

[1] Tao S, Wen-Xing Z, Shi-Bin S, et al. Distributed sliding mode control strategy for intelligent connected vehicle platoon in complex media[J]. Applied Mathematical Modelling, 2024, 125(PB): 404-424. DOI: 10. 1016/J. APM. 2023. 10. 028.

[2] Fangwu M, Yu Y, Jiawei W, et al. Eco-driving-based cooperative adaptive cruise control of connected vehicles platoon at signalized intersections[J]. Transportation Research Part D, 2021, 92DOI: 10. 1016/J. TRD. 2021. 102746.

[3] Guichuan L, Rongchen Z, Ming S. Cooperative Adaptive Sliding Mode Platoon Control of Intelligent Connected Vehicles under Communication Interruption[J]. International Journal of Automotive Technology, 2023, 24(2): 513-525. DOI: 10. 1007/S12239-023-0043-4.

[4] Wei Z, Dong C, Jun Y, et al. Multi-agent reinforcement learning for cooperative lane changing of connected and autonomous vehicles in mixed traffic[J]. Autonomous Intelligent Systems, 2022, 2(1): DOI: 10. 1007/S43684-022-00023-5.

[5] Li S, "David" W F. Performance of State-Shared Multiagent Deep Reinforcement Learning Controlled Signal Corridor with Platooning-Based CAVs[J]. Journal of Transportation Engineering, Part A: Systems, 2023, 149(8): DOI: 10. 1061/JTEPBS. TEENG-7768.

[6] Haotian S, Yang Z, Keshu W, et al. Connected automated vehicle cooperative control with a deep reinforcement learning approach in a mixed traffic environment[J]. Transportation Research Part C, 2021, 133DOI: 10. 1016/J. TRC. 2021. 103421.

[7] Haotian S, Danjue C, Nan Z, et al. A deep reinforcement learning based distributed control strategy for connected automated vehicles in mixed trafficplatoon[J]. Transportation Research Part C, 2023, 148DOI: 10. 1016/J. TRC. 2023. 104019.

[8] Pu Q, Xie K, Guo H, et al. Modeling interactive crash avoidance behaviors: A multi-agent state-space transformer-enhanced reinforcement learning framework. [J]. Accident; analysis and prevention, 2025, 226108334. DOI: 10. 1016/J. AAP. 2025. 108334.

[9] Jinjue L, Chunhui Y, Zilin S, et al. A survey on urban traffic control under mixed traffic environment with connected automated vehicles[J]. Transportation Research Part C, 2023, 154DOI: 10. 1016/J. TRC. 2023. 104258.

[10] Bellman R. Dynamic programming[M]. Princeton: Princeton University Press, 1957.

[11] Sutton R S, Barto A G. Reinforcement learning: An introduction[M]. 2nd ed. Cambridge, MA: MIT Press, 2018.

[12] Watkins C J C H, Dayan P. Q-learning[J]. Machine Learning, 1992, 8(3/4): 279-292.

[13] Williams R J. Simple statistical gradient-following algorithms for connectionist reinforcement learning[J]. Machine Learning, 1992, 8(3/4): 229-256.

[14] Konda V R, Tsitsiklis J N. Actor-critic algorithms[C]//Advances in Neural Information Processing Systems 13. Cambridge, MA: MIT Press, 2000.

[15] Silver D, Lever G, Heess N, et al. Deterministic policy gradient algorithms[C]//Proceedings of the 31st International Conference on Machine Learning. Beijing: JMLR. org, 2014: 387-395.

[16] Littman M L. Markov games as a framework for multi-agent reinforcement learning[C]//Proceedings of the 11th International Conference on Machine Learning. San Francisco: Morgan Kaufmann, 1994: 157-163.