ISSN 2616-5775 Vol. 6, Issue 1: 19-25, DOI: 10.25236/AJCIS.2023.060104

Research on Automatic Lane Changing Method for Electric Vehicles Based on Deep Deterministic Policy Gradient Algorithm

Yu Chen[a], Juntao Li[b,*]

School of Information, Beijing Wuzi University, Beijing, 101149, China
[a]51979145@qq.com, [b]ljtletter@126.com
[*]Corresponding author

Abstract: *Intelligent driving is an important feature of electric vehicles, and automatic lane-changing is an important auxiliary operation in the driving process. It involves complex data such as the external environment, vehicle status, and relationship characteristics of other vehicles. This paper proposes research on automatic lane-changing methods for electric vehicles based on a deep deterministic policy gradient algorithm aiming at the complexity of automatic lane-changing for electric vehicles. First, use the actor-critic model of deep reinforcement learning to realize the design of an automatic lane-changing algorithm. On this basis, the actor-critic model is further improved, and a deep deterministic policy gradient algorithm is proposed and applied to the automatic lane-changing strategy of electric vehicles, which improves the accuracy of automatic lane-changing and thus ensures the safety of vehicle driving.*

Keywords: *Intelligent driving; Automatic lane-changing; Deep learning; Actor-critic model; DDPG (Deep Deterministic Policy Gradient)*

1. Introduction

The automatic lane-changing technology of electric vehicles is an important part of intelligent driving technology, which is an auxiliary operation for intelligent driving. Through the sensors, 5G network communication equipment and cloud data platform carried by the electric vehicle, the vehicle can efficiently perceive the surrounding environment, vehicle position and distance information between other vehicles during driving. Based on the obtained data, better lane change and path planning can be performed to avoid a series of traffic accidents caused by the driver's personal factors and protect the safety of vehicles.

This paper studies the automatic lane-changing strategy of electric vehicles and realizes the autonomous lane-changing of vehicles through the deep learning model. It can be applied to the intelligent driving of electric vehicles and, at the same time, also to the theoretical research of the unmanned driving stage, alleviating the traffic pressure and providing a major guarantee for people's safe travel.

2. Overview of Intelligent Driving System for Electric Vehicles

In recent years, with the gradual popularization of electric vehicles, compared with traditional vehicles, in addition to being environmentally friendly, electric vehicles have applied more technologies such as big data and machine learning. Intelligent driving is achieved by analyzing large data samples through machine learning or deep models. Compared with general machine learning scenarios, it has higher timeliness and computing performance requirements. The emergence of 5G technology provides the basis for data perception and data communication for the intelligent driving of automobiles. The current intelligent driving [1] research covers the path rules of electric vehicles, dangerous driving supervision, automatic lane-changing, risk avoidance, object tracking, and other fields.

3. A Overview of Automatic Lane-changing Models for Electric Vehicles Based on Deep Learning

The traditional neural network realizes feature extraction through multiple activation functions,

ISSN 2616-5775 Vol. 6, Issue 1: 19-25, DOI: 10.25236/AJCIS.2023.060104

compounding, splicing, layering, and other steps. However, the automatic learning of the network reduces the workload of manual feature extraction. But the feature model extracted by the traditional neural network is one-dimensional, which is very destructive to some features with spatial properties. The application of automatic lane-changing of electric vehicles requires the processing of spatial features. Therefore, when studying automatic vehicle lane-changing, scholars have proposed a combination of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), long-short-term memory network (LSTM), gated recurrent unit (GRU), and other models are applied to the vehicle automatic lane-changing model. These deep learning models are applied to the automatic lane-change model of electric vehicles. In these network models, the characteristic data of the vehicle are obtained, including the longitudinal position, speed, distance between the acceleration and the road line, and the surrounding associated vehicle trajectories, to determine the learning rate, training times, etc.

Currently, the research on autonomous vehicle control is mainly divided into generative and research-judgment models. Firstly, the generative model adopts the method of target modeling and search matching. The controlled vehicle target is modeled according to the collected target characteristics, and similar targets matching the model are retrieved within the search range. Among the current generative model's typical algorithms such as the Kalman filter algorithm and the mean-shift algorithm. In addition, there are currently widely used particle filter algorithms; Another research and judgment model is to establish independent models for the current and the converted target vehicle and select a machine learning classifier for each model. The target object is regarded as a training sample, and the deep learning model is applied to train the sample, and the video frame in the next moment is used as a time series for classification training to obtain the position of the target area. This classifier can use SVM, and then the machine learning model is more and more perfect, and more and more models are to be chosen.

4. Automatic Lane-Changing Method for Electric Vehicles Based on DDPG Algorithm

4.1 Application of RNN in Intelligent Algorithm of Electric Vehicle

The RNN takes sequence data as the input of the model, and the output of each layer in the model can be used as part of the information content of the input item of the next layer. Therefore, RNN has the characteristics of memory context and is sensitive to the temporal features and semantics covered in the input data, so RNN can be widely used in the intelligent driving field of electric vehicles [2]. RNN can accept irregular sequence data items that are not specified in advance and receive any initial state of the vehicle. The output result is affected by the previous learning content, so the output result at a certain moment may be integrated into the previous learning to obtain features. From the perspective of input and output structures, one or more output sequences can be generated through one or more input sequences, and even under the same input conditions, the output results may be different. Figure 1 is a structural diagram of the RNN model.

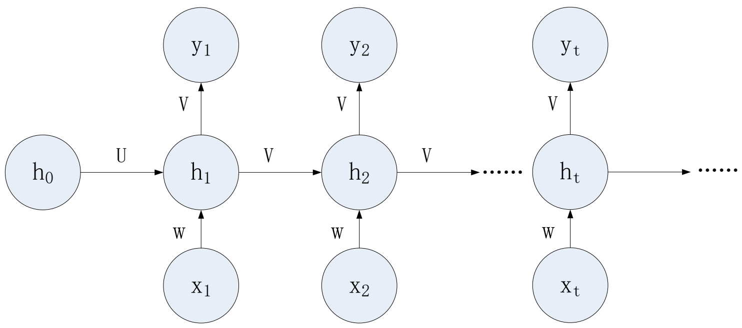

Figure 1: RNN model structure diagram

Figure 1 describes the RNN model structure, where W represents the weight value between the vehicle's perceived feature data xt and the vehicle state ht, U represents the weight value between the hidden layer ht-1 and ht, V represents the weight value from ht to output yt. Characteristic data such as radar data, vehicle coordinates or distance from the center line, etc. The state can be represented by vehicle direction angle, speed, acceleration, etc.

xt represents the input training sample of the t-th layer; ht represents the hidden features of the t-th layer. Its value is calculated by xt and ht-1; yt represents the output of the th layer, its only determined by ht. The calculation formula of the RNN model is shown in equations 1 and 2:

ISSN 2616-5775 Vol. 6, Issue 1: 19-25, DOI: 10.25236/AJCIS.2023.060104

$$h_t = f(Wx_t + Uh_{t-1}) \tag{1}$$

$$y_t = g(Vh_t) \tag{2}$$

Equation 1 represents the calculation of the vector output in the hidden layer of the RNN model, and Equation 2 represents the calculation of the output result.

4.2 Deep Reinforcement Learning Algorithm Modeling

In the reinforcement learning model, according to whether the learning and updating in the reinforcement learning algorithm model are the same strategies, the reinforcement learning model can be divided into value-based and policy-based learning methods. Value-based learning methods can be used to solve some low-dimensional space problems, while policy-based methods can be used to solve high-dimensional, high-frequency space problems. Policy-based methods can deal with complex high-dimensional space problems, but the learning efficiency of their single-step self-update is low. Therefore, the actor-critic (AC) algorithm is proposed to solve this problem. The AC algorithm fully uses the advantages of the two methods and can also realize single-step fast learning when dealing with continuous and high-dimensional spaces. In the actor-critic model, the policy gradient in the actor-network uses the value function as the benchmark iteration, which can directly interact with the external environment. After the current environment state s is collected, an action is selected according to the value of s. According to the critic network evaluation, the network model is adjusted according to the policy gradient to improve the reward value in the next evaluation. The initial state of the actor-critic model is random, and in the automatic driving lane-changing model, the actor is used to select the driving action, and the critic mainly implements the evaluation of the lane-changing driving behavior. In the model, the parameters of the actor and critic are adjusted through the reward function so that the final evaluation value of the critic network is more accurate, and the actor has more accurate decision support for changing lanes by adjusting the parameters (See Figure 2).

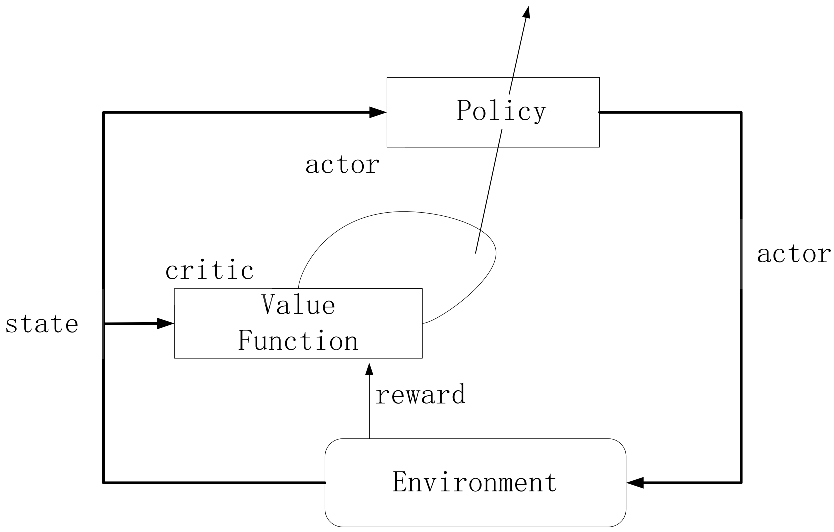

Figure 2: Principle of actor-critic algorithm

4.3 Research on Autonomous Driving Policy Based on DDPG Algorithm

The DDPG algorithm is improved based on the deep reinforcement learning algorithm. It adopts the thought of the policy gradient algorithm to realize that in the continuous action space, an action can be randomly selected according to the updated policy after learning. At the same time, a deterministic policy is added, that is, only a certain action value is allowed to be output. The DDPG algorithm uses the form of the Actor-Critic model. And the model includes two parts: actor and critic. The actor includes two neural networks of actor-network μ and target network μ', and the critic includes two neural networks of critic network Q and target network Q'. Figure 3 is a schematic diagram of the working principle of the DDPG model.

ISSN 2616-5775 Vol. 6, Issue 1: 19-25, DOI: 10.25236/AJCIS.2023.060104

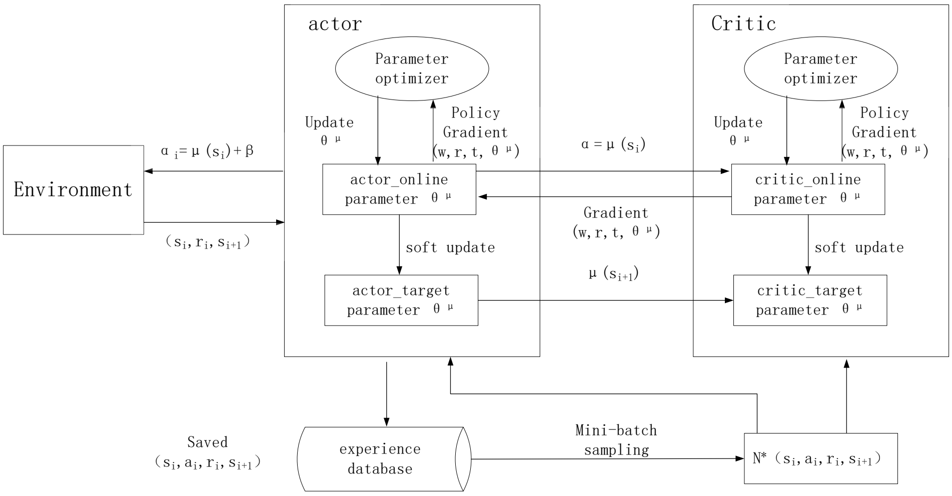

Figure 3: Working principle diagram of the DDPG model

In the initial state of the DDPG model, the weights of the current network, the size of the sample size, the learning rate, the discount rate, and other parameters are initialized. These initial values enter the actor-network, and the actor randomly generates an action according to the current weight value, and noise processing is added to the network to realize the rationality of the action. This noise action is used to calculate the environment, the reward value, and the update of the environment state. The learning results are stored in the empirical data pool. When the samples in the data pool reach the set capacity, the actor-network randomly selects N samples from the sample pool for learning and training.

In the DDPG model, the critic network parameters are first initialized, and the corresponding weights in the network model are set to θ^Q and θ^μ, initialize a random process, record the random process as N, and obtain the initial state s_1 of the automatic driving in the external environment, according to the strategy set in the current model, the executive action of noise selection is shown in Equation 3.

$$\alpha_t = \mu(s_t|\theta^\mu) + \mathrm{N}_t \tag{3}$$

In formula 3, α_t is the executive action at time t, s_t is the state at time t, and N_t is the random process generated at time t. When α_t is executed, the model returns the reward value r_t, at the same time, a new state is generated, that is, the state s_{t+1} of the next time series t+1, the transition process of this state is represented by an ordered data pair, and the state transition process is recorded as $< \mathrm{s}_t, \alpha_t, r_t, \mathrm{s}_{t+1} >$. This transfer process represents the conversion process of the state s from t to t+1, and this conversion process is stored in the playback memory unit R. R is the data set for network model training, randomly select M samples from R as the minimum batch of data samples, suppose the sample data is expressed as $< \mathrm{s}_i, \alpha_i, r_i, \mathrm{s}_{i+1} >$, and the output y_i is obtained by formula 4.

$$\mathrm{y}_i = \mathrm{r}_i + \gamma Q'(s_{i+1}, \mu'(\mathrm{s}_{i+1}|\theta^{\mu'})\theta^{Q'}) \tag{4}$$

Q' and μ' are the initial target network, and there are the following transformations in the formula: $\theta^{Q'} \leftarrow \theta^Q, \theta^{\mu'} \leftarrow \theta^\mu$.

During the automatic driving process of the vehicle, the high-frequency training process is realized, and the training and learning are used to update the state of the critic network. The procedure uses the formula for the mean square error to define the minimum loss function. The calculation of the minimum loss function L is shown in Equation 5.

$$\mathrm{L} = \frac{1}{M}\sum_i (y_i - Q(s_i, \alpha_i|\theta^Q))^2 \tag{5}$$

The update is achieved by using the policy gradient of the sample for the actor-network, and the policy gradient is shown in Equation 6.

$$\nabla_{\theta_\mu}\mu|_{s_i} \approx \frac{1}{\mathrm{M}}\sum_i \nabla_\alpha Q(s, \alpha|\theta^Q)\,|_{s=s_i, \alpha=\mu(s_i)} \nabla_{\theta_\mu}\mu(s|\theta^\mu)|_{s=s_i} \tag{6}$$

After updating the critic network and actor network, update the target network parameters accordingly, and assuming that the parameter update rate is τ(0<τ<1), the target network parameters can be obtained, see formulas 7 and 8.

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-r)\,\theta^{Q'} \tag{7}$$

ISSN 2616-5775 Vol. 6, Issue 1: 19-25, DOI: 10.25236/AJCIS.2023.060104

$$\theta^{\mu'} \leftarrow \tau\theta^{\mu} + (1 - r)\,\theta^{\mu'} \tag{8}$$

The target network is completed using multiple single-step learning iterations of the main network. The parameter values of the main network are assigned to the target network values after frequent iterations. Each step of DDPG requires a local update of the target network parameters. If the number of training reaches the minimum batch data with a capacity of M, the target network parameters are updated using the soft update algorithm.

4.4 Research on Policy Network Model of Automatic Lane-Changing Decision Controller

In the automatic lane-changing process of vehicle driving, the automatic control module of the vehicle is the main body of reinforcement learning [4]. The core control module of the vehicle needs to interact continuously with the environment to obtain the environmental state at any time to control the main body. Finally, calculate and decide the control amount of the vehicle movement according to the collected environmental state value. At the same time, in the state, the penalty value after the execution of the control amount is obtained and then enters the next state. Finally, the reinforcement learning model obtains the optimal lane-changing strategy. At the same time, this part describes the policy network model of the automatic lane-changing decision controller from the vehicle's state and action space, the design of target switching, and reward-punishment functions [5].

(1) Vehicle State Space Analysis and Design

The electric vehicle keeps driving in its lane before changing lanes. When there is a demand for lane changing and the external environment meets the conditions for lane changing, it can change lanes. When changing lanes, an electric vehicle should consider the driver's driving characteristics, such as stable and aggressive driving. This study does not consider the impact of the rear vehicle on the vehicle. In addition, in the entire reinforcement learning process, the defined state variables are set to Dline and DaheadVeh, represented respectively as the road line distance of vehicle radar detection and the distance between the vehicle and the vehicle ahead. Equation 9 defines the vehicle state space value:

$$\text{Spatiality} = \{\mathrm{D}_{line}, \mathrm{D}_{aheadVeh}\} \tag{9}$$

The range of vehicle radar detection is the left 90 degrees and the right 90 degrees in the front direction of the vehicle. Fig.4 is the control diagram of the vehicle during lane-changing. Within the detection range of the radar, if the forward direction of the vehicle is 0 degrees (it can be regarded as 0), then the left deviation is recorded as -90 degrees, and the right deviation is recorded as 90 degrees, and it is divided into 50 equal-angle units in [-90,90].

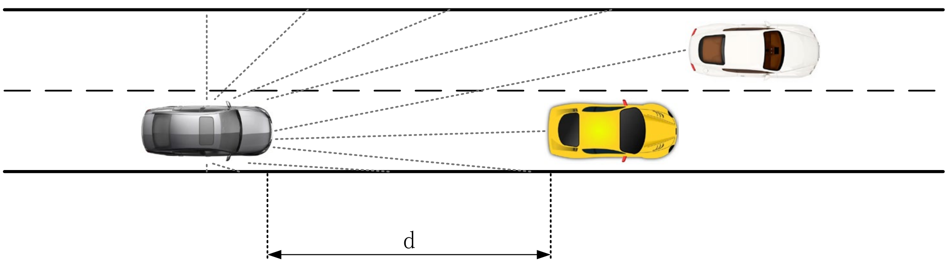

Figure 4: Schematic diagram of the control when the vehicle changes lanes

(2) Analysis and design of action space

Actions are actions taken by electric vehicles during lane changing. The lane-changing process involves the steering wheel angle, throttle, and braking control force. In the vehicle running model, the steering angle of the front wheels is the input value of the steering model. To simplify the processing, the DDPG model studied in this paper, its action space is continuous, and the steering wheel angle represents the value of the action space. And there is a positive mapping relationship between the steering wheel angle and the front wheel angle of the vehicle based on ratio conversion. The paper does not do too much research in this part, and the definition of the action space is shown in Equation 10.

$$\text{Action} = \{\theta_{\omega}\} \tag{10}$$

In formula 10, θ_{ω} is the steering wheel angle of the vehicle.

(3) Goal Switching and Reward and Penalty Function Design

ISSN 2616-5775 Vol. 6, Issue 1: 19-25, DOI: 10.25236/AJCIS.2023.060104

When the vehicle changes lanes, the state changes from lane-keeping mode to automatic lane-changing mode. Take the one-way two-lane as an example to describe the process of the automatic lane change. Figure 5 shows a schematic diagram of the one-way, two-lane automatic lane-changing. The target vehicle lane-changing target is to change from originally driving on the centerline of lane 1 (indicated by the dot-line in the figure) to driving on the centerline of lane 2. The distance between the target vehicle and the preceding vehicle is d.

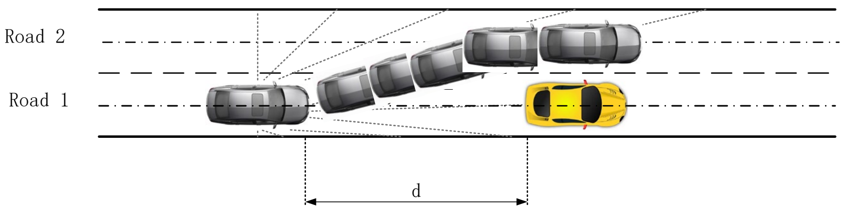

Figure 5: Schematic diagram of one-way, two-lane automatic lane-changing

Fig. 6 is a flow chart of automatic lane-changing processing, first, obtain the coordinates of the center point where the vehicle is located, and identify the lane where the current vehicle is located according to the center point coordinates. That is, the lane ID is used to represent the lane marking. The radar collects the distance d between the vehicle and the vehicle in front, and the model judges d and the ideal distance under ideal conditions. If d is smaller than the vehicle, it needs to change lanes. Otherwise, keep driving in this lane [6].

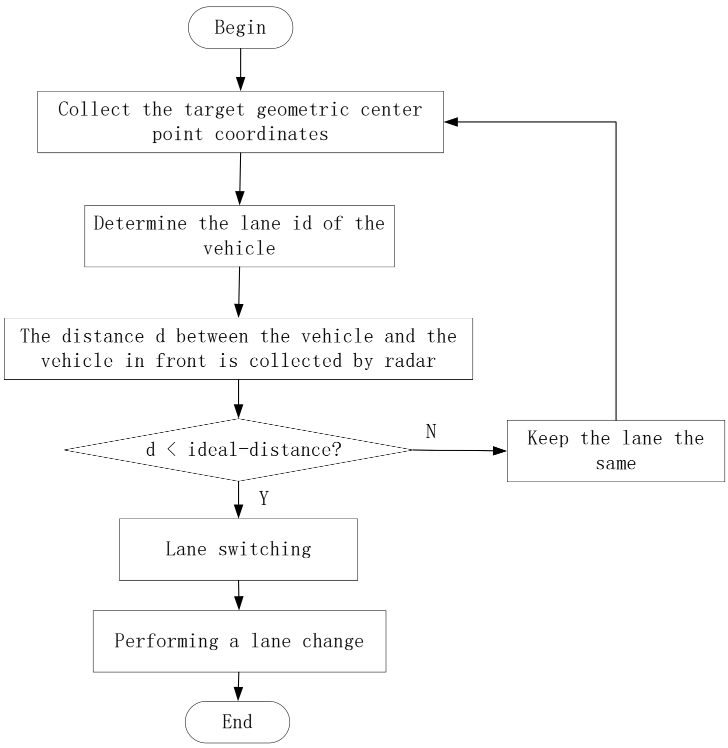

Figure 6: The flow chart of one-way, two-lane automatic lane-change processing

It is necessary to meet certain performance requirements based on completing the task to ensure the smooth completion of the deep learning task of the entire vehicle's automatic lane-changing process. Therefore, it is necessary to design a scientific reward function to improve the system's performance. The reward function in the vehicle's automatic lane-changing process involves many levels, such as traffic efficiency, safety, and vehicle comfort during driving etc. All of the above need to design the corresponding reward function. In this section, the reward function is designed by taking the vehicle's average speed as an example. And set $v_i(t)$ to be the speed of vehicle i at time step t, and the reward function is shown in Equation 11.

$$r(v) = \max\{0, \| v_{des} \| - \| v_{des} - v(t) \|\} \tag{11}$$

In formula 9, v_{des} is the expected average speed of the vehicle during driving, where the formula $\| v_{des} \| - \| v_{des} - v(t) \|$ makes the average speed of vehicles on the road as close as possible to the

ISSN 2616-5775 Vol. 6, Issue 1: 19-25, DOI: 10.25236/AJCIS.2023.060104

expected speed. The calculated result represents the degree of deviation between the average vehicle speed and the expected value.

5. Conclusion

The paper first analyzes the application status of the current electric vehicle intelligent driving system. According to the requirements of intelligent driving space characteristics, it proposes the application of deep learning in the research of electric vehicle automatic control and lane-changinig [3]. Subsequently, a car automatic lane-changing model based on the AC model was designed. Based on the application of this model, the DDPG algorithm adopted the idea of the strategy gradient algorithm to realize the conversion of actions in continuous space. The deep learning model is enriched, and finally, the policy network model of the automatic lane-changing decision controller is described from the state space of the vehicle, the action space, and the design of the target switching and reward-punishment functions. The scheme designed in this paper can be applied to the research of intelligent driving, unmanned driving theory and other related fields, which improves the accuracy of the automatic driving process.

Ackowledgement

The Paper was funded by National Natural Science Foundation of China (72101033, 71831001), Beijing Natural Science Foundation Project (KZ202210037046), Beijing Key Laboratory of Intelligent Logistics Systems (BZ0211), Canal Plan-Youth Top-notch Talent Project of Beijing Tongzhou District (YHQN2017014)

References

[1] Xia Q, Duan J, Gao F, et al. Test scenario design for intelligent driving system ensuring coverage and effectiveness[J]. International Journal of Automotive Technology, 2018, 19(4): 751-758.

[2] Duan J, Gao F, He Y. Test scenario generation and optimization technology for intelligent driving systems [J]. IEEE Intelligent Transportation Systems Magazine, 2022, 14(1).

[3] Alhariqi A, Gu Z, Saberi M. Calibration of the intelligent driver model (IDM) with adaptive parameters for mixed autonomy traffic using experimental trajectory data[J]. Transportmetrica B: transport dynamics, 2022, 10(1): 421-440.

[4] Li Y, Li L, Ni D, et al. Automatic Lane-Changing Trajectory Planning: From Self-Optimum to Local-Optimum [J]. IEEE Transactions on Intelligent Transportation Systems, 2022.

[5] Wang Lichao, et al. "A model of lane-changing intention induced by deceleration frequency in an automatic driving environment." Physica A: Statistical Mechanics and its Applications 604 (2022): 127905.

[6] Li, Ning, et al. "Reinforcement learning control method for real-time hybrid simulation based on deep deterministic policy gradient algorithm." Structural Control and Health Monitoring 29.10 (2022): e3035.