A Method for Predicting Diverse Lane-Changing Trajectories of Surrounding Vehicles Based on Early Detection of Lane Change

YUAN-YUAN REN, LAN ZHAO, XUE-LIAN ZHENG, AND XIAN-SHENG LI
School of Transportation, Jilin University, Changchun 130022, China
Corresponding author: Xue-Lian Zheng (zhengxuelian@jlu.edu.cn)
This work was supported by the National Key Research and Development Program of China under Grant 2018YFB1600502.

ABSTRACT
The trajectory prediction of surrounding vehicles is the basis for reasonable decision-making of autonomous vehicles (AV), which is helpful for improving their safety and comfort. Aiming to predict lane-changing trajectories, we propose a behavior-based method of predicting diverse lane-changing trajectories of surrounding vehicles, which includes two parts: lane-changing behavior recognition and diverse lane-changing trajectory prediction. Firstly, a lane-changing behavior recognition model based on the Continuous Hidden Markov Model (CHMM) is established to identify the lane-changing behavior of surrounding vehicles. Secondly, considering the driving styles will lead to diverse lane-changing patterns, a diverse lane-changing trajectory prediction method based on LSTM is proposed to predict three lane-changing trajectories when the driving style is unknown, which is composed of three LSTM trajectory generators representing three lane-changing patterns. Finally, the Next Generation Simulation (NGSIM) dataset is used to train, validate and test the behavior recognition model and the trajectory prediction model. The results show good accuracy and anticipative ability of the behavior recognition model. The average accuracy of surrounding vehicle behavior detection is 98.98%, the accuracy of surrounding vehicle behavior detection in 2s before lane change point is above 95%, the average anticipation time of left and right lane-changing behavior recognition is 3.24s and 3.71s, the average proportion of anticipation time in the lane-changing duration time is 46.78% and 55.54%. In the trajectory prediction section, with considering the diversity of lane changing trajectory caused by driving style, the proposed method for predicting diverse lane-changing trajectories reduces the error between the predicted and actual trajectories. The Root Mean Square Error (RMSE) and the Final Displacement Error (FDE) of the longitudinal and lateral positions are reduced by more than 21% over a 5s time horizon. In conclusion, the diverse trajectory prediction method based on the early detection of lane-changing behavior can provide AV with future trajectory of other vehicle under different driving styles, which is conducive to a more comprehensive and accurate driving risk assessment.

INDEX TERMS
Autonomous vehicle, behavior recognition, trajectory diversity, trajectory prediction, CHMM, LSTM.

I. INTRODUCTION
The structure of the autonomous driving system includes perception, prediction, decision-making, action planning and control. Among them, prediction plays a connecting role between perception and decision-making. It predicts the future states of surrounding vehicles according to the perceived information from target detection and tracking, and this predicted information can then be used for planning and decision-making. Therefore, determining how to accurately and reasonably predict the trajectories of surrounding vehicles on the road is very important, especially for the action planning of autonomous driving, which can make it more safe, efficient and comfortable.

Up to now, there are three representative research methods for trajectory prediction, shown in Fig.1, physics-based trajectory prediction, behavior-based trajectory prediction and learning-based end to end trajectory prediction. The early trajectory prediction method was physics-based trajectory prediction, which consists of three steps. First, it assumes
that the future trajectory of the vehicle is independent of the historical state; second, it takes the vehicle as a kinematic or dynamic model under certain physical rules; third, the future trajectory of the vehicle is obtained using the current state [1]–[4]. Because this method focuses on the current state of the vehicle, and the vehicle model is built under a lot of assumptions, it has achieved good results in short-term prediction, but less accuracy in long-term trajectory prediction. In recent years, with the development of computer technology, machine learning and neural network methods play a larger role in the research on vehicle trajectory prediction, therefore, data-driven trajectory prediction methods are widely used, which can be divided into behavior-based trajectory prediction and learning-based end to end trajectory prediction. Among them, learning-based end to end trajectory prediction method, which can generate vehicle’s future trajectory based on its historical information by establishing regression model and learning a large number of trajectory data [5]–[7]. For example, the deep neural network is used to learn the historical state information to generate the future trajectory of the vehicle. The main representatives of the above method are Recurrent Neural Network (RNN) [8],[9], Long Short-term Memory (LSTM) [10]–[12], Generative Adversarial Network (GAN) [13], and the Seq2Seq network based on attention [14]. However, when the driving intention is not clear, the output of the regression model tends to the average value of different intention trajectories, which obviously does not match the actual scene. Therefore, behavior-based trajectory prediction is widely used [15]–[17]. It regards the vehicle as a behavior entity, which is independent of other vehicles and subject to a series of conscious operations of the driver. In this method, the first step is to recognize the vehicle’s behavior according to its historical data. For example, vehicles may change lanes, turn or turn around. The next is to predict the future trajectory of the vehicle based on its behavior. This method takes the vehicle intention or behavior as a priori knowledge to reflect its future driving direction. Compared with the physics-driven trajectory prediction method, the behavior-based method has higher accuracy and longer prediction times (>2s). In this paper, we adopt behavior-based trajectory prediction method that include the behavior identification module and the trajectory prediction module. The former obtains the future driving behavior and driving direction of the surrounding vehicles, and the latter predicts the future trajectory under this behavior.

At present, research in trajectory prediction faces two major difficulties: modeling the interaction of different traffic participants on the road and inferring diverse future trajectories. Many scholars have studied the vehicle interaction problem in trajectory prediction, and contributed two solution methods. One is to extract the vehicle interaction modes based on vehicle trajectories [18], [19], while the other is to build a structural neural network model by imitating the position relationship between vehicles, and learn the interaction rules between vehicles based on this model [20]. In this paper, the influence of vehicle interaction on trajectory prediction is not considered. We weaken the interaction between vehicles and focus on providing a diverse lane changing trajectory prediction method.

In the mixed traffic flow of MVs and AVs, the trajectory prediction of MV is helpful for AV to evaluate the future driving risk and make decisions. However, each vehicle in the traffic scene is full of uncertainty and diversity, and the diversity of driving styles is bound to lead to diverse lane-changing trajectories. So, it is more helpful to decision-making to predict the diverse trajectories of surrounding vehicles than a single trajectory. Therefore, the main purpose of this paper is to predict diverse future trajectories according to the historical information of the vehicle. At present, there are two ways to study the prediction of diverse lane-changing trajectories. One is to add a part that can express trajectory diversity outside the trajectory prediction model, while the other is to reflect the trajectory diversity in the trajectory prediction model. The first method is widely used in both physics-driven and data-driven trajectory prediction. In the physics-driven trajectory prediction methods, the Gaussian noise method [21] and the Monte Carlo method [22] are used to represent the diversity of predicted trajectories. The former uses the unimodal Gaussian distribution, which expresses the deterministic vehicle position as a distribution at each moment; the latter first generates a regional trajectory, then uses the weighted division method to delete the regional trajectory based on artificially defined constraints, and finally obtains regional multiple trajectories to represent various trajectories. In the data-driven trajectory prediction methods, it is common to apply an MDN (mixture density network) layer in the trajectory output model [12]. The purpose is to use a probability distribution to represent the future positions of vehicles, rather than to just predict a certain vehicle position. The second stream of trajectory prediction research believes that trajectory diversity can be extracted from datasets, and then integrated into the trajectory prediction model so that a variety of predicted trajectories can be obtained. Taking the
prediction of lane-changing trajectories based on the prototype trajectory method [23]–[25] as an example, the approach is centered around the idea that the variety of lane-changing trajectories can be clustered into typical motion patterns, where each typical pattern can be represented by one or more prototype trajectories. By looking at the similarity between the given historical trajectory and each prototype trajectory, the remaining segment of the best-fitting prototype trajectory is used as an estimate for the future motion of the predicted vehicle. This method considers that the future lane changing trajectory of vehicle comes from the extracted prototype trajectory library, which has poor scene adaptability, and ignores the randomness of vehicle state changing with time in the process of lane changing.

Based on the above, this paper proposes a prediction method for diverse lane-changing trajectories by referring to the idea of prototype trajectories. The trajectory prediction method used in this paper is behavior-based trajectory prediction, which mainly includes two modules: behavior identification and trajectory prediction. Firstly, the future behavior of the surrounding vehicle is identified through the behavior identification module. Secondly, the diverse lane-changing trajectories under lane-changing behavior can be obtained through the trajectory prediction module. As for the trajectory prediction module, a diverse lane-changing trajectories prediction model based on LSTM (Long-Short term Memory) network is established, which can generate lane-changing trajectories that reflect three driving styles according to the historical trajectories of the surrounding vehicle for a certain period of time. The main contributions of this paper are as follows:

1) Determination of lane-change process. In the research on lane-changing trajectory prediction, the first step is to extract the lane-change process from the dataset. This paper adopts the unsupervised clustering method to determine the lane-changing process. Through unsupervised clustering of the time series, the lane-keeping states that occur before and after the lane-changing can be separated automatically from the lane-changing process, so that the lane-changing process can be obtained. This method saves time and reduces the influence of subjective factors while retaining the original randomness of the lane-change data.

2) Feature group extraction of lane-change behavior. In this paper, we propose a feature group extraction method for lane-change behavior. Based on the information gain rate and Pearson correlation coefficients, the features with high importance and low cross-correlation are selected to form the lane-change behavior feature groups, which allow us to clearly distinguish lane-changing from lane-keeping behaviors. This is then used to train the lane-change behavior recognition model, and achieves good results.

3) A diverse trajectory prediction model considering the lane-changing duration time is proposed. We believe that the lane-changing duration time reflects the diversity of the lane-change process to a certain extent, so we propose a diverse lane-changing trajectory prediction model considering the lane-changing duration time. The key is to establish the corresponding LSTM trajectory prediction models for different lane-changing patterns. This method produces multiple future trajectories according to the historical trajectories of the predicted vehicle, reflecting the lane-changing processes of different duration times. It not only ensures the accuracy of the trajectory prediction, but also represents the randomness and diversity of lane-change behavior caused by driving styles. Most importantly, it provides a reference for the danger area when the surrounding vehicle cuts in, which aids the advance decision-making and path-planning of the autonomous vehicle.

The remainder of this paper is divided into the following sections. A system overview is introduced in Section II. Section III introduces the dataset used in the paper. Section IV describes the process of lane-change behavior recognition. In Section V, the specific implementation of the proposed diverse trajectory prediction model and the results are discussed. Finally, Section VI concludes the paper with an outlook and a discussion of future work directions.

II. SYSTEM OVERVIEW

In this paper, a method for predicting diverse lane-changing trajectories is proposed, which includes a behavior recognition model and a trajectory prediction model, as shown in Fig.2. This paper mainly focuses on early detection of lane change and trajectory prediction of surrounding vehicles. Using the data-driven method, the lane-changing behavior recognition model is established based on the classification algorithm, and the trajectory prediction model is established based on the regression algorithm. First, the lane-change behavior recognition model for other vehicles is established, with the early detection of lane change through the historical driving state of the vehicle. Then, based on the LSTM network, the framework for predicting diverse lane-changing trajectories is established, which adopts the encoder-decoder method to model the relationship between the historical and future trajectories of the vehicle, and thus predicts the future trajectory based on the historical trajectory.

III. DATASET INTRODUCTION

A. BASIC CHARACTERISTICS OF DATASET

1) INTRODUCTION OF I80 AND US101 DATASETS

The dataset used in this paper is from the Next Generation Simulation (NGSIM), collected and published by the US Federal Highway Administration in 2005 [26]. More specifically, the data of two regions are used to train, validate and test the behavior recognition and trajectory prediction model. One is the I-80 freeway in Emeryville, California, the segment covered being approximately 500m in length and 6 lanes (3.66m or 12ft each) in width (see Fig.3). The other is the US101 freeway in Los Angeles, California, the segment covered being approximately 640m in length and 5 mainline lanes in width (see Fig.4).
Y.-Y. Ren et al.: Method for Predicting Diverse Lane-Changing Trajectories of Surrounding Vehicles

FIGURE 2. System overview: A diverse trajectory prediction method based on early detection of lane-changing behavior. Which includes behavior identification model and diverse trajectory prediction model. Lane-changing behavior identification model can judge the behavior of surrounding vehicles based on history information; the trajectory prediction module includes trajectory generation models of three lane-changing patterns (LSTM-LCL1, LSTM-LCL2, LSTM-LCL3), which are established based on LSTM, trained with the sample sets of three lane-changing patterns obtained by clustering the lane-changing samples. The trajectory prediction model generates diverse lane-changing trajectories based on behavior recognition results and vehicle historical trajectories.

FIGURE 3. Vehicle trajectory study area of I80 highway.

2) AVAILABLE INITIAL PARAMETERS
The dataset provides trajectories of individual vehicles with a sampling rate at 10 Hz. Each sample trajectory includes information such as instantaneous velocity, acceleration, longitudinal and lateral positions (both local and global), vehicle length and width, vehicle type, lane ID, and vehicle ID. The local coordinate is set at the upper-left point of the study area, where $x$ is the lateral position of the vehicle relative to the leftmost edge of the road, and $y$ its longitudinal position relative to the entry edge. The original data has certain error and noise, the symmetric exponential moving average filtering algorithm (sEMA) [27] is used to filter the original data.

B. ANALYSIS OF TRAFFIC FLOW CHARACTERISTICS
1) VELOCITY ANALYSIS
The statistical analysis results of the instantaneous velocities of the vehicles on the I-80 and US101 are shown in Fig.5.

Fig.5(a) shows the frequency distribution histogram of the instantaneous velocity on the I80. It can be seen that the instantaneous velocities of vehicles on the I80 are roughly in the range of [0m/s,15m/s], while Fig.5(b) shows that the instantaneous velocities of vehicles on the US101 are roughly within [0m/s,20m/s]. It should be noted that the reasons for the pulses in Fig.5 are: 1) the vehicle stops moving in a congested state, 2) the vehicle runs slowly at a low speed, and 3) the vehicle runs within a certain speed range when it is in free flow. And Fig.5 shows the frequency distribution histogram of instantaneous velocities, so the larger pulse will be formed if there is a certain range of vehicle velocity occur more frequently.

The results of the velocity analysis in the data collection areas show that the driving velocities of the vehicles are low, which may be due to the high density of vehicles and the
mutual restriction of the driving state by the vehicles. In this traffic environment, vehicle lane changing is relatively slow and lasts for a long time, which increases the randomness and diversity of lane-changing behavior, and is conducive to the research of diverse lane-changing trajectory prediction.

2) VEHICLE SPACING ANALYSIS
The vehicle spacing distribution in different velocity ranges of I-80 and US101 are shown in Fig.6. We can find that the greater the velocity, the greater the vehicle spacing. It can be seen from Fig.6(a) that the vehicle spacing on I80 is roughly in the range of [10m, 60m] under the velocity of [5, 25]m/s. Meanwhile, Fig.6(b) shows that the vehicle spacing on the US101 section is roughly within [10m, 40m] under the velocity of [5, 25]m/s.

Through the analysis of vehicle spacing, it can be seen that the distance between the vehicles is small, at far less than 150m [28]. In this traffic environment, the observed vehicles are closely related to the surrounding vehicles and interact with each other. On the one hand, the intentions of the lane-changing vehicles come from their own expectations; on the other hand, they come from the influence of surrounding vehicles. Therefore, the traffic flow state of the dataset provides a rich basis for the early detection of vehicle lane-change behavior.

3) LANE-CHANGING BEHAVIOR ANALYSIS
The trajectories of all the vehicles on the I80 and US101 are shown in Fig.7. It can be seen that lane-change behavior occurs frequently in the studied sections. The analysis results for vehicle spacing and instantaneous velocity also show that the studied sections have a high density of vehicles, and the mutual restriction between vehicles leads to frequent lane change behaviors and diverse trajectories, forming a wealth of lane-changing trajectories, which is conducive to the study of lane-change behavior recognition and trajectory prediction.

IV. EARLY DETECTION OF LANE CHANGE
Lane-change detection can be achieved by training a pattern classifier, which can distinguish between lane-keeping (LK), left-lane-changing (LCL) and right-lane-changing (LCR). This pattern classifier can be regarded as a lane-change behavior recognition model, which can automatically identify vehicle behavior according to the change trend of features. The recognition of lane-change behaviors is related to three factors: input sequence length, input features and model
Y.-Y. Ren et al.: Method for Predicting Diverse Lane-Changing Trajectories of Surrounding Vehicles

FIGURE 7. Trajectories of all vehicles on I80 and US-101 highway. Including left lane-changing trajectories, right lane-changing trajectories and lane-keeping trajectories.

type. Therefore, this section discusses the recognition of lane-change behaviors in three parts: the division of the lane-change stage, the selection of the lane-change behavior feature group and the construction of the model for recognizing lane-change behavior earlier.

A. LANE-CHANGING STAGE DIVISION

1) LANE-CHANGING STAGE DEFINITION

The schematic diagram of a complete lane-change process is shown in Fig. 8, which mainly includes the preparation stage before a lane change, the lane-change execution stage and the adjustment stage after a lane change. Taking the left lane change as an example, we first define the lane-change-preparation stage (T1-T2), from the time and distance that when the vehicle begins to approach the target lane until the vehicle reaches the lane boundary. In this stage, the driver’s lane-change intention becomes gradually clearer, and the feasibility of the lane change is tested. If at any time the driving environment does not satisfy the lane-change requirements, the driver may cancel the lane change. Second, we define the lane-change-execution stage (T2-T3), from the time and distance that when the vehicle reaches the lane boundary until the vehicle leaves the lane boundary line [29]. In this phase, the vehicle performs the lane-change operation and changes driving lanes. Third, we define the lane-change-adjustment stage (T3-T4), which is when the vehicle has changed to the target lane successfully and is adjusting its position in the lane.

The different lane-change stages represent different lane-change states of the vehicle, and are used in different studies. For example, the data for the lane-change-preparation stage (T1-T2) can be used to train a lane-change behavior recognition model, and the data for the whole of the lane-change stage (T1-T4) can be used to train a lane-changing trajectory prediction model. Therefore, extracting and dividing lane-changing process (T1-T4) from trajectory are of great significance for both lane-change behavior recognition and lane-changing trajectory prediction. The lane-change process can be divided into three stages by four key points T1-T4. Among them, the starting point T2 and the ending point T3 of the lane-change-execution stage are easy to determine, according to the lane line and the lateral position of the vehicle, however, for the start point of lane change preparation stage (T1) and the end point of lane change adjustment stage (T4), some researchers use the threshold method of fixed threshold to determine. It is considered that, if the vehicle reaches an artificially defined threshold, it has entered the lane-change process; otherwise, it is in the lane-keeping state. However, a lane change is a random and diverse process, so it is impractical to use a fixed threshold to identify it. Others pay attention to the use of turn signals [30] and the driver’s gaze characteristics [31]. The beginning of lane change intention can be defined as the turn-signal starting time or the last moment of glancing the view mirror before switching the turn light on. The above methods define the starting point of lane changing intention based on driver behavior. As a bystander, it is difficult for us to obtain driver’s information of surrounding vehicle, however, the trajectory can indicate that the spatial position of the vehicle changes with the time, and the trajectory difference between lane-keeping and lane-changing is obvious. Therefore, we propose a lane-changing process extraction method which is similar to the threshold method but suitable for diverse trajectory samples. In this method, T1 and T4 can be obtained by clustering vehicle trajectories, and then the lane change process can be extracted from the whole trajectory. For vehicle trajectories with different driving styles, we can obtain the lane changing process under the corresponding style, rather than hard dividing by a fixed threshold.

2) DETERMINATION OF KEY POINTS OF LANE-CHANGING BASED ON K-MEANS

Before carrying out the lane-change stage division, it is necessary to determine the initial range of the lane change for each sample. Because the lane-changing duration time is generally 1-16s [28], in order to ensure that the initial range of the lane
Y.-Y. Ren et al.: Method for Predicting Diverse Lane-Changing Trajectories of Surrounding Vehicles

FIGURE 8. Schematic diagram of lane-change stage division. The red vehicle is lane changing, other vehicles are surrounding vehicles. T1 is the start point of lane change preparation stage, T2 is the start point of lane change execution stage, T3 is the start point of lane change adjustment stage, and T4 is the end point of lane change process. The lane change process is divided into lane change preparation stage (T1–T2), lane change execution stage (T2–T3) and lane change adjustment stage (T3–T4).

change includes the complete lane-change process, this paper uses a 20s time series as the initial range of each lane-change sample, which is the 10s before and 10s after the vehicle crosses the line.

With the initial range of the lane change determined, each lane-changing trajectory is divided into three stages, which are the lane-keeping, lane-change process and lane-keeping, by K-means. Taking the left lane change as an example, the results are as shown in Fig.9. The upper right small figure in Fig.9 shows the clustering results of a single lane-changing trajectory, in which the blue parts are the lane-keeping stages before and after the lane change, and the red part is the lane-change process. According to this, the start point T1 and end point T4 of the lane change process can be obtained. It should be noted that the trajectories of lane changing vehicles involved in Fig.9 are from the local X, Y coordinate mentioned above.

In order to validate the accuracy of the K-means clustering in the lane-keeping and lane-changing stages, we evaluate the clustering effect from two aspects: the internal clustering effect and the external clustering effect.

The first is the evaluation of the internal clustering effect based on the silhouette coefficient. This can be used to evaluate the density degree between the same class and the dispersion degree between different classes after classification, and its variation range is \([-1, 1]\). The larger the value is, the better the clustering effect is. The silhouette coefficient value of each sample after clustering is shown in Fig.10, the average silhouette coefficient is 0.7636, which shows that the clustering effect is reasonable.

The second aspect is the external clustering effect evaluation based on a statistical analysis of the clustering stage and the actual stage, where the clustering effect is tested by comparing the similarity between them. Here, the actual stage is obtained using a manually defined parameter threshold. We randomly select 50 samples for the external clustering effect evaluation, and the statistical analysis results are shown in Fig.11 and Fig.12. Fig.11(a) is the frequency distribution histogram of the lane offset. Red is the lane offset of the manually divided stage, while green is the lane offset of the clustering stage, and brown is the overlapping part, reflecting the similarity of the two results. The overlapping part accounts for 91.3% of the real lane-keeping stage. Fig.11(b) is the boxplot of the lane offset under the two clustering methods. It can be seen that the statistical features of the two groups of data are similar, and the difference in the average values is about 0.22% of the average value of the real lane-keeping stage. Fig.12 shows the evaluation of clustering.
effect of left lane change, the coincidence rate is 94.7% and the average difference is 4.2%.

In conclusion, the starting point T1 and ending point T4 of a lane change process can be determined based on K-means clustering. Thus, all the key points in the lane-change process can be obtained, which is helpful for separating each lane-change stage of a vehicle’s lane change from the dataset conveniently and quickly, and these lane-change stages can then be used for vehicle behavior recognition and trajectory prediction.

**FIGURE 11.** Comparative analysis of clustered and actual lane-keeping stages. (a) shows the distribution comparison of laneoffset in LK-clustered (green) and the LK-labeled (red), the overlapping part (brown) shows the similarity between them (b) shows the statistical characteristics of the lane offset under two methods by the boxplot.

**B. DETERMINATION OF LANE-CHANGING-BEHAVIOR FEATURE GROUP**

In research on lane-change-behavior recognition, the determination of the lane-change-behavior feature group is as important as the division of the lane-change stage. The purpose of recognizing the lane-change behaviors of surrounding vehicles is to serve the decision-making process of autonomous vehicles. Therefore, the choice of the lane-change-behavior feature group should obey the following three principles:

1) It should be easy to obtain.

2) The selected features should be able to distinguish lane-change from lane-keeping in the early stage of lane change.

3) The cross-correlation between the selected features should be low to reduce the complexity of model training.

Based on the above three principles, this section constructs and selects lane-change-behavior features.

**1) CONSTRUCTION OF LANE-CHANGING-BEHAVIOR FEATURES**

We construct the driving state features X1-X7 of the lane-changing vehicle, based on the vehicle position (x, y) which reflects the state changes of the vehicle in the lane-change-preparation stage. It should be noted that feature X1 is the lateral distance between the lane change vehicle and the leftmost lane line in the lane change affected area, the lane change affected area is the lane where the lane change vehicle located and the adjacent lanes on both sides. At the same time, according to the positions of the lane-changing vehicle and surrounding vehicles, the feature group X8-X12 is constructed. Among them, X8-X10 reflect the relative driving states of the lane-changing vehicle and the surrounding vehicles, and X11 and X12 are used to represent the expected driving state of the lane-changing vehicle. The specific features and construction methods [32] are shown in Table 1. The construction formulas of X2-X6 are as follows, where the x and y in Formula (1)[12] are the lateral and longitudinal

**FIGURE 12.** Comparative analysis of clustered and actual lane-change stages. It is similar to the lane-keeping phase in Figure 10.
positions of the vehicle.

\[
\text{HeadAangle}_{s} = \arctan\left(\frac{x' - x'^{-3}}{y' - y'^{-3}}\right) \quad (1)
\]

\[
\text{HeadAangleRate}_{s} = \frac{\text{HeadAangle}_{s}^{(t)} - \text{HeadAangle}_{s}^{(t-1)}}{\Delta t} \quad (2)
\]

\[
V_{XS} = V_{s} \sin(\text{HeadAangle}_{s}) \quad (3)
\]

\[
V_{YS} = V_{s} \cos(\text{HeadAangle}_{s}) \quad (4)
\]

\[
a_{XS} = a_{s} \sin(\text{HeadAangle}_{s}) \quad (5)
\]

\[
a_{YS} = a_{s} \cos(\text{HeadAangle}_{s}) \quad (6)
\]

2) SELECTION OF LANE-CHANGING-BEHAVIOR FEATURES

In order to reduce the computational complexity of the model, it is necessary to select which features to include. In this paper, features are selected from those constructed in Table 1 based on their importance and the correlations between them.

The first element looked at is the importance of the features. We use the information gain rate [33] to quantify this. The higher the information gain rate is, the better the classification effect of the feature is, and the higher its importance. Because the lane-change-behavior features are continuous time series, including too many of them will increase the computational complexity of the information gain rate.

Considering that the duration time of the lane-change-preparation stage is short and the numerical fluctuation of the features is small, a series of statistical values can reflect the change trend in the time series. Thus, the continuous features are regarded as primary features. Then, by using the statistical value construction method to discretize the primary features, we construct the corresponding statistical sub-features based on the primary features in Table 1. The importance of the primary features is obtained by calculating the information gain rate of the statistical sub-features. The construction process is shown in Fig.13. 84 statistical sub-features can be obtained by constructing statistical values. The calculation steps for the information gain rates of each of the statistical sub-features are shown in formula (7)-(11) [34].

**FIGURE 13. Construction of statistical sub-features of primary features.**

There are 12 main features on the left, for each main feature, we use the statistical value construction method to construct 7 statistical sub-features, and finally 84 sub features can be obtained.

Firstly, 50 LK samples, 50 LCL samples and 50 LCR samples are randomly selected, and the 84 sub-features of each sample form the initial dataset D. The uncertainty of the dataset D is characterized by calculating the information entropy, \(\text{Info}_{D}\), as shown in formula (7), where m is the number of classes in the dataset D. Since there are three classes, m is 3.

\[
\text{Info}_{D} = - \sum_{i=1}^{m} p_{i} \log_{2} p_{i} \quad (7)
\]

Then, according to the value of statistical sub-feature A, the dataset D is divided into n sub-datasets. At this time, the information entropy of dataset D is \(\text{Info}_{A,D}\), as shown in formula (8), where \(|D_{j}|\) represents the number of samples in the \(j\)th sub-dataset, and |D| represents the total number of samples in the dataset before division.

\[
\text{Info}_{A,D} = \sum_{j=1}^{n} \frac{|D_{j}|}{|D|} \times \text{Info}_{D,j} \quad (8)
\]

After dataset D has been divided based on statistical sub-feature A, the information gain of dataset D is \(G_{A}\), as shown in formula (9):

\[
G_{A} = \text{Info}_{D} - \text{Info}_{A,D} \quad (9)
\]

However, there is a drawback when using information gain to evaluate the classification performance of features: it is easy to select features with more feature attributes as important features. Therefore, we use the information gain rate \(\text{GR}_{A}\) to quantify feature importance, obtained by dividing the information gain by the split information \(\text{SplitInf}_{A,D}\) as shown in formula (11):

\[
\text{GR}_{A} = \frac{G_{A}}{\text{SplitInf}_{A,D}} \quad (10)
\]

\[
\text{SplitInf}_{A,D} = - \sum_{j=1}^{n} \frac{|D_{j}|}{|D|} \log_{2} \frac{|D_{j}|}{|D|} \quad (11)
\]

According to the above steps, we obtain the information gain rates of the 84 statistical sub-features, and their ranking is shown in Fig.14, which takes the primary feature as the unit, and uses the information gain rate ranking of the statistical sub-features to represent the importance ranking of the primary feature, shown as a box in the figure. The horizontal axis of the figure represents the 12 primary features, and the vertical axis the information gain rate ranking of the statistical sub-features corresponding to each primary feature. In order to select the features with high importance, we use the 75th percentile, as shown by the red line in the figure, and therefore delete features X5, X9, X10 and X12 whose information gain rate rankings are outside that.

After the analysis of feature importance, it is necessary to analyze the correlation of the residual features in order to reduce feature redundancy. We use the Pearson correlation coefficient to characterize the cross-correlation between features. Table 2 shows them. If the correlation coefficient is greater than 0.8, it means that there is a strong correlation between the features. In order to reduce the complexity of model training, the features X6, X7 and X8 are excluded, as they are strongly correlated with other features under the...
C. LANE-CHANGING BEHAVIOR RECOGNITION MODEL BASED ON CHMM

Lane-changing is a continuous process with a strong time sequence, while lane changing behavior is difficult to determine at an early stage, but can be inferred indirectly from some continuous features. In this paper, the Continuous Hidden Markov Model (CHMM) [35], which is commonly used in the field of pattern recognition, is selected to establish the lane-change-behavior recognition model.

1) CHMM

A complete CHMM can be represented by a 7-tuple \( \lambda_c = \{N, M, \pi, A, C_{jm}, \mu_{jm}, \sum_{jm}\} \). It is necessary to give initial values to the parameters before model training. Among them, \( \pi \) and \( A \) are the initial hidden state and state transition matrix, and their initial values have little influence on the recognition effect of the model, so they can be assigned randomly. \( N \) is the number of hidden states. It is considered that there are three kinds of hidden states: LK, LCL and CLR, so the value of \( N \) is 3. Since the input variables are continuous, we use M Gaussian probability density functions to fit the relationship between the hidden state and the observed variables, and the parameters of that function are given initial values by K-means clustering.

2) MODEL DESIGN AND TRAINING

The lane-change-behavior recognition model includes LK hidden Markov model, LCL hidden Markov model and LCR hidden Markov model, with a supervised training pattern, and each model using the corresponding training data to train the model. The time-series features of the model input are as follows:

\[
O(t) = \{\text{laneoffset}_S(t); Vx_S(t); Vy_S(t); ax_s(t); \Delta y_{S, TP}/\Delta y_{S, P}(t)\}
\]  

(12)

The parameter design and dataset allocation of the model are shown in Table 3.

Based on the training set, the lane-change-behavior model and lane-keeping model are trained respectively to obtain the optimal parameter group of the model, and then the lane-change-behavior recognition model is obtained. Based on the verification set and test set, the recognition accuracy and recognition timeliness of the behavior recognition model are verified. Among them, the samples used in the train set and the verification set are feature sequences under specific behaviors, while the test set is a feature sequence including both the lane-keeping and lane changing, so there is no lane-keeping test set.
3) VALIDATION OF BEHAVIOR RECOGNITION

The performance of the behavior-recognition model is very important, and related to whether the trajectory-prediction model is timely and accurate. In this paper, we validate the performance of the lane-change-behavior recognition model from two aspects. One is to evaluate the performance of behavior recognition based on the validation set, while the other is to evaluate the time-efficiency of the behavior recognition based on the test set. There are two reasons for this. One is that the samples required are different. The verification of recognition accuracy is based on the verification set, and the analysis of anticipation time is based on the test set, the other is that the analysis methods are different. The former uses the feature sequence under specific behavior for off-line recognition, and the latter uses the sliding window method to continuously recognize the behavior corresponding to the feature sequence in each window.

The recognition result of the behavior-recognition model can be represented by a confusion matrix, as shown in Table 4. We use the commonly used evaluation indexes of the classification algorithm to evaluate the recognition performance of the model, namely precision, recall, F1-scores and accuracy. The results are shown in Table 5. It can be seen that the established model has good performance in identifying lane-keeping and lane-changing, and each index is above 0.95, which indicates that the lane-change-behavior-recognition model has good recognition ability.

For the misclassified cases, a reason analysis is required to better improve the model. Fig.15 is the trajectory of each misclassified case. It can be seen that these cases have abnormal fluctuations. LK cases 1 to 3, it was misclassified as lane change due to significant offset to the left or right. LC cases 4 to 5, it was misclassified as lane-keeping due to trans-line phenomenon of the lane changing vehicle. LCR case 6 was misclassified as LCL state, because the vehicle tended to return to the original lane after crossing the line.

In our future research, we will focus on the inspection of these factors that are easy to lead to misclassification. In the preprocessing step, delete the samples with excessive chattering and nonstandard driving as soon as possible, so as to obtain high-quality samples.

In addition to detecting lane-change behavior accurately, the lane-change-behavior recognition model should also have the ability to detect other vehicles’ lane-change behaviors as soon as possible. The common indicator used to evaluate the recognition timeliness of the model is the anticipation time, which is defined as the time interval from the time when the vehicle is detected the lane-change behavior to the time when the vehicle drives to the starting point T2 of the lane-change execution stage. We use the sliding-window method to gauge the dynamic recognition of lane-changing behavior. The sliding-window method is based on the test set, and uses a sliding window with a length of 2 and a step size of 1 to extract features continuously for lane-change recognition, and then obtains the dynamic recognition results. The results are shown in Fig.16. Fig.16(a) shows an example of LCL behavior recognition, and Fig.16(b) shows an example of LCR behavior recognition. Using the sliding-window
method, we identify the lane-change behavior dynamically based on the test set, and the lane-change anticipation times are shown in Fig.17. Fig.17(a) shows the frequency distribution histogram of the anticipation time. It is found that the average anticipation time of the LCL and LCR behavior recognition models are 3.24s and 3.71s. Fig.17(b) shows the accuracy of the model at 3, 2, 5, 2, 1.5, 1, 0.5 and 0s before the lane change execution start point (T2). It can be seen that the behavior recognition accuracy increases gradually as the predicted vehicle drives to the lane change point. When the vehicle moves to 2s before the lane change point, the lane-changing behavior can be detected with an accuracy rate of more than 95%.

The datasets collected under different traffic flows will have significant differences in their lane-changing duration times, which will further affect the length of the anticipation time. As a result, it would be inconvenient to compare different research results with the anticipation time directly. Therefore, this paper proposes that the anticipation time proportion is used. It is defined as the ratio of the anticipation time to the lane-changing duration time. This index can eliminate differences in lane-changing duration times caused by different traffic environments, so as to compare the research results of different datasets. Fig.18(a)(b) shows the anticipation time proportion of LCL and LCR. In the figure above, blue is the lane-changing duration time of each sample, and red is the anticipation time. It can be seen that the anticipation time proportion is more intuitive. The figure below shows the statistical distribution histogram of the anticipation time proportion. The results of 100 test samples show that the average anticipation time proportion of LCL and LCR recognition

In addition, considering that lane-change behavior is easily affected by the traffic environment and driving style,
models are 46.78% and 55.54%. This index can provide a benchmark for the comparison of the conclusions of different studies in the future.

**D. MODEL DOMINANCE ANALYSIS**

In order to prove that the lane-change-behavior recognition model in this paper has some advantages, we compare its results with those of other studies [12], [36], which are all based on the same dataset. The comparison is shown in Table 6. Based on the same evaluation index, we can conclude that the proposed lane-change-behavior recognition model based on the CHMM outperforms the SVM and LSTM.

**V. TRAJECTORY PREDICTION**

In the behavior-based trajectory prediction method, the behavior identification module is applied firstly. Once the vehicle’s lane-changing behavior is identified in the early stage, the trajectory prediction module can be applied to predict the lane-changing trajectories of the surrounding vehicle. Considering that the surrounding vehicles are mainly driven by manual drivers, the lane-changing behavior is very random due to the influence of different driving styles and environments, which leads to a diversity of lane-changing trajectories. Therefore, compared with predicting a certain trajectory, the diverse trajectory prediction results are more meaningful for the decision-making of autonomous vehicles. As one of the manifestations of lane-change diversity, lane-change duration time reflects this diversity indirectly. Therefore, we take the LCL as an example, from the perspective of lane-changing duration time, this section predicts the diversity of lane-changing trajectories and expresses the future positions of surrounding vehicles as comprehensively as possible.

**A. LANE-CHANGING DIVERSITY ANALYSIS CONSIDERING DURATION TIME**

There are differences among different drivers in terms of the lane-change process, and quantifying and characterizing these differences is very important. In this section, from the perspective of lane-change duration time, we use fuzzy C-means (FCM), which is an unsupervised clustering method, to analyze the diversity of lane-changing trajectories. First, we cluster all the lane-changing sample, and obtain the kinds of lane-changing trajectory sets. Then, the diversity of the lane-changing sample is analyzed to obtain the kinds of lane-changing trajectory clusters and prototype trajectories.

1) **LANE-CHANGING DURATION TIME ANALYSIS**

The duration time of a lane change is defined as the time interval from the time when the vehicle begins to move laterally to the time when the vehicle is driving stably in the target lane, that is, the time interval between T1 and T4. Based on the lane-change key points T1 and T4, the lane-change duration times of all left-lane-changing samples was calculated. The results are shown in Fig.19 and 20. The frequency distribution histogram of the lane-changing duration time is shown in Fig.19. It can be seen that the lane-change duration time of the left-lane-changing samples is mainly distributed in the range of 3s-13s. Fig.20 shows the actual lane-changing trajectories of all of the lane-changing samples.

---

**TABLE 6. Lane-changing behavior recognition performance evaluation.**

| Approach   | Precision | Recall | F1-scores | Accuracy |
|------------|-----------|--------|-----------|----------|
| Ji Xue-wu  | SVM       | 0.8418 | 0.8319    | 0.8412   | 0.8319   |
| (2019)     | LSTM      | 0.8787 | 0.8756    | 0.8756   | 0.8756   |
| Berents[5]  | SVM       | 0.958  | 0.957     | 0.957    | 0.971    |
| (2019)     | ANN       | 0.983  | 0.983     | 0.983    | 0.988    |
| Proposed   | CHMM      | 0.9913 | 0.9898    | 0.9905   | 0.9898   |

---

**FIGURE 18.** Anticipative ability of the recognition model. In (a) and (b), the upper figure shows the lane changing duration (blue) and anticipation time (red) of each test sample, the lower figure shows the proportion of anticipation time in the lane changing duration.
We can see that the lane-changing trajectories of the different vehicles are different, and the lane-changing duration time shows obvious diversity. Therefore, the lane-changing duration time is an important factor in lane-changing trajectory prediction.

2) LANE-CHANGING DIVERSITY ANALYSIS BASED ON FCM

The different duration times of lane changes will lead to uneven lengths of lane-changing samples, resulting in different lane-changing trajectories. Therefore, we use the lane-changing duration time to represent the diversity of lane-changing trajectories, and analyze the diversity of the lane-changing samples based on FCM. The FCM algorithm [37] is similar to K-means, both being clustering analysis methods based on partition. However, compared with the hard partition of K-means, FCM is an advanced soft-clustering method, which uses the membership degree to determine the degree to which each data point belongs to a certain cluster, and the FCM clustering results are more consistent with the objective reality.

The key to the FCM clustering algorithm is to determine the number of clusters C in advance. In order to reasonably characterize the diversity of lane-changing, it is necessary to divide the left-lane-changing samples into an appropriate number of categories. We use the experimental method to determine the number of clusters. This method involves taking the number of clusters C to be 1-10, using the silhouette coefficient to represent the clustering effect, carrying out the cluster test, and finally selecting the number of categories with the largest silhouette coefficient as the most appropriate number of clusters. Fig.21 shows the curve between the silhouette coefficient and the number of clusters C. It can be seen that, when the lane-changing duration times of all the samples are clustered into three groups, the silhouette coefficient is the largest, at about 0.7103, making the optimal number of clusters three.

The left lane-changing samples are clustered into three categories with different lane-changing duration times, and the clustering results are shown in Fig.22. The horizontal axis represents the sample number, the vertical axis represents the lane-changing duration time of the lane-changing samples, and the different colors represent the different lane-changing categories.
Based on the FCM clustering results, we analyze the diversity of the lane changes from two aspects: lane-changing trajectory and lane-changing preparation stage.

First, we analyze the lane-changing trajectory. The different types of lane-changing samples have different lane-changing duration times. The number of samples and lane-changing duration times of each type are shown in Table 7.

In order to further observe the lane-changing trajectory, the lateral offset of the three types of lane-changing samples is visualized in Fig.23. In the figure, the horizontal axis represents the sequence of sample points, the vertical axis the lateral offset of the lane-changing vehicles, and the different colors the categories of lane-changing samples. It can be seen that the lane-changing trajectories of the three types of lane-change samples are significantly different.

Second, we analyze the lane-change preparation stage. Based on the lane-change key points T1 and T2, we can obtain the length of the lane-change preparation stage in the three types of lane-changing sample. The frequency distribution histogram is shown in Fig.24. We can see that the lane-changing preparation time of the different types of lane-change behavior are different.

In particular, it is found that the length of the lane-change preparation stage is linearly related to the lane-change duration time, as shown in Fig.25. The longer is the lane-changing duration time, the longer is the lane-change preparation stage.

We can explain that driving style factors affect the whole lane-changing process by influencing each stage, such as the lane-changing preparation stage.

3) FITTING OF PROTOTYPE TRAJECTORY
In order to further analyze the influence of lane-changing duration time on lane-changing trajectory, we perform a statistical analysis on the three kinds of lane-changing samples based on a box graph. Based on the mean curve method, three prototype trajectories are fitted for the three kinds of lane-changing samples. Each prototype trajectory represents a typical trajectory generated under the lane-changing duration time.

The fitting process of the prototype trajectory is divided into two steps. First, the trend of the lane offset is visualized based on the box plot, which involves three types of lane-changing samples, as shown in Fig.26. In the figure, the horizontal axis represents the sampling point of the lane-changing trajectory, from T1 to T4, while the vertical axis represents the lane offset of the lane-changing vehicle.

At every moment of the lane-changing process, the lane offset of the same kind of sample lane change is represented by a box plot, so that the lateral motion trajectory cluster of lane-changing vehicles can be obtained. The green, blue and red colors in the graph represent the three types of lane offset trajectory with different lane-changing duration times. Secondly, three prototype trajectories are fitted based on the mean curve method, as shown in Fig.27. At each time, different lane-changing samples have different lane offsets, and the mean of each kind of sample lane change roughly
reflects the average lane offset in the corresponding class. Therefore, we can obtain three prototype trajectories of lane changes based on the mean curve method, which represent the trend of the lateral trajectory under different lane-changing duration times.

B. FRAMEWORK OF DIVERSE LANE-CHANGING TRAJECTORY PREDICTION

Based on the above analysis of lane-changing diversity, it is found that the lane-changing samples can be clustered into three categories with different duration times, and they have different lane-changing trajectories, lengths of lane-changing stages and lateral offsets. Therefore, when predicting the lane-changing trajectories of other vehicles, it is necessary to consider their diversity, which is caused by different lane-changing duration times. In this section, based on the encoder-decoder structure and LSTM neural network, we propose a trajectory prediction model considering the diversity of lane-changing behavior caused by driving styles. Different from the traditional LSTM encoder-decoder structure, the proposed method establishes three LSTM encoder-decoder models. Each model can predict the lane-changing trajectory under a certain lane change duration time, and thereby achieve a variety of lane-changing trajectories that represent different driving styles.

1) LSTM-BASED ENCODER-DECODER FRAMEWORK

In this paper, a trajectory prediction framework considering lane-changing duration is proposed, which is based on the LSTM-based encoder-decoder framework. LSTM was first proposed by Hochreiter and Schmidhub in 1997 [38]. After years of efforts of many generations of scholars, a relatively systematic and complete LSTM framework was finally formed. The specific structure is shown in Fig.28. LSTM improves RNN by introducing a “gate control” structure, and solves the problem of RNN producing gradient disappearance in long-term prediction to a certain extent. The “gate control” structure can combine the short-term memory and long-term memory of time series, which gives a strong information-mining ability and a deep representation ability. The gate structure of LSTM consists of three parts: the forgetting gate, input gate and output gate. See the part marked by a red box in Fig.28.

Based on the LSTM neural network, we build the model for predicting diverse lane-changing trajectories as shown in Fig.29. On the left is a common trajectory prediction model, which does not consider the driving styles. On the right is the trajectory prediction model proposed in this paper, which does consider the driving styles.

In the structure of encoder-decoder trajectory prediction, the working process is as follows: The encoder encodes the predicted vehicle’s historical trajectory information into a context vector, which contains the encoder’s understanding and memory of the historical trajectory. The decoder extracts important information from the context vector and generates the future position of the predicted vehicle. The classical trajectory prediction method is to use the encoder-decoder, which thinks that all lane-changing processes follow only one lane-changing rule, and this general lane-changing rule can be summarized to guide the decoding model to generate a deterministic trajectory. However, the analysis of the lane-changing samples shows that the lane-changing process is diverse. Therefore, in the proposed trajectory prediction framework, by using the lane-changing trajectory encoder-decoder, we summarize three kinds of lane-changing rules based on the three types of lane change samples with different
Y.-Y. Ren et al.: Method for Predicting Diverse Lane-Changing Trajectories of Surrounding Vehicles

FIGURE 29. Framework for predicting diverse lane-changing trajectories. The left represents the LSTM trajectory prediction framework without considering driving styles. The right represents the trajectory prediction method proposed in this paper, including three models similar to the left side, which represent the trajectory prediction models under three patterns of lane-changing, so as to predict the diverse lane-changing trajectories.

durations. Then, three possible lane-changing trajectories can be predicted based on the historical vehicle trajectories, which represents the diversity of the lane-changing process.

We used the single-layer LSTM cell in the encoder-decoder model with a fixed hidden state dimension of 15. SeLU was used as the activation function for the LSTM cell. A batch size of 128 was used and the network was trained for 2000 epochs using the Adam optimizer with an initial learning rate of 0.001. The proposed LSTM model was trained on a single GTX1650 GPU with a Tensorflow-2.3 implementation.

C. PREDICTING DIVERSE LANE-CHANGING TRAJECTORIES OF SURROUNDING VEHICLES BASED ON MULTI-LSTM

Based on the framework for predicting diverse lane-changing trajectories proposed in this paper, the NGSIM dataset is used to design input variables and train the lane-changing trajectory prediction model with multiple LSTM.

1) DESIGN OF INPUT VARIABLES

The input variables of the lane-changing trajectory prediction model and the control model are designed as shown in Table 8. The historical trajectory of the other vehicle \((x_{T_0}, y_{T_0})\) is taken as the input variable of the trajectory prediction model, and the output is the future trajectory \((x_{T_f}, y_{T_f})\) of the other vehicle. Because the lane-changing durations of the three types of lane-changing samples are different, in order to ensure that the lane-changing trajectory prediction model can learn the rules of the three types of lane-changing trajectories, the history series \(T_p\) and prediction series \(T_f\) of the LSTM network are set to different lengths to represent the three types. The value of \(T_p\) is based on the length of the lane-changing preparation stage obtained from the above analysis.

2) MODEL TRAINING

In order to make full use of the sample lane change, the sliding-window method is used to intercept the lane-changing trajectory segments, and thus train the lane-changing trajectory prediction model. In the three types of lane-changing trajectories, the difference in durations leads to a difference in sample length. Therefore, it is very important to choose an appropriate width of sliding window. The width of the sliding window is defined as the sum of the input and output sequences of the model, and the sliding step size is 1. In this way, the three types of lane-changing samples can be intercepted separately. 80% of them are taken as the training set, the remaining 20% as the test set, and the data are standardized to facilitate the neural network training.

D. VALIDATION OF TRAJECTORY PREDICTION PERFORMANCE

In order to validate the trajectory prediction effect of the proposed model, we analyze it from three aspects: trajectory prediction error, diversity of trajectory prediction results and comparison with the classical model.

1) TRAJECTORY PREDICTION ERROR ANALYSIS

Firstly, the prediction accuracy of the trajectory prediction model is evaluated based on the trajectory prediction evaluation index.

Because the trajectory prediction method in this paper considers the diversity of lane changing, the output of the model is multiple trajectories. Different from the prediction error analysis of a single trajectory, the analysis of multiple trajectory prediction results has a special evaluation index. In this paper, \(\text{minFDE}\) (minimum final displacement error) and \(\text{minRMSE}\) (minimum root mean square error) are used to evaluate the prediction performance of the diversity trajectory prediction model. The \(\text{minFDE}\) is the minimum value of the final displacement error between the predicted trajectory and the actual trajectory, as shown in formula (13). The calculation process of \(\text{minRMSE}\) is divided into two steps. First, find the trajectory with minimum final displacement error between the predicted trajectory and the actual trajectory, then calculate the \(\text{RMSE}\) of the trajectory, and the result is the \(\text{minRMSE}\). The calculation method is shown in formula (14).

\[
\text{minFDE} = \min(\abs{O_{T_{\text{true}}}} - O_{T_{\text{true}}}), \quad i = 1, 2, 3 \quad (13)
\]
TABLE 8. Model input and output parameter design.

|                      | Multiple LSTM-based model considering lane change duration time | Single LSTM-based model |
|----------------------|---------------------------------------------------------------|-------------------------|
|                      | LSTM-LC1 | LSTM_LC2 | LSTM_LC3 | LSTM     |
| Samples              | 169      | 406      | 428      | 334      |
| Trajectory segments  | 1690     | 1899     | 1629     | 1956     |
| Input length $T_p$   | 40       | 30       | 20       | 40       |
| Output length $T_r$  | 50       | 40       | 30       | 50       |
| Input variable       | Vehicle lateral and longitudinal position $(x, y)$ |
| Output variable      | Vehicle lateral and longitudinal position $(x, y)$ |

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (O_{pre}^t - O_{true}^t)^2}$$  \hspace{1cm} (14)$$

where $n$ is the length of $T_{pre}$.

The test set was used to evaluate the lane-changing trajectory prediction error, and the resulting minFDE and minRMSE are shown in Fig.30. Fig.30(a) and Fig.30(b) respectively show the prediction errors of the lateral position and the longitudinal position in the next 1-5s. The first two columns are the FDE and RMSE of the three predicted trajectories, and the last column is the minFDE and minRMSE. The results show that the proposed lane-changing trajectory prediction model can ensure high prediction accuracy over a short time, but with the increase of the prediction time, the prediction error increases gradually.

2) TRAJECTORY PREDICTION DIVERSITY ANALYSIS

Secondly, the diversity of the trajectory prediction results is analyzed.

The prediction method of diverse lane-changing trajectories proposed in this paper not only achieves trajectory prediction accuracy, but also obtains the diverse lane-changing trajectories of the lane-changing vehicles in the future. Fig.31 and Fig.32 give an example diagram of diverse lane-changing trajectories. The yellow area in the figure is the possible lane-changing area in the future, which is composed of three predicted trajectories. For the rear vehicles in the target lane, this area is the danger area due to the possibility of other vehicles cutting in in the future. In the field of autonomous driving, the prediction of danger areas in advance is helpful for the decision-making of autonomous vehicles.

Therefore, the method of predicting diverse lane-changing trajectories proposed in this paper can obtain diverse trajectories, which reflect the randomness and diversity of the lane-changing process, and are closer to the actual driving situation of vehicles. More importantly, it can be applied to the field of autonomous driving, which will be helpful for the advance decision-making and path-planning of autonomous vehicles.

3) MODEL DOMINANCE ANALYSIS

Finally, we perform a comparative analysis of the trajectory prediction results between the proposed trajectory prediction model and the contrast model. Based on the LSTM network with the same super parameters, a trajectory prediction contrast model is established, which does not distinguish the lane-changing samples by lane-changing duration, and one third of the sample of lane changes are randomly selected as a training sample. The input parameters of the model are set as shown in Table 8. The prediction error of the proposed model is compared with that of the contrast model in the next 5 seconds. The prediction errors of the two models based on the same test set are shown in Tables 9 and 10 and Fig.33. In the 5s prediction time, compared with the contrast model, the RMSE and FDE of the lateral position of the
FIGURE 31. Schematic diagram of trajectory prediction results. Red car represents the lane-changing vehicle, green car represents the rear vehicle in target lane. Black trajectory is the historical trajectory of lane changing vehicle, the yellow area is the area composed of diverse lane-changing trajectories obtained by the lane changing trajectory prediction method proposed in this paper, which represents the driving area of lane changing vehicles in the target lane in the future.

FIGURE 32. Diverse lane-changing trajectory prediction results. The left figure shows the prediction results of lane-changing trajectory at the early stage of the lane-changing behavior, the right figure shows the prediction results of lane-changing trajectory when the vehicle reach the boundary of the lane.

FIGURE 33. Comparison of prediction results between classical model and model proposed in this paper. (a) represents the lateral position error (RMSE and FDE) under the prediction time of 1–5 s. (b) represents the longitudinal position error (RMSE and FDE) under the prediction time of 1–5 s. It should be noted that the RMSE and FDE of lane changing trajectory obtained by proposed model is the minRMSE and minFDE of the three predicted trajectories.

The proposed trajectory prediction model are reduced by more than 21%, the FDE of the longitudinal position is reduced by more than 24%, and the RMSE is reduced by more than 39%.

The above results show that, compared with the classical LSTM model that does not consider the lane-changing duration, the proposed method reduces the trajectory prediction error and significantly improves the prediction accuracy.
In addition, the diverse lane-changing trajectory prediction model, which is composed of three LSTM trajectory prediction models under three lane-changing patterns, can generate three possible lane-changing trajectories caused by driving style.

VI. CONCLUSION AND DISCUSSION

A. CONCLUSION

In this paper, a lane-changing trajectory prediction method based on early lane-changing behavior identification is proposed, which mainly includes behavior identification module and trajectory prediction module. Firstly, the future behavior of the surrounding vehicle is identified. Secondly, the diverse lane-changing trajectories can be obtained through the trajectory prediction module, which considers the diversity of driving styles.

In the lane-changing behavior recognition part, firstly, a method of extracting lane-changing process automatically is proposed based on unsupervised clustering, which divides the key points of the lane-changing process automatically according to the rules of the lane-changing trajectory, and then extracts the lane change stage. This method can reduce the subjective influence of artificially defined rules. Secondly, a group of feature parameters is constructed, and then parameters with high importance and low cross-correlation are selected to represent the diverse lane-changing behaviors of surrounding vehicles. Then, a model for recognizing other vehicles’ lane-changing behaviors is established, and the NGSIM dataset is used to train the model, and evaluate the accuracy and timeliness of lane-changing behavior recognition. The results show that the recognition accuracy is 88.96%, the accuracy of surrounding vehicle behavior detection in 2s before lane change point is above 95%, the average anticipation time is 3.12s before lane change execute, and the anticipation time as a proportion of the lane-changing duration is average 44.13%. Compared with the recognition results in other studies, the proposed lane-changing behavior recognition model has better recognition performance.

Based on the lane-changing behavior recognition results, a diverse lane-changing trajectories prediction model considering the diversity of driving styles is proposed. Firstly, the lane-changing duration time is taken as the representative feature of the diverse lane-changing process under different driving styles, and the unsupervised clustering method is used to analyze the diversity of the lane-changing samples. Then, three kinds of lane-changing samples, with significant differences in their lane-changing durations, are obtained. Further, based on the mean curve method, the three kinds of lane-changing samples are fitted to three prototype trajectories. Secondly, based on the encoder-decoder framework commonly used in trajectory prediction, three LSTM lane-changing trajectory prediction models are established, which represent the diverse lane-changing process under different driving styles. The model is also trained and validated using the NGSIM public dataset. The results show that the prediction error is significantly reduced by the proposed method. Compared with the classical LSTM model, the prediction accuracy of the proposed model is improved by more than 21%. The trajectory prediction method can obtain three predicted trajectories and the influence area of the lane-changing vehicle. The three predicted trajectories represent the diversity of lane-changing behavior caused by driving styles, while the prediction area represents the danger area caused by other vehicles cutting into the target lane, which can aid early decision-making and path-planning for rear vehicles in the target lane.

In conclusion, the proposed lane-changing trajectory prediction method uses the lane-changing duration to represent the diversity of lane-changing, which is caused by the driver’s style, the surrounding environment and so on. The lane-changing trajectory obtained is consistent with the actual lane-changing process of vehicles. At the same time, the diverse lane-changing trajectories form a reference danger area for vehicles that are behind in the target lane, caused by the lane-changing behavior of surrounding vehicles. In the field of autonomous driving, this method can help autonomous vehicles to make more accurate and comprehensive driving decisions in advance, when faced with surrounding vehicles cutting into their lane. In addition, the lane-changing trajectory prediction method considering...
driving styles proposed in this paper can provide a reference for researchers using deterministic trajectory prediction models (such as LSTM). It shows that the combination of deterministic models can be used to predict lane-changing trajectories under diverse driving styles, and make the trajectory diversity more interpretable.

In the practical implementation, the proposed lane-changing trajectory prediction method is used for autonomous vehicles on the premise that AV vehicles can obtain the driving information of surrounding vehicles in real time. That is, the behavior of surrounding vehicles (such as lane-keeping, lane-changing) is judged by the behavior identification model in real time. If it is detected that the surrounding vehicles are about to change lanes, the diverse lane-changing trajectories prediction model is triggered to predict their future lane-changing trajectory. At each sampling point, the input historical trajectory information is updated to obtain the latest predicted trajectory.

B. DISCUSSION AND FUTURE WORKS

This research process of this paper has the following shortcomings:

1) There are two challenges in trajectory prediction: one is the modeling of the bidirectional interaction between the lane-changing vehicle and surrounding vehicles, while the other is the modeling of lane-changing trajectory diversity. This paper only analyzes and studies the latter, without considering the bidirectional interaction between the lane-changing vehicle and surrounding vehicles, which may affect the accuracy of our trajectory prediction.

2) In order to obtain accurate and reliable future behavior of other vehicles, in this article, we achieved the early detection of lane changing behavior of surrounding vehicles. However, before the vehicle’s lateral position changes, the lane changing intention of the vehicle may be inferred from the relative speed and distance of surrounding vehicles and their behaviors.

3) In the trajectory prediction part, the input variable of the model is the historical trajectory (x, y) of the predicted vehicle, with no consideration of the other driving state parameters of the lane-changing vehicle or the relationship between it and the surrounding vehicles. There is a lack of analysis of the influencing mechanism between the input parameters and the trajectory prediction effect.

Therefore, future works will include early intention reasoning of surrounding vehicles by analyzing the influencing factors of lane-changing decision and vehicles interaction by adding other features related to vehicle neighborhood or using structural LSTMs.

REFERENCES

[1] D. Helbing and P. Molnar, “Social force model for pedestrian dynamics,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 51, no. 5, p. 4282, 1995.

[2] A. Houenou, P. Bonnifait, V. Cherfaoui, and W. Yao, “Vehicle trajectory prediction based on motion model and maneuver recognition,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Nov. 2013, pp. 4365–4369.

[3] J. Ziegler, P. Bender, and M. Schreiber, “Making bertha drive—An autonomous journey on a historic route,” IEEE Intell. Transp. Syst. Mag., vol. 6, no. 2, pp. 8–20, Summer 2014.

[4] S. Lefèvre, D. Vasquez, and C. Laugier, “A survey on motion prediction and risk assessment for intelligent vehicles,” Robomech J., vol. 1, no. 1, pp. 1–14, Dec. 2014.

[5] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, “Social LSTM: Human trajectory prediction in crowded spaces,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 961–971.

[6] N. Lee, W. Choi, P. Varnaza, C. B. Choy, P. H. S. Torr, and M. Chandraker, “DESIRE: Distant future prediction in dynamic scenes with interacting agents,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 336–345.

[7] M. Bansal, A. Krizhevsky, and A. Ogale, “Chauffeurnet: Learning to drive by imitating the best and synthesizing the worst,” 2018, arXiv:1812.03079.

[8] S. Choi, J. Kim, and H. Yeo, “Attention-based recurrent neural network for urban vehicle trajectory prediction,” Proc. Comput. Sci., vol. 151, pp. 327–334, Jan. 2019.

[9] A. Benterki, V. Judalet, M. Choubeila, and M. Bouchkifne, “Long-term prediction of vehicle trajectory using recurrent neural networks,” in Proc. 45th Annu. Conf. IEEE Ind. Electron. Soc. (IECON), Oct. 2019, pp. 3817–3822.

[10] A. Benterki, M. Bouchkifne, V. Judalet, and C. Mmaoui, “Artificial intelligence for vehicle behavior anticipation: Hybrid approach based on maneuver classification and trajectory prediction,” IEEE Access, vol. 8, pp. 56992–57002, 2020.

[11] X. N. Zhang, “Research on vehicle trajectory prediction based on LSTM network,” Automobilie Technol., vol. 45, no. 22, pp. 32–33, 2020.

[12] X. Ji, C. Fei, X. He, Y. Liu, and Y. Liu, “Intention recognition and trajectory prediction for vehicles using LSTM network,” China J. Highway Transp., vol. 32, no. 6, pp. 34–42, Jun. 2019.

[13] H. Y. Wen, W. G. Zhang, and S. Zhao, “Vehicle lane change trajectory prediction model based on generative adversarial networks,” J. South China Univ. Technol. (Natural Sci. Ed.), vol. 48, no. 5, pp. 32–40, 2020.

[14] H. Hu and T. Xie, “Lane change trajectory prediction of vehicles in highway interweaving area using Seq2Seq-attention Network,” China J. Highway Transp., vol. 33, no. 6, pp. 106–118, 2020.

[15] L. Xin et al., “Intention-aware long horizon trajectory prediction of surrounding vehicles using dual LSTM networks,” in Proc. IEEE Int. Conf. Intell. Transp. Syst. (ITSC). IEEE, 2018.

[16] M. Schreier, V. Willert, and J. Adamy, “An integrated approach to maneuver-based trajectory prediction and criticality assessment in arbitrary road environments,” IEEE Trans. Intell. Transp. Syst., vol. 17, no. 10, pp. 2751–2766, Oct. 2016.

[17] W. M. Xu, “Research on lane-changing trajectory prediction based on driver’s driving intention,” M.S. thesis, Dept. Vehicle. Eng., Hunan Univ., Hunan, China, 2015.

[18] C. Zhang, J. Zhu, W. Wang, and J. Xi, “Spatiotemporal learning of multivehicle lane-change patterns in lane-change scenarios,” IEEE Trans. Intell. Transp. Syst., early access, Feb. 18, 2021, doi: 10.1109/TITS.2021.3057645.

[19] X. Jia, L. Sun, M. Tomizuka, and W. Zhan, “IDE-Net: Interactive driving event and pattern extraction from human data,” IEEE Robot. Autom. Lett., vol. 6, no. 2, pp. 3065–3072, Apr. 2021.

[20] L. Hou, L. Xin, S. E. Li, B. Cheng, and W. Wang, “Interactive trajectory prediction of surrounding road users for autonomous driving using structural-LSTM network,” IEEE Trans. Intell. Transp. Syst., vol. 21, no. 11, pp. 4615–4625, Nov. 2020.

[21] S. Ammoun and F. Nashashibi, “Real time trajectory prediction for collision risk estimation between vehicles,” in Proc. IEEE 5th Int. Conf. Intell. Comput. Commun. Process., Aug. 2009, pp. 417–422.

[22] A. Broadhurst, S. Baker, and T. Kanade, “Monte Carlo road safety reasoning,” in Proc. IEEE Intell. Vehicles Symp., Jun. 2005, pp. 319–324.

[23] S. Anev, G. Miller, and N. Papanikolopoulos, “Clustering of vehicle trajectories,” IEEE Trans. Intell. Transp. Syst., vol. 11, no. 3, pp. 647–657, Sep. 2010.

[24] C. Hermes, C. Wohrer, K. Schenk, and F. Kummert, “Long-term vehicle motion prediction,” in Proc. IEEE Intell. Vehicles Symp., Jun. 2009, pp. 652–657.

[25] D. Augustin, M. Hofmann, and U. Konigorski, “Prediction of highway lane changes based on prototype trajectories,” Forschung im Ingenieurwesen, vol. 83, no. 2, pp. 149–161, Jan. 2019.
[26] US Department of Transportation. (2008). NGSIM: Next Generation Simulation. [Online]. Available: http://www.ngsim.fhwa.dot.gov

[27] T. Christian, M. Treiber, and A. Kesting, “Estimating acceleration and lane-changing dynamics from next generation simulation trajectory data,” Transp. Res. Rec., J. Transp. Res. Board, vol. 2088, pp. 90–101, Dec. 2008.

[28] W. Xuesong, Y. Minming, and D. Hurwitz, “Analysis of cut-in behavior based on naturalistic driving data,” Accident Anal. Prevention, vol. 124, no. 1, pp. 127–137, Mar. 2019.

[29] R. D Worrall and A. Bullen, “An empirical analysis of lane change on multilane highways,” Highway Res. Rec., vol. 9, no. 303, pp. 30–43, 1970.

[30] Y. Guo, Q. Sun, R. Fu, and C. Wang, “Improved car-following strategy based on merging behavior prediction of adjacent vehicle from naturalistic driving data,” IEEE Access, vol. 7, pp. 44258–44268, 2019, doi: 10.1109/ACCESS.2019.2908422.

[31] X. Li, W. Wang, and M. Roetting, “Estimating driver’s lane-change intent considering driving style and contextual traffic,” IEEE Trans. Intell. Transp. Syst., vol. 20, no. 9, pp. 3258–3271, Sep. 2019.

[32] Y.-Y. Ren et al.: Method for Predicting Diverse Lane-Changing Trajectories of Surrounding Vehicles

[33] Z. Q. Bai, “The study on feature engineering construction and application based on user behavior,” M.S. thesis, Dept. Probab. Statist., South China Univ. Technol., Guangzhou, China, 2018.

[34] L. I. Ling, “Multi-label feature selection algorithm based on information gain,” Comput. Sci., vol. 42, no. 7, pp. 52–56, 2015.

[35] W. Huan, “Feature selection model for harmfulness prediction of clone code,” J. Comput. Appl., vol. 37, no. 4, pp. 1135–1142, 2017.

[36] X. J. Lu, “Fuzzy association rules mining from time series based on FCM clustering,” J. Dalian Univ. Technol., vol. 50, no. 5, pp. 806–810, 2010.

[37] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.