An Intelligent Lane Changing Decision Method for Connected Vehicles

Zheyu Cui
Department of Automation, Tsinghua University, Beijing, China
czyfaiz@foxmail.com

Jianming Hu*
Department of Automation, Tsinghua University, Beijing, China
hujm@mail.tsinghua.edu.cn

ABSTRACT

Aiming at the problem of driving scenarios redundancy in the lane change decision-making, this paper proposes a deep reinforcement learning method (DRL) for lane change decision based on embedded attention mechanism (CADQN). The algorithm introduces the Convolutional Attention Mechanism Module (CBAM) into the DQN network to optimize the scenarios in time and space dimensions, and assist connected vehicles in making lane changing decisions. The algorithm is verified by the traffic simulation platform under the highway environment, and the results show that CADQN is helpful to improve the global traffic efficiency, and with the increase of traffic flow density, the benefit is more significant. Moreover, the visualization results of the attention layer in the CADQN can guide the optimization of the driving scenario.

CCS CONCEPTS

• **Computing methodologies**; • **Artificial intelligence**; • **Control method**; • **Motion path planning**;

KEYWORDS

Lane change decision-making, Scenarios optimization, Deep reinforcement learning, Attention mechanism

ACM Reference Format:

Zheyu Cui and Jianming Hu. 2021. An Intelligent Lane Changing Decision Method for Connected Vehicles. In *The 5th International Conference on Computer Science and Application Engineering (CSAE 2021), October 19–21, 2021, Sanya, China.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3487075.3487175

1 INTRODUCTION

The two main microscopic behaviors of a vehicle during driving are car following behavior and lane changing behavior [1]. Lane changing behavior is the most influential behavior on traffic flow. It is defined as: due to the influence of the surrounding vehicle driving state or road condition information, the driver has the intention of lane changing and changes the vehicle from the original lane to the target lane. More macro driving behaviors, such as overtaking the preceding vehicle and vehicle confluence, can be decomposed into the most microscopic car-following behavior and lane-changing behavior for research. Literature [2-5] pointed out that the main cause of traffic jams is caused by unreasonable or too frequent lane changing operations. The reason is that when the vehicle enters the target lane from the existing lane, such a lane change action will make the vehicle a moving obstacle, resulting in reduced highway capacity and safety [5] [6]. In addition, the number of traffic accidents caused by inappropriate lane changes is about one-tenth of the total number of accidents [7]. Therefore, studying the method of lane change decision has far-reaching significance to the improvement of traffic efficiency.

Lane changing decision model means that the driver decides whether to produce lane changing intentions based on the expected driving state (such as expected speed, acceleration, etc.) and the state of surrounding traffic elements (such as the state of the car following the car, the state of obstacles, road information, etc.) model. The driver's intention to change lanes is roughly divided into two categories: mandatory lane change and arbitrary lane change. Compulsory lane change refers to the driver's intention to change lanes when there are obstacles in front of the vehicle, roads converge, and traffic rules restrict access to the target lane.

With the gradual development of real-scene traffic data collection technology, since the early 1980s, vehicle lane changing decision models have received more and more attention [8]. The lane change model has a wide range of applications. Its main applications can be roughly divided into two categories: adaptive cruise control and computer simulation. The lane change model in adaptive cruise control mainly focuses on the development of assisted driving models, which can be further divided into collision avoidance models and autonomous driving models. The collision avoidance model is used to control the driver's lane changing operation and help the driver complete the lane changing operation safely. The automation model is used to automatically adjust the steering wheel angle of the vehicle to perform safe lane change operations [9-16]. For computer simulation applications, it is mainly to reproduce the driver's driving decision on the computer in order to restore the real driving scenario. These lane changing models are mainly divided into three categories: 1. Rule-based models, for example: Gibbs [17] [18] proposed in 1986 the lane changing decision model on highways and urban streets; 2. Discrete choice Models, for example: Ahmed [19] [20] introduced the discrete choice model to the lane changing decision problem and proposed a dynamic hierarchical discrete choice model; 3. Artificial intelligence model, for example: Hunt [21] used a neural network model on two lanes to predict the driver's lane change decision. In 2009, Dumbuya [22] and others designed a neuro-driving agent to model lane changing behavior. The network input is the current direction of the vehicle, the current speed, the distance to the vehicle, the preferred speed and the

CSAE 2021, October 19–21, 2021, Sanya, China

ACM ISBN 978-1-4503-8985-3/21/10…$15.00
https://doi.org/10.1145/3487075.3487175

current lane, and the output is the new direction and speed. The training of the model is completed through the back propagation algorithm. In 2016, Xin LI [23] introduced the Q-learning algorithm of reinforcement learning to the overtaking decision problem of unmanned vehicles under highways.

Deep reinforcement learning (DRL) is the fusion of the neural network model and reinforcement learning model in machine learning. It uses a neural network model to replace the traditional tabular function to fit the state-action value function, which greatly enriches the ability of the reinforcement learning agent to explore the optimal solution in the infinite-dimensional state space. It is a useful tool for studying and exploring the lane changing decision problem with infinite dimensional state space [24]. The Connected vehicle can provide self-driving vehicle decision makers with a relatively complete perception of environmental conditions through sensors and computer communications, and use deep reinforcement learning models to replace human drivers to explore the global optimal lane changing strategy.

For the connected vehicle, the perceived driving environment is provided by sensors, computer communication and other means. These data will constitute a complex and huge but redundant driving environment state, which contains many redundant traffic elements, such as convective vehicles and traffic light information of the last intersection. The complex and redundant driving environment makes it more difficult for the agent to find the optimal solution of the decision. Therefore, this study introduces attention mechanism into the state space of deep reinforcement learning to complete the optimization of complex scenario redundancy, so as to better assist the vehicle driving decision.

In conclusion, this study focuses on the lane change decision of highway scenario. The scenario optimization model based on DRL embeded attention mechanism is introduced to assist the vehicle to make better driving decisions by reducing the redundancy of the scenario. Finally, the performance of the algorithm is verified on the simulation platform, and the visualization of scenario optimization is realized. Strive to explore more safe, efficient, energy-saving vehicle lane change driving decision-making.

2 MODELING OF LANE CHANGE DECISION

Before explaining the deep reinforcement learning decision-making algorithm based on the embedded attention mechanism, it is necessary to mathematically model the lane-changing decision problem and turn it into a Markov decision process for reinforcement learning.

2.1 Markov Decision Process

Markov decision process is a mathematical description of decision type. The core idea is memory-lessness, that is, the current decision depends only on the current state, and has nothing to do with the historical state and decision-making. The Markov decision-making process uses five-tuples E $=< S, A, P, R, \pi >$to describe. S is the state space of the decision-making process task; A is the action space of the decision-making process, P is the transition probability, which is defined as$P(S \times A- > S')$. That is, the probability of the current state S transitioning to S'under action A. For the decision-making process strategy π, it determines the next action A in the state S. R is the reward. The mathematical description of the memory-lessness of Markov's decision-making process is shown in equation (1).

$$P(S_{t+1} = s'|S_t = s_t, A_t = a_t, S_{t-1} = s_{t-1}, A_{t-1} = a_{t-1}, \dots S_0 = s_0) \\ = P(S_{t+1} = s'|S_t = s_t, A_t = a_t) \tag{1}$$

$$G_\pi = E_\pi[\sum_{t=0}^{T} \gamma_t R_t] \tag{2}$$

The goal of reinforcement learning is to maximize the expected value G of cumulative rewards by learning the optimal strategy π^*, where γ_t is the attenuation factor. As shown in equation (2).

The lane changing decision problem can also be modeled as a Markov decision problem. The vehicle makes a real-time decision A according to the driving environment state S, and the environment accepts the action A to transfer the state to S' and feeds back the reward value R at the same time. The decision maker continuously adjusts its own strategy π according to the reward value R to obtain the optimal strategy. Next, the S, A, and R defined under the lane change decision will be introduced.

2.2 State Space

When reinforcement learning is applied to vehicle lane changing decision-making, there are usually three methods for vehicle environment modeling. 1. Vehicle dynamics parameter vector representation; 2. Road network grid modeling; 3. Original image data of simulation. The first two state spaces filter data according to human driving behavior, not the original environmental information. In order to study the effect of attention mechanism on scenario optimization, we use the third state space representation: the most original frame of the simulation program. Because the simulation platform is a color image, in order to reduce the overall calculation of the model, and the decision-making problem is not strongly dependent on color, we do gray processing on the original image. The weighted average method is used in gray processing, and its mathematical expression is shown in equation 3.

$$I(x, y) = W_r \times I_r(x, y) + W_g \times I_g(x, y) + W_b \times I_b(x, y) \tag{3}$$

Irepresents the pixel value after grayscale processing. W_r,W_g,W_b represent the conversion weights of red, green and blue. I_r,I_g,$I_b$represent the pixel values of red, green and blue. In the decision-making process, the speed information of vehicles in the environment plays an important role in the decision-making process, so we use continuous frames to represent the speed information of vehicles in this period. Finally, we use continuous gray frames as the state representation of lane change decision problem. Figure 1 shows the states of three consecutive moments at the beginning of simulation, in which each state contains information of four consecutive moments.

2.3 Action Space

Actions represent decisions made by the vehicle based on the current state. In order to compress the search space of the entire problem, we designed a discrete action space. In the final action space, we divide lane-changing behavior into longitudinal and lateral dimensions for research.

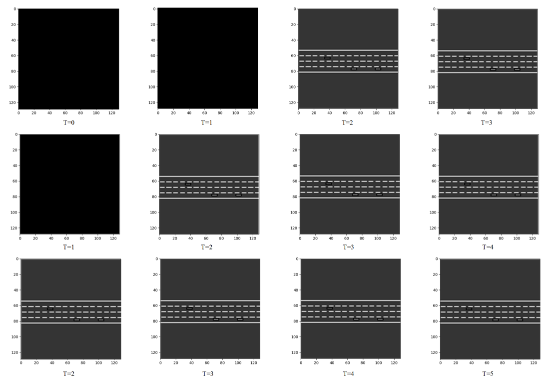

Figure 1: State Space Diagram.

In the longitudinal direction, the control of vehicle mainly adopts adaptive cruise control. Its main three actions are: acceleration, uniform speed and deceleration.

In the lateral direction, when vehicles change lanes, they mainly take three actions: turn left to change lanes, keep lanes and change lanes right. Considering the longitudinal and lateral dimensions, the final action space is defined as five discrete actions: 1. No action; 2. Acceleration; 3. Deceleration; 4. Left lane change; 5. Right lane change.

2.4 Reward Function

The reward function represents the score of an action in the current environment. Since the objective function of reinforcement learning is to maximize the cumulative reward, that is, equation 2, the design of reward function determines the strategy learned by the agent. In lane changing decision-making, the reward function represents the score of a certain lane changing behavior, and the value of the reward function is expected to exist both positive and negative, which is convenient for the training of neural network. The design of reward function determines the final training result.

For connected vehicles, they are connected with each other and drive together as a whole. Compared with single agent intelligence, the design of reward function not only considers the influence of the planned vehicle itself, but also considers the influence on the overall transportation system. So the reward function is divided into two parts. Part of it comes from the improvement of system performance by lane changing, and part of it comes from the improvement of vehicle itself by lane changing. The form of reward function is shown in equation 4.

$$R_{total} = w_1 R_{system} + w_2 R_{individual} \tag{4}$$

R_{total}represents the total reward, R_{system}represents the system reward, $R_{individual}$represents the individual reward, $w_1$and $w_2$represent the weight coefficient corresponding to the reward.

For the system reward, this study considers the impact on the original lane and the target lane at the same time. For the original lane and the target lane, this paper selects the traffic density and the average speed of road vehicles to represent the traffic efficiency of the lane. Obviously, the smaller the traffic density is, the higher the average speed of road vehicles is, the higher the traffic efficiency of the system is. The overall expression of system reward is shown in equation 5.

$$\begin{aligned} R_{system} = {} & w_{11}(Density_{bf}^{t\arg et} - Density_{af}^{t\arg et}) \\ & + w_{12}(Density_{bf}^{original} - Density_{af}^{original}) \\ & + w_{13}(\bar{v}_{af}^{t\arg et} - \bar{v}_{bf}^{t\arg et}) \\ & + w_{14}(\bar{v}_{af}^{original} - \bar{v}_{bf}^{original}). \end{aligned} \tag{5}$$

For the reward of individual vehicle, this paper considers the reward of speed, collision penalty and lane change penalty. It has its maximum speed for different lanes. However, vehicles in the free driving state tend to drive at the maximum speed of the lane. Therefore, the maximum speed of the lane is set to the desired speed. The difference between the lane changing speed and the expected speed is used as a reward and normalized. Collision penalty is used to punish collision behavior during lane changing. Due to the fluctuation of traffic flow caused by lane changing, it is easy to produce negative effects, so generally, lane changing behavior is not encouraged, so lane changing penalty is introduced. In conclusion, the individual reward function is equation 6.

$$R_{individual} = \alpha\left(\boldsymbol{v}_{des} - \boldsymbol{v}\right)/(\boldsymbol{v}_{des} - \boldsymbol{v}_{\min}) + \beta R_{col} + \gamma R_{lc} \tag{6}$$

α,β,$\gamma$are the corresponding weight coefficients, R_{col} is the collision penalty, and R_{lc} is the lane change penalty.

$$R_{col} = \begin{cases} -R_{coll}\text{ifcollisionhappen} \\ 0\text{otherwise} \end{cases} \tag{7}$$

$$R_{lc} = \begin{cases} -R_{lanechange}\text{ifvehiclechangeslane} \\ 0\text{otherwise} \end{cases} \tag{8}$$

3 DEEP REINFORCEMENT LEARNING ALGORITHM EMBEDDED ATTENTION MECHANISM

3.1 DQN

DQN is a powerful tool which combines traditional Q-learning and deep learning[25]. Its main purpose is to solve the bottleneck problem of using Q table to replace Q function in Q-learning algorithm. In the original Q-learning algorithm, we need to maintain a two-dimensional Q table at all times. The dimensions of the two-dimensional Q table represent states and actions. But in reality, the state space can't be counted, and the simple Q table can't express it at all. So neural network is introduced to fit Q function. The DQN network framework is shown in Figure 2, in which the q-eval network adopts a multi-layer perceptron structure.

The whole network is composed of two networks with different parameters but the same network structure, namely Q_EVAL network and Q_TARGET network. Q_EVAL and Q_TARGET share delay parameters. The Q_EVAL update formula of traditional Q-learning algorithm is equation 9.

$$Q(s,a) \leftarrow Q(s,a) + \alpha[R + \gamma \max_{a'} Q(s',a') - Q(s,a)] \tag{9}$$

Equation 9 can be regarded as an updated formula of gradient descent. The loss function of DQN network is obtained by integrating the right gradient term

$$L(\theta) = (R + \gamma \max_{a'} Q(s',a'|\theta^-) - Q(s,a|\theta))^2 \tag{10}$$

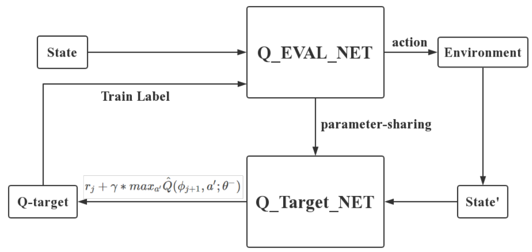

Figure 2: DQN Network.

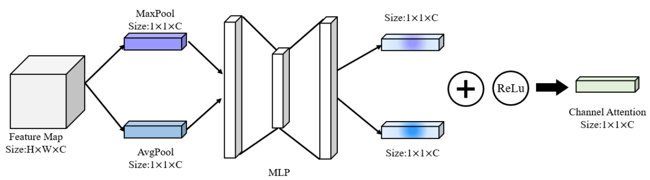

Figure 3: Channel Attention.

γ is the discount factor, θ^- is the parameter of Q_TARGET network, θ is the parameter of Q_EVAL network.

3.2 Convolutional Block Attention Module

CBAM proposed by s woo [26] is used for attention tagging in image classification tasks. It mainly combines channel attention mechanism and spatial attention mechanism. Because the state space uses four consecutive gray images. Therefore, this study introduces CBAM to allocate the attention of the input image. It aims at optimizing scenario in time and space.

Channel attention is to allocate attention on the channel for the feature map. It means to pay attention to what kind of characteristics are meaningful. In order to compress the information content in a channel, the channel attention module adopts global average pooling and global maximum pooling to compress the information, as shown in Figure 3.

The size of the input feature map f is H×W×C. After global maximum pooling and global average pooling, the compressed vector with size of 1×1×C is obtained. The two vectors pass through the multi-layer perceptron to get the 1×1×C feature vector. Finally, two different feature vectors are added and the Relu activation function is used to get the channel attention vector with the final size of 1×1×C. Its mathematical expression is equation 11

$$A_c(F) = \delta(MLP(AvgPool(F) + MLP(MaxPool(F))) \tag{11}$$

Spatial attention mechanism is to distribute attention in space, which means to pay attention to which points on a feature map are meaningful. Similar to channel attention mechanism, spatial attention mechanism adopts maximum pooling and average pooling to compress information in channel dimension. As shown in Figure 4. The mathematical expression is 12

$$A_s(F) = \delta(f^{7\times7}([AvgPool(F); MaxPool(F)])) \tag{12}$$

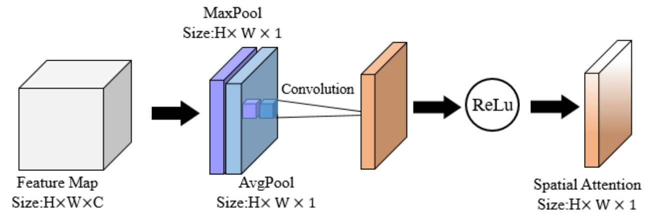

Figure 4: Spatial Attention.

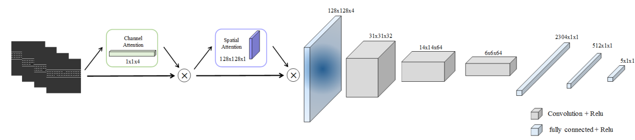

Figure 5: CADQN Structure.

3.3 CADQN

Based on the above-mentioned DQN network framework and CBAM, this paper proposes a CADQN (Convolution Attention Deep Q-network) network structure for lane change decision model. The framework of CADQN model and DQN model is the same, which is divided into Q_EVAL network and Q_TARGET network. The delay parameters are shared between Q_EVAL network and Q_TARGET network, and the network structure is the same. Therefore, only one network structure is shown. The network structure of Q_EVAL is shown in Figure 5

Four consecutive frames of gray-scale image F as the input, its size is 128×128×4, through the channel attention mechanism module, we get the channel attention vector of 1×1×4 and multiply it to get F'. The feature map F'' is obtained by the product of the spatial attention and F'.The mathematical expression process is equation 13,14.

$$F' = A_c(F) \otimes F \tag{13}$$

$$F'' = A_s(F') \otimes F' \tag{14}$$

F'' as the input, through the module composed of three-layer convolution layer and Relu function, the final feature map of 6×6×64 is obtained. The design of convolution module is shown in Table 1. Finally, the obtained feature map is expanded into one-dimensional vector and the action Q value of 5×1×1 is obtained through three full connected layers.

4 SIMULATION

4.1 Simulation Platform

In the part of algorithm verification, we use the traffic simulation platform developed by Leurent [27]. The highway simulation environment is selected to verify the algorithm. After parameter adjustment, the whole highway environment is set as a 4-lane highway, as shown in Figure 6.

The whole traffic simulation platform relies on Python's Gym library. In terms of vehicle control, the bottom control model of the vehicle adopts the kinematic bicycle model for simulation; in terms of lateral control of the vehicle, Leurent designs the steering

Table 1: CADQN Convolution Module

Input Shape	Kernel Num	KernelSize	Stride	Output shape
128×128×4	32	8×8	4	31×31×32
31×31×32	64	4×4	2	14×14×64
14×14×64	64	3×3	1	6×6×64

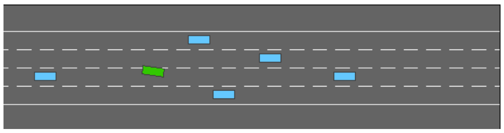

Figure 6: Simulation Platform.

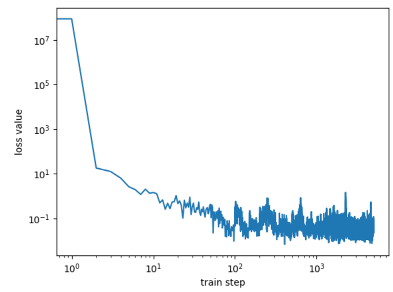

Figure 7: Loss Value in Training.

controller to track the designed target route; in terms of longitudinal control, it adopts the intelligent driving model proposed by Treiber et al. [28] in 2000. The top-level lane change decision is replaced by the proposed CADQN model to verify the performance of the algorithm.

4.2 Experimental Parameter Setting

This section mainly displays the operation parameters of the simulation platform and the CADQN network training parameters, as shown in Table 2 and table 3.

4.3 Algorithm Performance Comparison

In order to verify the performance of the proposed decision network CADQN, we set up a comparative experiment with DQN. Observe the change value of reward function with the increase of training rounds under the same conditions. The training epoch occurred during the simulation.Figure 7 shows the convergence curve of the loss function of CADQN. The number of training rounds can converge when the order of magnitude is 10^3.

At the same time, set the density of traffic flow under CADQN and DQN decision network to 1200 vehicle / h. with the progress of network training, the change curve of reward function is obtained, as shown in Figure 8.

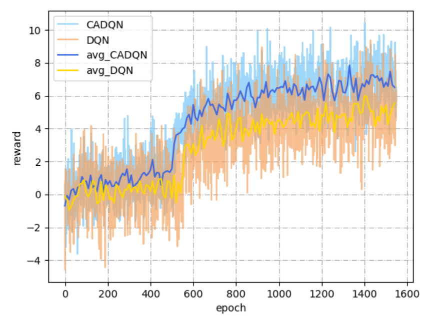

Figure 8: Reward Value in Training.

Both CADQN and DQN keep the greedy strategy. As a result, the fluctuation range of reward value is very large, and the value range covers each other. For the convenience of observation, we calculate the average value of the curve of CADQN and DQN every 10 steps to get the curve of avgCADQN and avgDQN in the Figure 8. Analysis of Figure 8 shows that: in the early stage of training, the reward values of the two networks are in a state of shock, and there is little difference; in the late stage of training, CADQN network can obtain higher reward values than DQN network. It can be

Table 2: Simulation Parameters

Parameter	Value
Simulation Time	10000epoch
Frame Size	128×128
Decision Frequency	1Hz
Conversion Weights	[0.2989,0.5870,0.1140]
Simulation Frequency	15Hz
Lane	4
R_{coll}	10
$R_{lanechange}$	1

Table 3: CADQN Parameters

Parameter	Value
Learning Rate	0.005
Greedy Factor	0.9
Discount Factor	0.9
Replacement episode	50
Replay Memory Size	2000
Batch Size	32

Table 4: Travel Time Comparison

Traffic Density	DQN	CADQN	Improvement
600vehicles/h	10.47s	10.29s	1.74%
800 vehicles /h	11.58s	10.96s	5.66%
1000 vehicles /h	12.03s	11.32s	6.27%
1200 vehicles /h	12.85s	11.73s	9.54%
1400 vehicles /h	13.88s	12.59s	10.24%
1600 vehicles /h	14.91s	13.64s	9.31%
1800 vehicles /h	16.11s	14.77s	9.07%

seen that the introduction of attention mechanism in the scenario to allocate attention to traffic elements can indeed improve the decision-making effect of vehicles compared with the same weight of each traffic element.

In order to study the influence of different traffic density on decision-making effect, we set up comparative experiments of CADQN network and DQN network under different density. The average travel time after network training is used as a comparison. The experimental results are shown in Table 4.

According to the analysis of the experimental results in Table 4, the performance of CADQN is better than that of DQN under different traffic density. With the increase of traffic density, the effect is gradually obvious, the increase percentage is not big at low density, and the increase effect reaches the maximum at 1400 vehicles/h. The value is between medium density and high density. The reason is that with the increase of vehicle density, the traffic participants in the scenario gradually increase. If attention mechanism is not introduced, the attention of vehicles will be distracted by more and more vehicles, that is, the weight of all vehicles is the same. But for lane changing decision, it is obvious that the influence of vehicles in the environment is different. With the increase of traffic volume, this disadvantage is gradually enlarged, so the effect of CADQN is getting better and better.

4.4 Attention Visualization

In order to explore the practical significance of attention, we extract the attention layer of a scenario in the simulation process. The visualization is realized by using heat map. Observe the results of scenario optimization. The visualization example is shown in Figure 9. From the heat map in Figure 9, we can see that the attention mechanism not only optimizes the scenario redundancy, but also strengthens the blind spots in the perception field that we don't pay attention to in our usual driving habits.

In Figure 9, the mapping relationship between the value of attention weight and color is on the far right. The darker the color, the more important the scenario elements are. The running direction of the vehicle is from left to right. Therefore, the attention weight of the rightmost border (that is, the destination of the vehicle) is very high. In the original scenario, the dark gray vehicle in the upper left corner is the control vehicle. Because the driving rule of the simulation software is to keep left, the fast lane is located in the right lane of the road. In this case, there are three driving strategies (straight ahead, left turn and right turn). However, left turn will not get any positive benefits, but will also be punished because of the penalty factor of frequent lane changing behavior. Therefore, there is no attention point on the left road, which is redundant for the scenario. In order to compare the benefits of straight ahead decision and right turn decision, the attention network predicts the driving trajectory of control vehicles in the original lane and the right lane after two frames, and allocates high attention, which represents the front and rear of the vehicle at a certain time in the form of two points. The prediction results are shown in the red box in Figure 10.

Trajectory prediction in scenario optimization is based on the results of optimization in time dimension. This is incomparable to the rule-based scenario optimization. In the spatial dimension, the redundancy of the leftmost lane is optimized. The attention mechanism also allocates the attention of the outermost lane line corresponding to the obstacle vehicle to 0, as shown in the red box of Figure 11

The understanding is also very intuitive, because the existence of obstacle vehicles has played a role similar to the lane line, and it is impossible for autonomous control vehicles to cross the obstacle vehicles and drive out of the lane, so this part is redundant for the scenario. There shouldn't be too much attention.

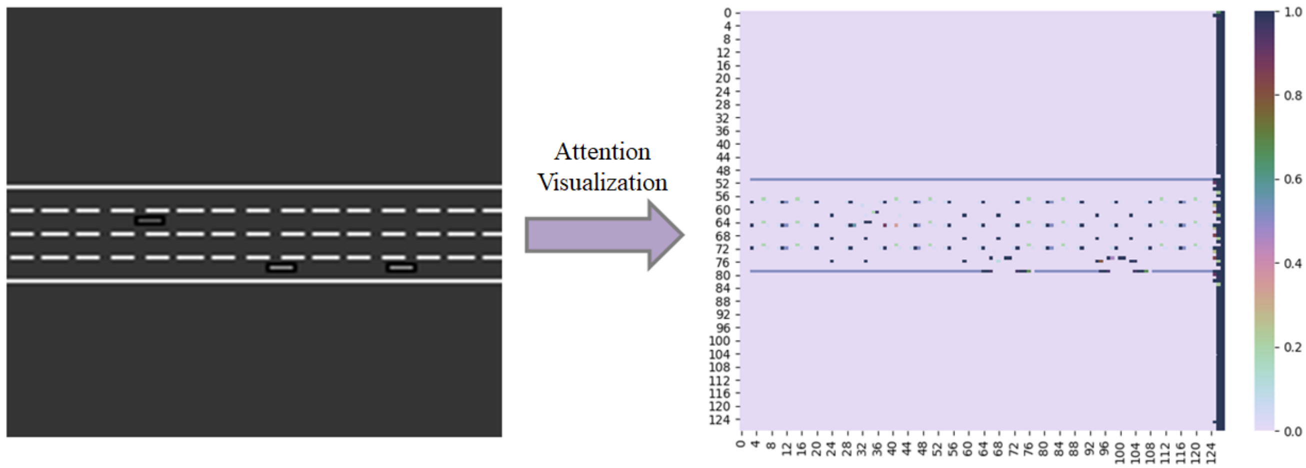

Figure 9: Attention Visualization.

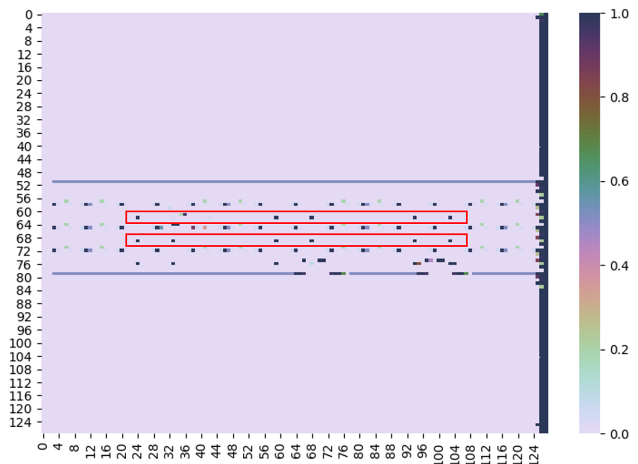

Figure 10: Trajectory Prediction in scenario.

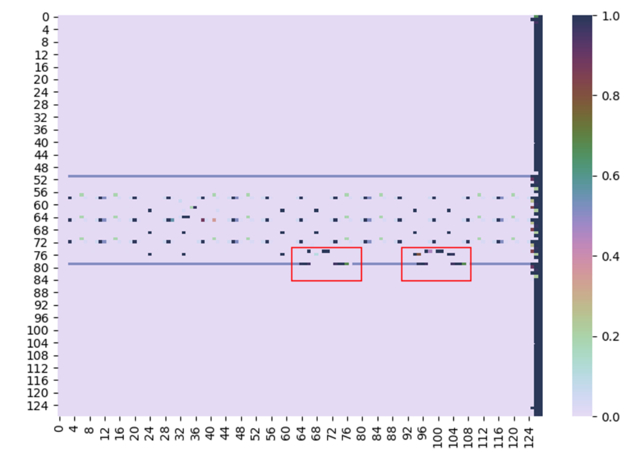

Figure 11: Lane Line Optimization In Scenario.

The above scenario is just a driving scenario of vehicles in the highway environment. According to the experimental results, we can see that the deep reinforcement learning network embedded with attention mechanism can not only optimize the redundancy of time and space. At the same time, high attention will be allocated to the blind spots that we usually don't pay attention to but play a role in the overall traffic efficiency, so as to improve the decision-making effect.

5 CONCLUSION

This paper proposes a deep reinforcement learning lane changing decision method based on embedded attention mechanism, namely CADQN network. In this algorithm, attention mechanism is introduced into lane changing decision-maker to realize the attention weight allocation of scenario traffic elements. So as to complete the scenario optimization and improve the decision-making effect. Simulation results show that compared with the deep reinforcement learning decision-maker without attention mechanism, this algorithm can improve the reward of decision making. Furthermore, in the comparative experimental study under different traffic density, it can be seen that CADQN can improve the overall traffic efficiency of the whole traffic system, and the improvement effect is more significant with the increase of traffic density. Finally, through the visualization of the attention matrix, we find the blind spots that human driving habits can not pay attention to, but the scenario components that play a role in the overall traffic efficiency. The experimental content of the algorithm can provide guidance for scenario optimization before lane change decision.

ACKNOWLEDGMENTS

This work is supported by National Key R&D Program in China (2018YFB1600802).

REFERENCES

[1] L Li, R Jiang, B Jia and X M Zhao, 2010. Modern Traffic Flow Theory and Applications, Vol. I, Highway Traffic. Tsinghua University Press, Beijing, China. ISBN 978-7-302-23807-2 (in Chinese).

[2] Cassidy, M. J., & Bertini, R. L. (1999). Some traffic features at freeway bottlenecks. *Transportation Research Part B: Methodological*, *33*(1), 25-42.

[3] R. L. Bertini and M. T. Leal (2005). Empirical study of traffic features at a freeway lane drop. *Journal of Transportation Engineering*, *131*(6), 397-407.

[4] Y. Hou, P. Edara and C. Sun (2013). Modeling mandatory lane changing using Bayes classifier and decision trees. *IEEE Transactions on Intelligent Transportation Systems*, *15*(2), 647-655.

[5] J. A. Laval and C. F. Daganzo (2006). Lane-changing in traffic streams. *Transportation Research Part B: Methodological*, *40*(3), 251-264.

[6] J. C Munoz and C. F. Daganzo (2002). *Moving bottlenecks: a theory grounded on experimental observation*. Emerald Group Publishing Limited.

[7] N. H. Gartner, P. J. Tarnoff, and C. M. Andrews (1991). Evaluation of optimized policies for adaptive control strategy. *Transportation Research Record*, (1324).

[8] M. Brackstone„ M. McDonald and J. Wu (1998). Lane changing on the motorway: Factors affecting its occurrence, and their implications.

[9] J. Lygeros, D. N. Godbole and S. Sastry (1998). Verified hybrid controllers for automated vehicles. *IEEE transactions on automatic control*, *43*(4), 522-539.

[10] K. Nagel, D. E. Wolf, P. Wagner and P Simon (1998). Two-lane traffic rules for cellular automata: A systematic approach. *Physical Review E*, *58*(2), 1425.

[11] S. Maerivoet and B. De Moor (2005). Cellular automata models of road traffic. *Physics reports*, *419*(1), 1-64.

[12] A. Eidehall, J. Pohl, F. Gustafsson and J. Ekmark (2007). Toward autonomous collision avoidance by steering. *IEEE Transactions on Intelligent Transportation Systems*, *8*(1), 84-94.

[13] D. D. Salvucci, H. M. Mandalia, N. Kuge, and T. Yamamura. (2007). Lane-change detection using a computational driver model. *Human factors*, *49*(3), 532-542.

[14] A. Doshi, and M. Trivedi (2008). A comparative exploration of eye gaze and head motion cues for lane change intent prediction. In *2008 IEEE Intelligent Vehicles Symposium* (pp. 49-54). IEEE.

[15] R. J.Kiefer and J. M. Hankey (2008). Lane change behavior with a side blind zone alert system. *Accident Analysis & Prevention*, *40*(2), 683-690.

[16] L S Jin, W P Fang, Y N Zhang, S B Yang and H J Hou (2009). Research on safety lane change model of driver assistant system on highway. In *2009 IEEE Intelligent Vehicles Symposium* (pp. 1051-1056). IEEE.

[17] P. G. Gipps (1986). A model for the structure of lane-changing decisions. *Transportation Research Part B: Methodological*, *20*(5), 403-414.

[18] P. G. Gipps (1981). A behavioural car-following model for computer simulation. *Transportation Research Part B: Methodological*, *15*(2), 105-111.

[19] K. Ahmed, B. M. Akiva, H. Koutsopoulos, and R. Mishalani (1996). Models of freeway lane changing and gap acceptance behavior. Transportation and traffic theory, 13, 501-515.

[20] K. I. Ahmed, (1999). Modeling drivers' acceleration and lane changing behavior (Doctoral dissertation, Massachusetts Institute of Technology).

[21] J. G. Hunt and G. D. Lyons (1994). Modelling dual carriageway lane changing using neural networks. *Transportation Research Part C: Emerging Technologies*, *2*(4), 231-245.

[22] A. Dumbuya, A. Booth, N. Reed, A. Kirkham, T. Philpott, J. Zhao and R. Wood (2009). Complexity of traffic interactions: Improving behavioural intelligence in driving simulation scenarios. In *Complex systems and self-organization modelling* (pp. 201-209), Springer, Berlin, Heidelberg.

[23] X. Li, X. Xu and L. Zuo (2015). Reinforcement learning based overtaking decision-making for highway autonomous driving. In *2015 Sixth International Conference on Intelligent Control and Information Processing (ICICIP)* (pp. 336-342). IEEE.

[24] Y. Li, (2017). Deep reinforcement learning: An overview. *arXiv preprint arXiv: 1701.07274*.

[25] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, ... and D. Hassabis (2015). Human-level control through deep reinforcement learning. *nature*, *518*(7540), 529-533.

[26] S. Woo, J. Park, J. Y. Lee and I. S. Kweon (2018). Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 3-19).

[27] E. Leurent (2018) An environment for autonomous driving decision-making. https://github.com/eleurent/highway-env.

[28] M. Treiber, A. Hennecke and D. Helbing (2000). Congested traffic states in empirical observations and microscopic simulations. *Physical review E*, *62*(2), 1805.