CHALMERS
UNIVERSITY OF TECHNOLOGY

Spatiotemporal Interaction Pattern Recognition and Risk Evolution Analysis During Lane Changes

Accepted by IEEE Transactions on Intelligent Transportation Systems

Spatiotemporal Interaction Pattern Recognition and Risk Evolution Analysis during Lane Changes

Yue Zhang, Yajie Zou, Selpi, Yunlong Zhang, and Lingtao Wu

***Abstract*—In complex lane change (LC) scenarios, semantic interpretation and safety analysis of dynamic interaction pattern are necessary for autonomous vehicles to make appropriate decisions. This study proposes a learning framework that combines primitive-based interaction pattern recognition and risk analysis. The Hidden Markov Model with the Gaussian mixture model (GMM-HMM) approach is developed to decompose the LC scenarios into primitives. Then K-means clustering with Dynamic Time Warping (DTW) is applied to gather the primitives into 13 LC interaction patterns. Finally, this study considers time-to-collision (TTC) of two conflict types involved in the LC process. And the TTC is used to analyze the risk of interaction patterns and extract high-risk LC interaction patterns. The LC events obtained from the Highway Drone Dataset (highD) demonstrate that the identified LC interaction patterns contain interpretable semantic information. This study identifies the dynamic spatiotemporal characteristics and risk formation mechanism of the LC interaction patterns. The findings are useful to comprehensively understand the latent interaction patterns, which can then be used to design and improve the decision-making process during lane changes and enhance the safety of autonomous vehicle.**

***Index Terms*—Lane change, Interaction pattern, Traffic risk, Driving primitive**

I. Introduction

LANE change (LC) is a daily-routine but challenging driving task which involves vehicle movements in both longitudinal and lateral directions and close interactions between multiple vehicles. Crash data in the U.S. from 2010 to 2017 show that the sudden lane change caused about 17.0% of total severe crashes [1, 2]. With the advantages of perception and information acquisition, autonomous vehicles can significantly reduce traffic crashes. However, how autonomous vehicles make reasonable decisions in complex LC scenarios is a major challenge at present [3, 4]. On the one hand, human behavior is heterogeneous and stochastic. On the other hand, the behavior of each vehicle in the LC scenario is dynamically affected by surrounding vehicles. Therefore, it is necessary to understand the multi-vehicle interaction behavior patterns and reveal the interaction mechanism to facilitate autonomous vehicles to make reasonable decisions.

As one of the most important microscopic behaviors of traffic flow, LC behavior has been analyzed in many studies. For example, Woo et al. [5] defined four LC phases according to the relationship between the vehicle and the center line and other features, namely keeping, changing, arrival and adjustment. The study assumed that the LC behavior is not affected by surrounding vehicles. In order to improve the accuracy of the LC behavior simulation, some studies examined the impact of surrounding environment on the subject vehicle. For example, Gipps [6] considered adding safety gap and presence of heavy vehicles to the LC decision. Later, various factors indicating surrounding vehicles were considered, such as relative speed [7, 8], gap acceptance [9], and turning signal [10], etc. However, these studies only focused on the impact of other vehicles on the subject vehicle, but ignored the impact of the subject vehicle on surrounding vehicles. However, the interaction between vehicles plays an important role in the decision-making process. In order to consider the interaction of vehicles in LC scenarios, some studies classified LC scenarios based on prior knowledge. Hidas [9] and Zhang et al. [11] categorized LC into three classes: free, forced, and cooperative lane changing [12]. Halati, Lieu and Walker [13] classified LC maneuvers into three: mandatory lane changes, discretionary lane changes, and random lane changes. Due to the uncertainty of behaviors, the multi-vehicle interaction patterns are dynamically changing. To capture the impact of traffic context, Wang et al. [14] proposed a real-time multi-vehicle collaborative learning approach to model spatial and temporal information among multiple vehicles. Other researchers introduced game theory to simulate the LC interaction behaviors [15, 16]. However, these real-time decision-making methods heavily rely on the sensor data, and it is difficult to explain the mechanism of interaction in this way, resulting in incomprehensible decisions. Therefore, it is necessary to identify the interaction patterns as a prior information for

This research was funded by the National Natural Science Foundation of China (Grant No. 71971160), the Shanghai Science and Technology Committee (Grant No. 19210745700) and the Fundamental Research Funds for the Central Universities (Grant No. 22120220013). (Corresponding author: Yajie Zou)

Yue Zhang is with key Laboratory of Road and Traffic Engineering of Ministry of Education, Tongji University, No. 4800 Cao'an Road, Shanghai 201804, China (e-mail: zhangyue18@tongji.edu.cn).

Yajie Zou is with key Laboratory of Road and Traffic Engineering of Ministry of Education, Tongji University, No. 4800 Cao'an Road, Shanghai 201804, China (e-mail:yajiezou@hotmail.com).

Selpi is with the Department of Computer Science and Engineering, Chalmers University of Technology, SE-412 96 Goteborg, Sweden (e-mail: selpi@chalmers.se).

Yunlong Zhang is with Zachry Department of Civil Engineering, Texas A&M University, 3136 TAMU, College Station, TX 77843-3136, United States (e-mail: yzhang@civil.tamu.edu).

Lingtao Wu is with Texas A&M Transportation Institute, Texas A&M University System, 3135 TAMU College Station, Texas 77843-3135, USA (e-mail: wulingtao@gmail.com).

decision-making in advance.

Because the LC scenarios have complex and high-dimensional features, it is challenging to extract interaction patterns from it. Some researchers addressed this issue by decomposing the entire scenarios into finite semantic segments, such as [17]. These studies are based on the view that human activity should be decomposed into building blocks which belong to an "alphabet" of elementary actions [18]. Bregler [19] found partial motions were always the same and repeatable segments of trajectoryin periodic or stereotypical motions. Ding, Xu, Tu, Zhao, Moze, Aioun and Guillemard [20] also believed that drivers share a pool of driving states. Therefore, it is assumed that the lane changing interaction scenario can also be decomposed into a finite number of interaction patterns that are heterogeneous and can be shared. The Gaussian mixture model (GMM) is commonly used to study the heterogeneity in data. For instance, the GMM was applied to split the intersection encounter scenarios into discrete segments [21, 22]. The dynamic transition between these discrete segments is described by stochastic processes such as the Markov model [23, 24]. This solution is conducive to improve learning performance from massive data [25]. Zhang et al. [26] proposed a primitive-based framework to learn interaction patterns by segmenting the LC scenarios. However, the study did not analyze the risk of interaction patterns, which may lead to autonomous vehicles learning improper operation and making dangerous decisions.

In recent years, surrogate safety measures have been widely used to quantify the risk of micro-behavior and severity of interaction. Saunier et al. [27] used some surrogate safety indicators for the analysis of spatial interactions between vehicles at intersections, and interactions are classified into four categories: head-on, rear-end, side and parallel. Common surrogate safety indicators included Time-to-collision (TTC), Post-Encroachment Time (PET), Gap Time (GT), and Deceleration-to-Safety Time (DST) [28]. Among these indicators, TTC is one of the most commonly applied indicators for risk analysis [29-32]. TTC refers to the time for a vehicle to collide with the preceding one without changing its current direction and speed. Wu et al. [33] recognized real-time LC risk level based on TTC. According to the results of Laureshyn et al. [34], the risk in an interaction scenario was not constant. For example, for a dangerous interaction scenario, some time periods may have high risk level, while some time periods may not have risk. The decomposition method also focuses on the most dangerous segment (interaction pattern) in a LC interaction scenario.

According to the previous studies, interaction behavior plays an important role in making reasonable lane changing decisions for vehicles. Exploring different interaction patterns not only is useful for developing heterogeneous lane changing models, but also is one of the essential elements to build a driving scenario library for autonomous vehicle test. Furthermore, identifying high-risk interaction patterns among many interaction patterns is conducive to enhance the safety of driving decisions. However, the existing classification methods based on interaction (e.g., classifying lane changes into free, synchronized, and cooperative lane changes [35]) are subjective, ambiguous and rough [20]. In addition, complex and high-dimensional interaction scenarios are difficult to understand. The lack of consideration of dynamic interaction mechanism results in inaccurate understanding of lane changing behavior for autonomous vehicles [24]**.** Motivated by these two issues, this study proposes a framework based on the semantic decomposition method, including the interaction pattern recognition, semantic interpretation and risk analysis. The contributions of this study are:

1. A learning framework is presented to integrate primitive-based interaction pattern recognition and risk analysis, which can automatically extract a finite number of heterogeneous and interpretive spatiotemporal interaction patterns between vehicles existing in massive data.

2. The mapping relationship between the LC interaction scenarios and the real world is established to form an interaction pattern pool to enrich the driving scenario library. To understand the human driving behavior, the spatiotemporal evolution mechanism of the interaction patterns is analyzed.

3. Different interaction types between two vehicles are considered to calculate TTC and identify high-risk LC interaction patterns.

The main advantages of the proposed framework are:

1. The method proposed in this study can identify the dynamic spatiotemporal interaction pattern between the vehicles involve in a lane change process. The method requires low computational cost and no manual scene annotation.

2. The 13 interaction patterns obtained from this research can clearly explain the complicated LC interaction scenarios using the highD data. The proposed modeling framework can provide important information for understanding the human driving behavior and identifying the high-risk lane changing interaction patterns, which can promote the decision-making of autonomous vehicles.

The rest of this paper is organized as follows: Section 2 introduces the methodology used in this study. Section 3 documents the data collection and preprocessing. Section 4 discusses the learning process of models and analyzes the results of the proposed methods. Section 5 provides the conclusion and future work of this study.

II. METHODOLOGY

The framework proposed in this paper is shown in Figure 1. Firstly, specific LC scenarios are extracted from naturalistic driving data. Then the primitives are segmented and clustered to obtain interaction patterns. Finally, semantic explanation and risk analysis of the interaction pattern are conducted.

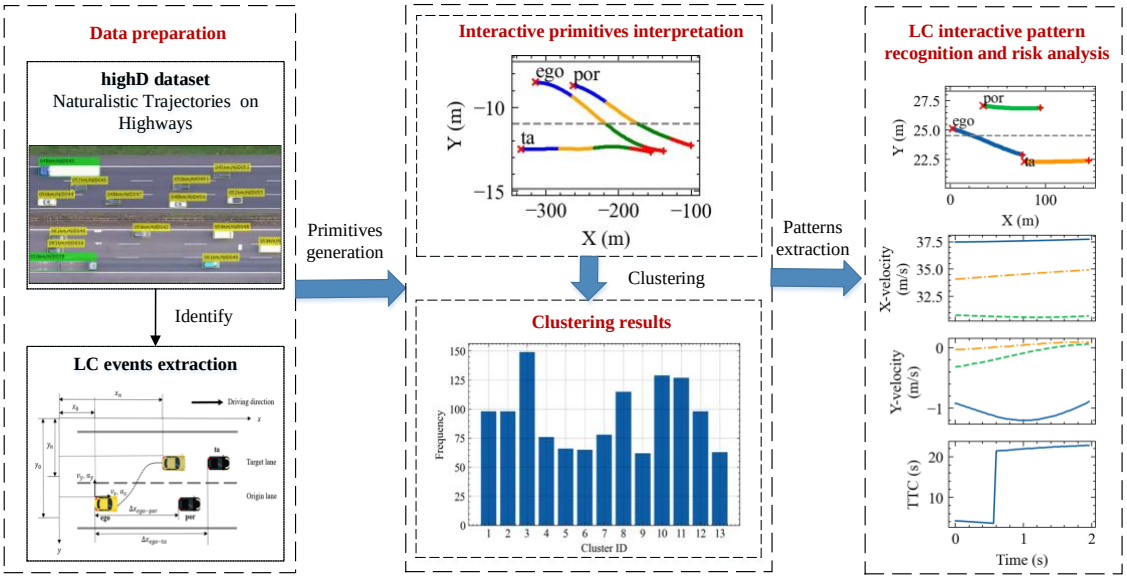


Figure 1 Framework of LC interaction patterns recognition and risk analysis

A. *LC scenario and its primitives*

The sequential LC event data includes the trajectories of three vehicles (i.e., the LC vehicle, the preceding vehicle in the original lane and the vehicle in the target lane), expressed as

$$O = \{o_1, o_2, \cdots o_t, \cdots o_T\} \tag{1}$$

O can be called a LC interaction scenario, where $o_t = \left\{x_t^{(1)}, y_t^{(1)}, x_t^{(2)}, y_t^{(2)}, x_t^{(3)}, y_t^{(3)}\right\} \in \mathbb{R}^6$. $x_t^{(1)}$, $x_t^{(2)}$ and $x_t^{(3)}$ are the longitudinal position of three vehicles at time *t*, respectively. $y_t^{(1)}, y_t^{(2)}$ and $y_t^{(3)}$ are the lateral position of three vehicles at time *t*. *T* represents the time length of a LC scenario *O* (or the duration of a LC event), and *t* represents any timestamp during LC. The primitive of *O* is formulated as Eq. (2).

$$P_i = \{p_m, \cdots p_n\} \qquad (1 \le m \le n \le T) \tag{2}$$

Where $P_i \subseteq R$, *i* is the *i*-th primitive.

B. *Segment LC interaction primitives*

LC interactions are dynamic process with stochastic human behavior. To analyze the complex scenarios with high-dimensional data, the Hidden Markov Model with the Gaussian mixture model (GMM-HMM) is adopted to decompose scenarios into semantic primitives. The GMM-HMM has been widely used in natural language processing (NLP) and automatic speech recognition (ASR) to process sequence data and mine the hidden states such as phonemes. The HMM in the GMM-HMM has a strong advantage in modeling dynamic behavior. In order to establish the relationship between driving scenarios and types of interaction patterns, the GMM is used to model the state-output distribution on the basis of HMM. The mixture component of GMM corresponds to the hidden state in the HMM. The entire process of the model is illustrated in Figure 2. The components of the GMM-HMM are described in detail below.

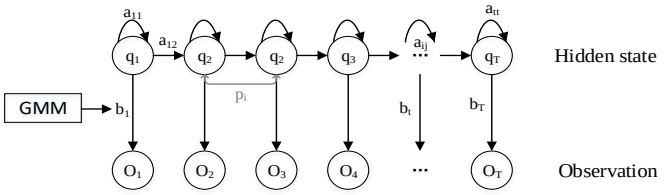


Figure 2 The schematic diagram of the proposed GMM-HMM model.

1) *Gaussian Mixture Model*

The trajectory data in LC interaction scenario are continuous, and the observation is heterogeneous due to the existence of multiple interaction patterns. Due to the strong ability of fitting multiple distributions and mining heterogeneity in data, the GMM is selected to model the emission probability in the HMM so as to establish the relationship between interaction patterns and LC interaction scenarios. The joint probability density function generated by multi-dimensional GMM with *M* components is

$$\begin{aligned} P(o_t) &= \sum_{i=1}^{M} \omega_i N(o_t|\mu_i, \Sigma_i) \\ &= \sum_{i=1}^{M} \omega_i \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \\ &\quad \times exp\left\{-\frac{1}{2}(o_t - \mu_i)^T(\Sigma_i)^{-1}(o_t - \mu_i)\right\} \end{aligned} \tag{3}$$

$$\sum_{i=1}^{N} \omega_i = 1 \tag{4}$$

Where $N(o_t|\mu_i, \Sigma_i)$ is the i-th multivariate Gaussian distribution (dimension d=6) and $o_t = \left\{x_t^{(1)}, y_t^{(1)}, x_t^{(2)}, y_t^{(2)}, x_t^{(3)}, y_t^{(3)}\right\}$ is the observation of the LC interaction scenario at time t. M is the number of components. $\mu_i$is mean of the i-th Gaussian component, and Σ_i is covariance matrix. ω_i is weight of the i-th component. In Eq. (3), μ_i and Σ_i

are the parameters that needs to be estimated in the GMM. The Expectation maximization (EM) algorithm is used for estimating these parameters. The convergence criterion is that the difference between the log likelihood values of two adjacent iteration steps is less than 10^{-10}.

2) Hidden Markov Model and Decoding

In the GMM-HMM, each component of the GMM is regarded as the hidden state of the HMM. The aim is to segment primitives of a given observation sequence O using HMM. Some concepts and symbols involved in the HMM should be defined in advance.

Hidden states: the hidden states are denoted as $S = \{S_1, S_2, \cdots, S_N\}$, where N is the number of hidden states and $q_t \in S$ is the state at time *t*. Although the states are hidden, there are often physical meaning in applications [36]. In this study, the states are described as types of interaction patterns.

Observations: the observable state sequence is $O = \{o_1 \cdots, o_t \cdots, o_T\}$, $o_t = \left\{x_t^{(1)}, y_t^{(1)}, x_t^{(2)}, y_t^{(2)}, x_t^{(3)}, y_t^{(3)}\right\}$ is the observation at time t.

Initial state distribution: $\pi = \{\pi_i\}$, where $\pi_i = P[q_1 = S_i], 1 \le i \le N$. π_i is the probability which the Markov chain will start in state *i*.

Emission probabilities: $B = \{b_j(o_t)\}$ are the sequence of probability of an observation o_t generated from state j. In the GMM-HMM, the emission probability is generated by Eq. (3).

State transition probability distribution: $A = \{a_{ij}\}$ is the state transfer matrix, which represents the probability of transition from state i to state j. And the transfer probability for (i, j) pairs can be expressed as Eq. (5).

$$a_{ij} = P\left(q_{t+1} = S_j | q_t = S_i\right), \quad \text{s.t. } i \ge 1, j \le N \text{ and } \sum_{j=1}^{N} a_{ij} = 1 \ \forall\, i \tag{5}$$

Thus, the HMM can be expressed as $\lambda = (N, T, A, B, \pi)$, where N is a hyper-parameter. As one of the three classical HMM problems, decoding problem is the core of composing LC interaction primitives. The decoding problem is to find the optimal hidden state sequence given the observation sequence O and model λ, which can be formulated as Eq. (6).

$$\gamma_t(i) = P(q_t = S_i | O, \lambda) \tag{6}$$

Based on the forward-backward algorithm, Eq. (6) can be transformed into Eq. (7).

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)} \tag{7}$$

Where $\alpha_t(i)$ accounts for a part of observations $\{o_1, o_2, \cdots o_t\}$ and $\beta_t(i)$ accounts for the residual observations $\{o_{t+1}, o_{t+2}, \cdots o_T\}$ given S_i at time t. $P(O|\lambda)$ is the normalization factor to make sure

$$\sum_{i=1}^{N} \gamma_t(i) = 1 \tag{8}$$

Then most likely state q_t can be solved as Eq. (9).

$$q_t = \underset{1 \le i \le N}{argmax}[\gamma_t(i)], 1 \le t \le T \tag{9}$$

According to the most likely state sequence $Q = \{q_1, q_2 \cdots q_t\}$ obtained by Eq. (9), if the state sequence corresponding to the observation value is continuous and consistent, these observations are segmented into a primitive (because they contain the same semantic information). The Viterbi algorithm [36] is used to find the best hidden state sequence in the GMM-HMM. Viterbi is a dynamic programming method, which is used to search the most probable path by taking the maximum probability value of all possible previous hidden state sequences.

C. Clustering of interaction patterns

Since a large number of primitives are extracted by the GMM-HMM, this study uses a clustering method to separate them into homogeneous interaction patterns.

1) Data scaling and normalization

The length of primitives extracted by the GMM-HMM is different. Before applying the clustering model, these primitives should be scaled into the same length *l*. This study uses linear interpolation for scaling-down and scaling-up, which ensures that the scaled trajectory is similar to the original trajectory. Given the data point p_0 at time t_0 and p_1 at time t_1, the unknown data point at time $t \in (t_0, t_1)$ can be calculated by Eq. (10).

$$o_t = o_0 + (t - t_0)\frac{o_1 - o_0}{t_1 - t_0} \tag{10}$$

Since the dimensions of longitude and latitude of the vehicle trajectory are different, normalization is a necessary step. Therefore, each sample of the input sequences is standardized using Min-Max normalization, so o_t is within a standard range [-1, 1].

2) DTW distance based K-means clustering

The K-means clustering can provide satisfactory results in terms of unsupervised feature learning [37]. The K-means is used to cluster LC interaction patterns. The primitive is a time series composed of the temporal feature of the vehicle trajectory. It is inappropriate to simply calculate the center of a trajectory using the traditional K-means clustering method. Therefore, this study proposes to use the K-means clustering method based on Dynamic Time Warping (DTW) to classify the categories according to the similarity between time series.

DTW is a shape-based similarity measure for sequence data. DTW distance is a length of the optimal alignment between two given primitives which are time-series data. It uses dynamic programming to find an optimal path with a minimum distance between two given time series.

Before clustering, the input primitive $P_i \in \mathbb{R}^{6\times l}$ should be reshaped as $P_i \in \mathbb{R}^{1\times 6l}$.

$$P_i = [o_1, o_2 \cdots o_a, \cdots o_l],\ o_a \in \mathbb{R}^6 \tag{11}$$

Then the DTW distance based K-means method is used to cluster primitives. Assuming all primitives $P = \{P_1, P_2, \cdots P_N\}$ are grouped into K patterns $C = \{C_1, C_2, \cdots C_K\}$ and the center of each cluster is μ_i. The goal of the proposed method is to minimize the within-cluster sum-of-squares by Eq. (12).

$$\lambda_w = min \sum_{i=1}^{K} \sum_{P \in C_i} \|P_i - \mu_i\|^2 \quad (12)$$

Finally, the primitives with similar temporal and spatial features are grouped into a cluster.

D. Interaction pattern risk calculation

In order to assess the risk of each interaction pattern and identify high-risk interaction patterns, a risk calculation method based on TTC is introduced in this section. The TTC at time *t* is defined as the remaining time between two vehicles driving in the current state until a collision. The main reasons that TTC is selected as the risk indicator are: (1) it can intuitively reflect the driver's perceived risk, (2) it exists in all types of interactions, and (3) it evolves continuously during each interaction [38, 39].

In LC scenarios, there are two types of interactions between two vehicles according to the relative position and their lanes (Figure 3). The type A interaction refers to two vehicles driving in the same lane and the speed of the following vehicle is greater than that of the preceding vehicle. The type C interaction is vehicles in different lanes experiencing side interaction when their expected path cross and projected positions overlap. In type C, it is necessary to consider the situation that the vehicles reach the collision point in both the longitudinal and lateral directions at the same time. Therefore, a two-dimension TTC is used as the risk indicator of type C interaction, and its calculation method is based on the work conducted by [34, 38].

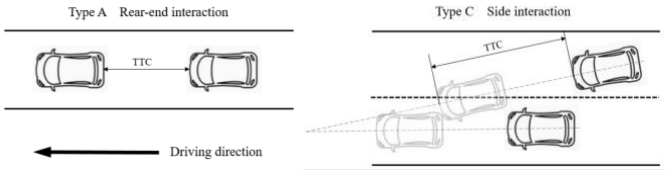


Figure 3 Interaction types between two vehicles in the LC interaction scenario.

After obtaining the TTC of the interaction between two vehicles at each moment, a primitive's risk could be aggregated by TTC values of all interaction pairs at all moments in the primitive, which is used in the *LC interaction pattern risk analysis* section. The risk of every two pairs is represented by a minimum TTC at each moment (one measure per interaction) and the risk of a primitive is represented by the mean TTC of three interactions (one measure per primitive).

III. DATA DESCRIPTION

A. Data collection

The data employed in this study are naturalistic driving trajectory dataset extracted from drone video called highD. It is a large-scale public dataset widely used in traffic flow modelling and safety analysis in recent years [40-44]. Compared with other naturalistic driving datasets, it has the characteristic of high accuracy - the position error does not exceed 10cm [45]. The highD dataset was collected on German freeways around Cologne during 2017 and 2018 [45]. Table 1 shows the information of five basic segments selected in this study. The data are measured at the frequency of 25 Hz.

TABLE 1
DESCRIPTION OF THE SELECTED SEGMENTS

Location ID	Lane width (m)	Segment length (m)	Number of lanes (both direction)
1	4.07; 3.88; 4.15; 4.24; 3.80; 4.15	420	6
2	4.08; 3.84; 3.96; 3.84	420	4
3	3.82; 3.65; 3.73; 3.74; 3.56; 3.91	420	6
4	3.97; 3.63; 3.62; 3.53; 3.80; 3.89	420	6
5	4.09; 3.84; 3.96; 3.97;	420	4

[1] The lane width column shows the width of each lane, separated by ";".

B. Vehicle type classification

Due to different vehicle dynamic performance, the interaction patterns of different vehicle types should also be discussed separately. First of all, the vehicle types need to be classified, based on length and width of the vehicles. When calculating TTC of interaction patterns, vehicle width is also an indispensable factor. For two-dimensional variables, grouping the vehicle type based on threshold values of vehicle length and width may result in unreasonable classification results. For example, some categories may only contain very few observations (e.g., short-wide or long-narrow vehicles). Therefore, this study applies the K-means clustering method to avoid the limitation of fixed threshold values. The number of clusters was set to three to represent passenger cars (PC), heavy-duty vehicles (HV), and over-sized truck (OT). Three is chosen based on [46] and the fact that there are few vehicles longer than 22.5 m in our data. Figure 4 shows the distribution of the three vehicle types, in which the red triangle symbol is the centroid of each cluster.

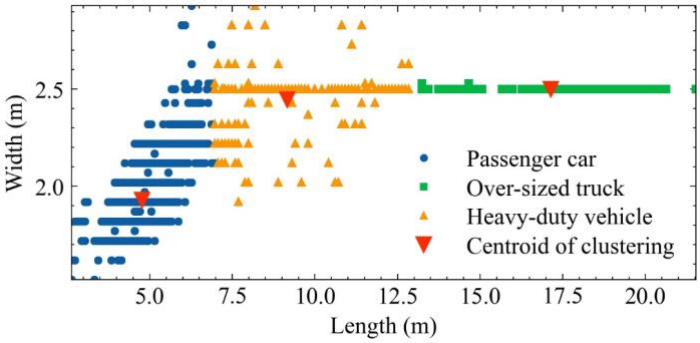


Figure 4 Clustering results of three vehicle types

C. LC events extraction

Figure 5 shows a common LC process. The lane where the LC vehicle starts the LC maneuver is called original lane, and the lane where the LC process ends is called target lane. In order to explore the interaction between LC vehicle and the vehicle in the original lane and vehicle in the target lane, this study fixed the interested vehicles as three vehicles within a certain distance of LC vehicles. Note that the interaction patterns are not comparable when the number of vehicles in the place of interest is different. These three vehicles are the LC vehicle

(*ego*), the lead vehicle in the original lane (*por*) and the lead/lag vehicle in the target lane (*ta*). Previous studies demonstrate that one of the most common types of LC scenarios is a vehicle changes lanes to pass a slower lead vehicle to maintain current speed or gain speed advantage [47, 48]. Thus, the preceding vehicle in the original lane (*por*) is determined. The LC scenarios can be automatically extracted by the following steps.

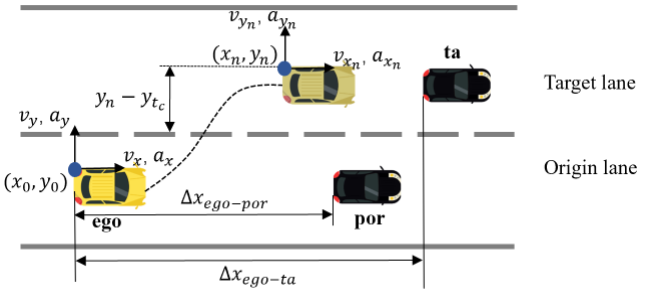


Figure 5 A schematic diagram of a typical LC interaction scenario

Step 1: Determine the *ego* vehicle. If the vehicle's driving lane ID changes (cross-lane), then the vehicle is marked as *ego* at this time stamp (t_c).

Step 2: Extract the complete LC process of the *ego* vehicle. According to the cross-lane timestamp recorded in step 1, the beginning (t_b) and ending (t_e) of LC process are searched forward and backward respectively in the trajectory record. We define the ending of LC as follows:

- Discontinuous increase of lateral displacement.
- $\left|y_{t_c} - y_{t_e}\right| > 0.9\,m$, where y_{t_c} is y position of the vehicle at time crossing lane. Assuming the vehicle width is 1.8m, when half of the vehicle width crosses the lane marker, it is considered that the vehicle has successfully changed to the target lane (it will not return to the original lane).
- $\left|a_{y_{t_e}} - a_{y_{t_{e-1}}}\right| < 0.01\ m/s^2$

And the beginning of LC is defined as follows:

- Discontinuous increase of lateral displacement.
- $\left|y_{t_c} - y_{t_b}\right| > 0.9\ m$
- $\left|a_{y_{t_b}} - a_{y_{t_{b+1}}}\right| < 0.01\ m/s^2$

Step 3: Determine three-vehicle interaction scenario. The filter rules are $\Delta x_{ego-por} < 120\text{m}$ and $\Delta x_{ego-ta} \in [-100\text{ m}, 100\text{ m}]$ at the beginning of LC.

Step 4: Classify interaction scenarios according to vehicle types. Based on the result of vehicle type classification, there are a total of 27 interactions between different vehicle types. In all interaction events, there are 578 *ego*(PC)-*por*(PC)-*ta*(PC) interaction events, 153 *ego*(PC)-*por*(OT)-*ta*(PC) interaction events, and 104 *ego*(PC)-*por*(HV)-*ta*(PC) interaction events. The frequency of other interaction types is less than 30. Therefore, the *ego*(PC)-*por*(PC)-*ta*(PC) interaction LC events are selected as the final samples in this study.

IV. Results and Analysis

A. LC interaction primitive extraction result

Firstly, the GMM-HMM is implemented to the 578 LC interaction events and generates 1224 primitives. In the training process of the GMM-HMM, the log-likelihood value is used to evaluate the learning performance of the model. In order to determine the optimal value of *N* in Eq. (3) and Eq. (7), the value of *N* increases by 1 from 1 in the training process until the state sequence cannot be calculated. Because when *N* increases, some state transition probabilities may be 0 so that there is no optimal state sequence. Among these *N* values, the *N* with the smallest log likelihood value of the model is selected. According to Wang, Xi and Zhao [49], *N* is equal to the number of GMM components *M*. Figure 6 shows an example of the GMM-HMM learning process for one LC interaction scenario, which illustrates that the model has been trained to converge. Figure 7 shows distribution of the duration of all primitives. Short duration contains limited information, which cannot reflect the evolution of interaction pattern. Thus, the primitives with less than 10 frames (0.4 s) are removed.

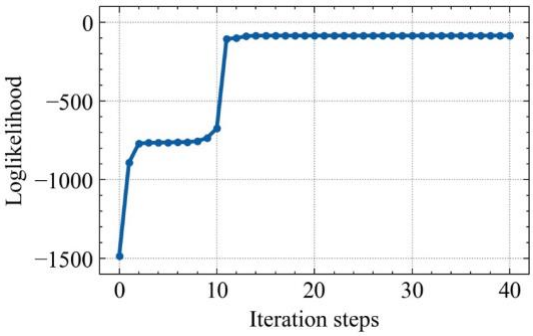


Figure 6 The learning process of the GMM-HMM to segment one of the LC interaction scenarios

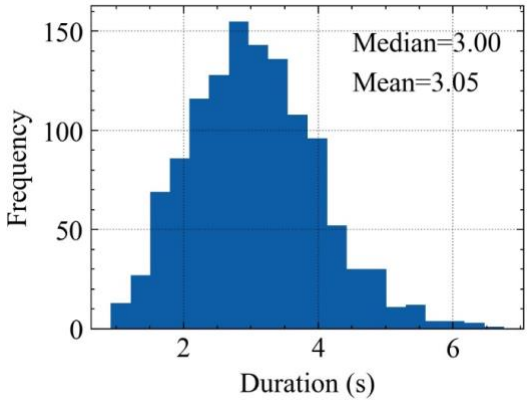


Figure 7 Distribution of LC interaction primitive duration

Figure 8 displays three common LC interaction events with extracted primitives. In order to clearly interpret the generated driving primitives, not only the trajectory but also the longitudinal speed, lateral speed and risk value are shown in top-down order in Figure 8. Different colors represent different primitives in an LC interaction scenario, and the primitives in different LC interaction scenarios are independent. The results

show that the GMM-HMM can automatically identify the boundaries of primitives with different semantic information from complex scenarios. In detail, the results are discussed from the following four aspects:

● The number and attribute of primitives in different LC interaction scenarios are significantly contrasting. For example, the LC interaction scenario of Figure 8 (b) consists of four primitives while that of Figure 8 (a) and (c) include two primitives. Note that the same color does not represent the same semantic segment in different events. For example, the blue primitives in Figure 8 (a) and (b) are obviously different.

● The primitives are interpretable. Taking Figure 8(b) as an example, the *ego* and *por* vehicles change lanes at the same time, and the *ta* vehicle with higher longitudinal speed in the target lane is behind the *ego* vehicle at the beginning of LC. After the LC process, all vehicles are in front of the *ta* vehicle in the target lane. The entire interaction process is segmented into four primitives. In the orange primitive, *ego* and *por* vehicle start LC with increasing speed in lateral direction. The *por* vehicle has the lowest speed and the *ta* vehicle has the highest speed in longitudinal direction, and the *ta* vehicle starts decelerating in order to avoid collisions. In the blue primitive, *ego* and *por* vehicles are approaching the lane marker while adjusting the lateral speeds. In the green primitive, *ego* and *por* vehicles cross the lane marker to reach the target lane and decelerate in the lateral direction. At the same time, *ego* and *ta* vehicles quickly decelerate in the longitudinal direction in order to avoid collision. In the red primitive, *ego* and *por* vehicles driving steadily on the target lane and the lateral speed of them tends to be zero. Due to the small gap between the *ego* vehicle and the *por* vehicle, the *ta* vehicle continues to decelerate until its speed is less than the speed of *ego* vehicle in longitudinal direction.

● The extracted primitives accurately distinguish the LC stage. In Figure 8(b) and (c), the primitives are segmented before and after cross-line. When two vehicles are in LC process but not in the same stage (Figure 8(a)), the primitives can also automatically make a trade-off between two vehicles' LC stage.

● The extracted primitives also have significant differences in risk level. In the process of interaction between vehicles, the risk level changes dynamically. The primitives can help identify high-risk semantic segments and the TTC value used to reflect their risk level. In the TTC graph of Figure 8(b), the blue primitive shows high risk, the orange primitive shows medium risk, the green primitive shows low risk, and the red primitive shows no risk.

Therefore, the complex LC interaction scenarios can be decomposed into explanatory primitives automatically extracted by the GMM-HMM to facilitate the understanding of interaction patterns.

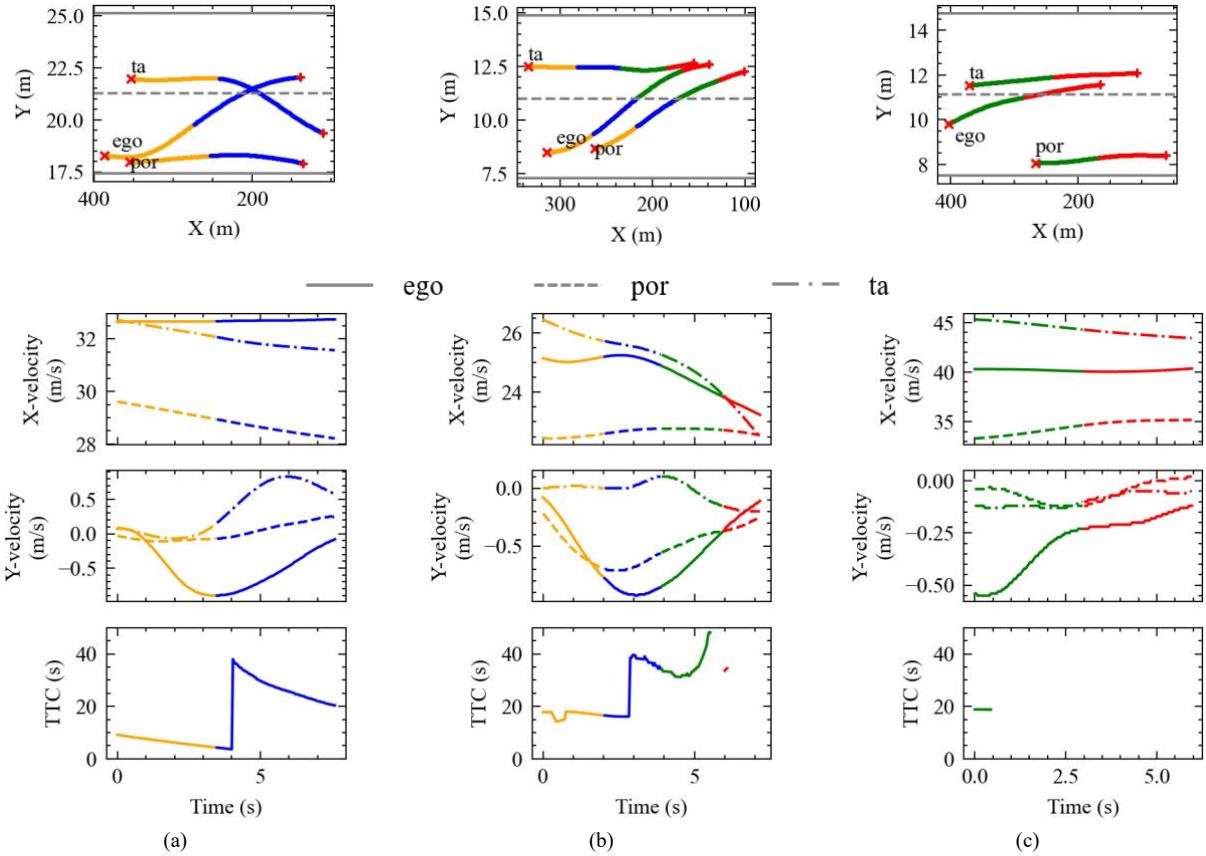


Figure 8 Decomposition results of three typical LC interaction scenarios. In the trajectory subgraph, "×" represents the starting of the trajectory, and "+" represents the ending of the trajectory. In the speed and risk value subgraphs, the solid line represents *ego* vehicles, the dashed line represents *por* vehicles, and the dotted line represents *ta* vehicles. Different colors represent different primitives in an LC interaction scenario.

B. Clustering evaluation

In order to summarize numerous primitives into finite interaction patterns, the DTW distance based K-means clustering is adopted to cluster the primitives. Before clustering, these primitives need to be scaled into the same length as the input of the clustering model. It is challenging to determine the optimal length of the input: a short length result in loss of the semantic information, while a long length increases the computational cost. In order to balance the learning performance and computational cost, this study chooses the median of all primitive durations (3 s) as the expected primitive duration (see Figure 7). Note that the primitives are restored to the original length after obtaining category labels. Figure 9 shows the scaled data fit the original data very well, which can minimize the information loss from scaling-down.

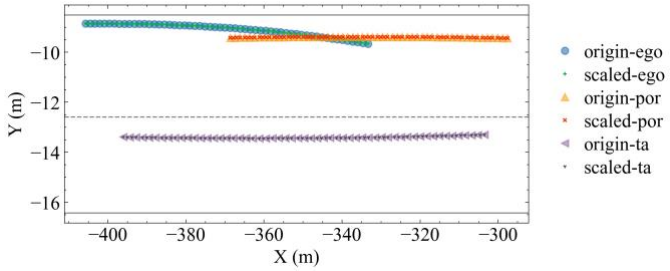


Figure 9 The scaled data and original data of one primitive

The number of clusters in the DTW distance based K-means clustering is determined based on the within-cluster sum-of-squares criterion. The within-cluster sum-of-squares criterion (λ_w) can be recognized as a measure of how internally coherent clusters are. The smaller the value, the better the clustering result. Figure 10 shows the results of change rate λ_w, and smoothed change rate of λ_w after quadratic polynomial fitting. Figure 10 indicates that with the increase of k, the clustering algorithm converges and the improvement of clustering effect (change rate of λ_w) decreases gradually. Considering the performance of the model and computational cost, k=13 is selected as the best number of clusters (When k=13, λ_w tends to be flat and change rate of λ_w approaches 0).

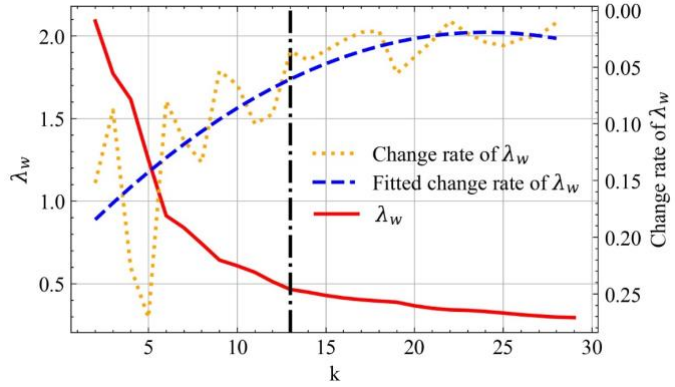


Figure 10 The curve of λ_w over the number of clusters k

After clustering the 1,224 primitives, the frequency and duration distribution of each cluster of interaction patterns are shown in Figure 11 and Figure 12 respectively. In Figure 11, it is not difficult to find that cluster #3 is the most common interaction pattern. Cluster #9 and cluster #13 are the least common interaction patterns. The following sections focus on these three clusters of interaction patterns.

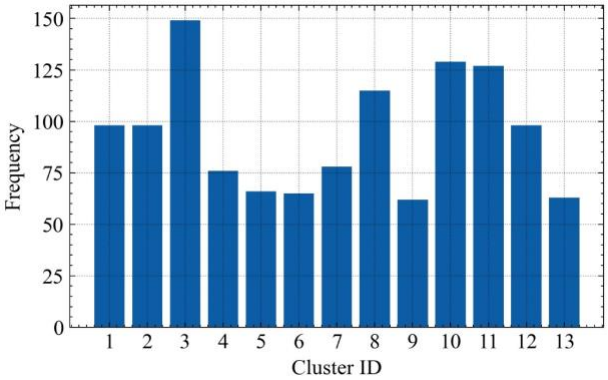


Figure 11 The frequency distribution of each cluster of interaction patterns

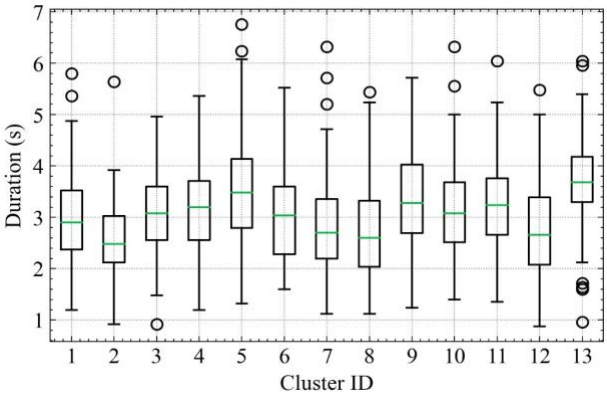


Figure 12 The duration distribution of each cluster of interaction patterns

C. Interaction patterns analysis

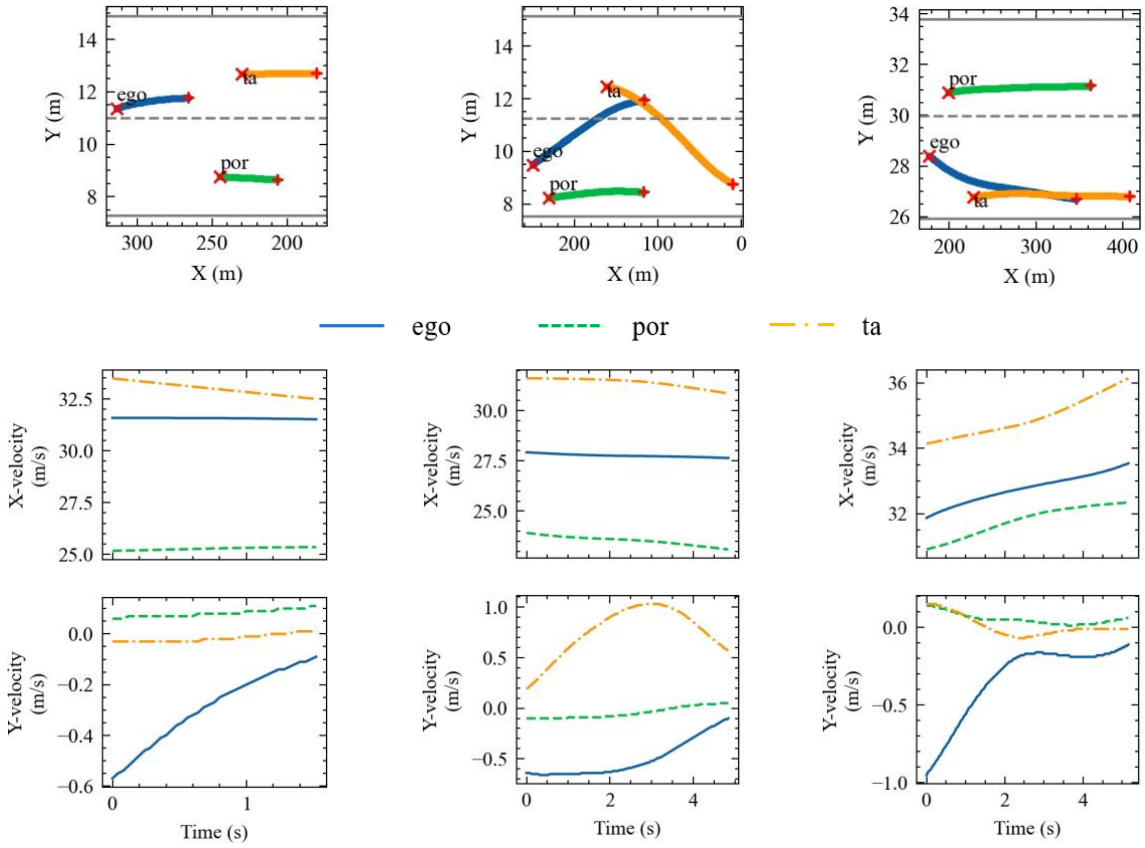
1) LC interaction pattern interpretation

Figure 13 (a) displays the most common interaction pattern (cluster #3). In the lateral direction, the *ego* vehicle just crossed the lane marker and reduce its lateral speed with a constant acceleration, while the lateral speed of the *por* and *ta* vehicles remain around zero. In the longitudinal direction, the *ta* vehicle has the highest speed while the *por* vehicle has the lowest speed. Figure 13 (b) and (c) represent the least common interaction pattern (cluster #9 and cluster #13). Figure 13 (b) shows the *ta* and *ego* vehicles change lanes to each other's lanes respectively (their lateral speed directions are opposite), and they are both at the stage of crossing the lane marker. Figure 13 (c) indicates the adjustment stage after the completion of LC. In this interaction pattern, the lateral speed of the *ego* vehicle gradually decreases until it becomes stable, while that of the *ta* and *por* vehicles remain around 0. Due to the long adjustment stage in the target lane, cluster #13 rarely occurs in real traffic condition.

2) LC interaction pattern risk analysis

According to the previous results, it was found that different primitives imply different risk levels. Further, the risk of interaction patterns is explored. The TTC value is selected to indicate both the risk of interaction patterns and the degree of interaction [43]. Figure 14 shows the distribution of each cluster of interaction patterns' risk. It is obvious that cluster #12 and cluster #10 interaction patterns are more concentrated in the lower TTC range than the other clusters. This suggests that the

risk level of these two interaction patterns are theoretically higher.



(a) cluster #3 (b) cluster #9 (c) cluster #13

Figure 13 Trajectory and velocity graphs of three clusters of primitives. In the trajectory subgraph, "×" represents the beginning of the trajectory, and "+" represents the ending of the trajectory. In the velocity subgraphs, the solid line represents ego vehicles, the dashed line represents por vehicles, and the dotted line represents ta vehicles.

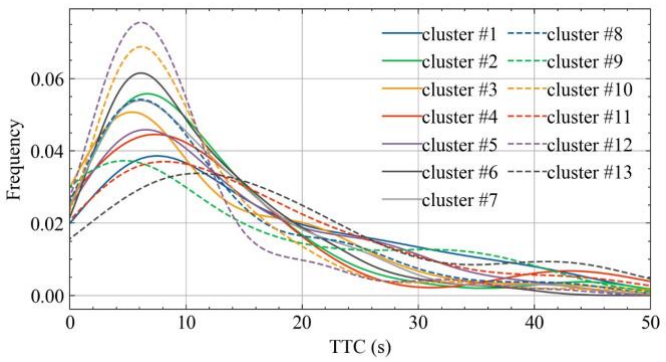
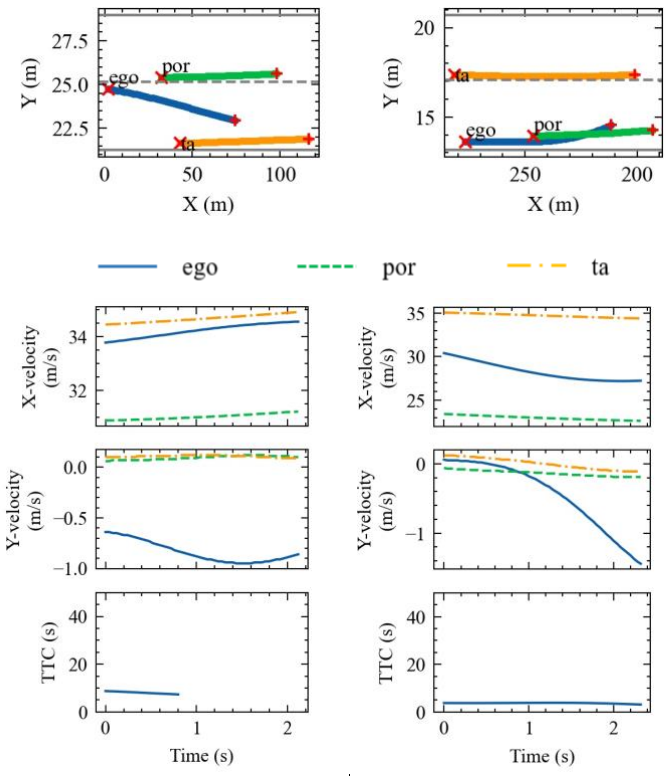


Figure 14 Distribution of TTC observation according to the cluster of interaction primitives

In order to perform in-depth analysis of two clusters of high-risk interaction patterns, Figure 15 shows their dynamic behavior and risk evolution in the interaction process. Figure 15 (a) represents the later stage of the *ego* vehicle crossing the lane marker. Before the *ego* vehicle crosses the lane marker, the interaction risk between the *ego* and the *por* vehicles is high. The high risk comes from the close distance and large speed difference between these two vehicles. There is also interaction risk between the *ego* and *ta* vehicles, because their paths cross and the *ego* vehicle's lateral speed is high. During the *ego* vehicle keep away from the lane marker, there is no interaction between the *ego* and *por* vehicles, but only between the *ego* and *ta* vehicles. However, the interaction risk between the *ego* and *ta* vehicles is low because of the large spacing. Figure 15 (b) shows the beginning stage of LC. At first, there are interaction risk between the *ego* and *por* vehicles (the longitudinal speed of *ego* vehicle is higher than that of *por* vehicle). In order to obtain a higher speed advantage, the *ego* vehicle starts to change lanes when the *ta* vehicle is approaching close to the *ego* vehicle from behind in the target lane. Thus, the *ego* vehicle reduces the longitudinal speed to wait for the chance of LC. During this period, the interaction risk between the *por* and *ego* vehicles decreases slightly. At about 1.5 s, the *ego* vehicle stops decelerating, so that the interaction risk between the *ego* and *por* vehicles continue to increase.

Therefore, the extracted driving primitives can also be used to thoroughly understand the mechanism and evolution of risk in LC interaction patterns.



(a) cluster #12 (b) cluster #10

Figure 15 Trajectory, speed and risk evolution graph of two examples from clusters #12 and #10 .

V. Conclusion

This study proposes a learning framework that combines primitive-based interaction pattern recognition methods and risk analysis methods. The GMM-HMM is used to decompose LC interaction scenarios into primitives, and TTC is used as an indicator for risk analysis. According to the analysis results of highD naturalistic driving dataset, the following findings are summarized:

(1) The complete LC interaction scenarios can be segmented into interpretable primitives, thereby identifying finite types of understandable interaction patterns.

(2) The interaction patterns reproduce the real interaction scenarios and reflect the interaction mechanism and evolution law.

(3) The proposed framework is suitable to analyze the high-risk interaction patterns in LC scenario and explain the risk formation process.

Therefore, the framework proposed in the study can be applied to analyze interaction patterns, understand human behavior, and provide prior knowledge for autonomous vehicle decisions, thereby promoting safe and smooth interaction between autonomous vehicles and human vehicles.

The practical application of the proposed method includes four aspects. 1) Vehicle interaction relationship is one of the important elements in the design of automatic driving test scenarios. The proposed method can be used to generate interaction pattern pool and enrich the driving scenario library. 2) In this study, several types of interaction patterns are obtained through clustering. Calibrating the driving decision models for these homogeneous patterns can improve the model performance [25] and enhance the decision-making of autonomous vehicles. 3) In the aspect of trajectory planning for autonomous vehicles, some probabilistic methods may generate multiple trajectories for each vehicle. Comparing the generated trajectories with the prior trajectories extracted by our method can help eliminate unreasonable trajectories. 4) For the identified high-risk interaction patterns, the test frequency of these patterns can be increased when designing the automatic driving test scenario to improve the proportion of effective tests.

The interaction pattern recognition and risk analysis framework proposed in this paper is flexible. It might be suitable for identifying other scenarios involving multiple traffic participants, such as pedestrian-bicycle-vehicle interaction at intersections and on urban roads. This study only identifies the interaction patterns between human driven vehicles. Future work will consider analysis of the interaction pattern between autonomous vehicles and human driven vehicles under mixed traffic conditions to further improve the decision-making of autonomous vehicles.

References

[1] M. Shawky, “Factors affecting lane change crashes,” *IATSS Research*, vol. 44, no. 2, pp. 155-161, 2020.

[2] T. A. Dingus, S. G. Klauer, V. L. Neale, A. Petersen, S. E. Lee, J. Sudweeks, M. A. Perez, J. Hankey, D. Ramsey, and S. Gupta, *The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment*, United States. Department of Transportation. National Highway Traffic Safety Administration, 2006.

[3] L. Yue, M. Abdel-Aty, and Z. Wang, “Effects of connected and autonomous vehicle merging behavior on mainline human-driven vehicle,” *Journal of Intelligent and Connected Vehicles*, 2021.

[4] T. Peng, X. Liu, R. Fang, R. Zhang, Y. Pang, T. Wang, and Y. Tong, “Lane-change path planning and control method for self-driving articulated trucks,” *Journal of intelligent and connected vehicles*, 2020.

[5] H. Woo, Y. Ji, H. Kono, Y. Tamura, Y. Kuroda, T. Sugano, Y. Yamamoto, A. Yamashita, and H. Asama, “Lane-Change Detection Based on Vehicle-Trajectory Prediction,” *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 1109-1116, 2017.

[6] P. G. Gipps, “A model for the structure of lane-changing decisions,” *Transportation Research Part B: Methodological*, vol. 20, no. 5, pp. 403-414, 1986.

[7] J. Tang, S. Yu, F. Liu, X. Chen, and H. Huang, “A hierarchical prediction model for lane-changes based on combination of fuzzy C-means and adaptive neural network,” *Expert systems with applications*, vol. 130, pp. 265-275, 2019.

[8] J. Tang, F. Liu, W. Zhang, R. Ke, and Y. Zou, “Lane-changes prediction based on adaptive fuzzy neural network,” *Expert Systems with Applications*, vol. 91, pp. 452-463, 2018.

[9] P. Hidas, “Modelling vehicle interactions in microscopic simulation of merging and weaving,” *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 1, pp. 37-62, 2005.

[10] R. Ponziani, *Turn signal usage rate results: A comprehensive field study of 12,000 observed turning vehicles*, 0148-7191, SAE Technical Paper, 2012.

[11] Y. Zhang, L. E. Owen, and J. E. Clark, “Multiregime approach for microscopic traffic simulation,” *Transportation Research Record*, vol. 1644, no. 1, pp. 103-114, 1998.

[12] G. Wang, J. Hu, Z. Li, and L. Li, “Cooperative lane changing via deep reinforcement learning,” *arXiv preprint arXiv:1906.08662*, 2019.

[13] A. Halati, H. Lieu, and S. Walker, "CORSIM-corridor traffic simulation model."

[14] Y. Wang, S. Zhao, R. Zhang, X. Cheng, and L. Yang, "Multi-Vehicle Collaborative Learning for Trajectory Prediction With Spatio-Temporal Tensor Fusion," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1-13, 2020.

[15] A. Ji, and D. Levinson, "Estimating the Social Gap With a Game Theory Model of Lane Changing," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1-10, 2020.

[16] A. Ji, and D. Levinson, "A review of game theory models of lane changing," *Transportmetrica A: Transport Science*, vol. 16, no. 3, pp. 1628-1647, 2020/01/01, 2020.

[17] N. Lyu, Y. Wang, C. Wu, L. Peng, and A. F. Thomas, "Using naturalistic driving data to identify driving style based on longitudinal driving operation conditions," *Journal of Intelligent and Connected Vehicles*, 2021.

[18] D. Del Vecchio, R. M. Murray, and P. Perona, "Decomposition of human motion into dynamics-based primitives with application to drawing tasks," *Automatica*, vol. 39, no. 12, pp. 2085-2098, 2003/12/01/, 2003.

[19] C. Bregler, "Learning and recognizing human dynamics in video sequences." pp. 568-574.

[20] Z. Ding, D. Xu, C. Tu, H. Zhao, M. Moze, F. Aioun, and F. Guillemard, "Driver Identification Through Heterogeneity Modeling in Car-Following Sequences," *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[21] F. Havlak, and M. Campbell, "Discrete and Continuous, Probabilistic Anticipation for Autonomous Robots in Urban Environments," *IEEE Transactions on Robotics*, vol. 30, no. 2, pp. 461-474, 2014.

[22] V. Gadepally, A. Krishnamurthy, and U. Ozguner, "A Framework for Estimating Driver Decisions Near Intersections," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 2, pp. 637-646, 2014.

[23] K. Tang, S. Zhu, Y. Xu, and F. Wang, "Modeling Drivers' Dynamic Decision-Making Behavior During the Phase Transition Period: An Analytical Approach Based on Hidden Markov Model Theory," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 1, pp. 206-214, 2016.

[24] W. Zhang, and W. Wang, "Learning V2V interactive driving patterns at signalized intersections," *Transportation Research Part C: Emerging Technologies*, vol. 108, pp. 151-166, 2019/11/01/, 2019.

[25] B. Higgs, and M. Abbas, "Segmentation and Clustering of Car-Following Behavior: Recognition of Driving Patterns," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 81-90, 2015.

[26] C. Zhang, J. Zhu, W. Wang, and J. Xi, "Spatiotemporal Learning of Multivehicle Interaction Patterns in Lane-Change Scenarios," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1-14, 2021.

[27] N. Saunier, T. Sayed, and K. Ismail, "Large-Scale Automated Analysis of Vehicle Interactions and Collisions," *Transportation Research Record Journal of the Transportation Research Board*, vol. 2147, pp. 42-50, 2010.

[28] Y. Ni, M. Wang, J. Sun, and K. Li, "Evaluation of pedestrian safety at intersections: A theoretical framework based on pedestrian-vehicle interaction patterns," *Accident Analysis and Prevention*, vol. 96, no. nov., pp. 118-129, 2016.

[29] N. Nadimi, D. R. Ragland, and A. Mohammadian Amiri, "An evaluation of time-to-collision as a surrogate safety measure and a proposal of a new method for its application in safety analysis," *Transportation letters*, vol. 12, no. 7, pp. 491-500, 2020.

[30] Y. Li, D. Wu, J. Lee, M. Yang, and Y. Shi, "Analysis of the transition condition of rear-end collisions using time-to-collision index and vehicle trajectory data," *Accident Analysis & Prevention*, vol. 144, pp. 105676, 2020/09/01/, 2020.

[31] E. Jeong, and C. Oh, "Evaluating the effectiveness of active vehicle safety systems," *Accident Analysis & Prevention*, vol. 100, pp. 85-96, 2017/03/01/, 2017.

[32] S. Noh, and K. An, "Risk assessment for automatic lane change maneuvers on highways." pp. 247-254.

[33] J. Wu, H. Wen, and W. Qi, "A new method of temporal and spatial risk estimation for lane change considering conventional recognition defects," *Accident Analysis & Prevention*, vol. 148, pp. 105796, 2020/12/01/, 2020.

[34] A. Laureshyn, Å. Svensson, and C. Hydén, "Evaluation of traffic safety, based on micro-level behavioural data: Theoretical framework and first implementation," *Accident Analysis & Prevention*, vol. 42, no. 6, pp. 1637-1646, 2010/11/01/, 2010.

[35] W. J. Schakel, V. L. Knoop, and B. van Arem, "Integrated Lane Change Model with Relaxation and Synchronization," *Transportation Research Record*, vol. 2316, no. 1, pp. 47-57, 2012/01/01, 2012.

[36] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257-286, 1989.

[37] A. Coates, and A. Y. Ng, "Learning feature representations with k-means," *Neural networks: Tricks of the trade*, pp. 561-580: Springer, 2012.

[38] P. St-Aubin, L. Miranda-Moreno, and N. Saunier, "An automated surrogate safety analysis at protected highway ramps using cross-sectional and before–after video data," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 284-295, 2013/11/01/, 2013.

[39] N. Saunier, T. Sayed, and K. Ismail, "Large-Scale Automated Analysis of Vehicle Interactions and Collisions," *Transportation Research Record*, vol. 2147, no. 1, pp. 42-50, 2010/01/01, 2010.

[40] A. Pierson, W. Schwarting, S. Karaman, and D. Rus, "Learning Risk Level Set Parameters from Data Sets for Safer Driving." pp. 273-280.

[41] V. Kurtc, "Studying Car-Following Dynamics on the Basis of the HighD Dataset," *Transportation Research Record*, vol. 2674, no. 8, pp. 813-822, 2020/08/01, 2020.

[42] F. Kruber, J. Wurst, S. Chakraborty, and M. Botsch, "Highway traffic data: macroscopic, microscopic and criticality analysis for capturing relevant traffic scenarios and traffic modeling based on the highD data set," *arXiv preprint arXiv:1903.04249*, 2019.

[43] F. Wirthmüller, J. Schlechtriemen, J. Hipp, and M. Reichert, "Towards Incorporating Contextual Knowledge into the Prediction of Driving Behavior." pp. 1-7.

[44] P. Schneider, M. Butz, C. Heinzemann, J. Oehlerking, and M. Woehrle, "Scenario-based threat metric evaluation based on the highd dataset." pp. 213-218.

[45] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems." pp. 2118-2125.

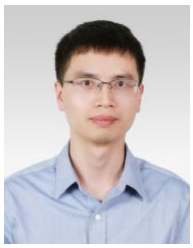
[46] S. Turner, *Traffic Monitoring in Recreational Areas*, United States. Federal Highway Administration. Western Federal Lands Highway …, 2010.

[47] S. Bogard, *Analysis of data on speed-change and lane-change behavior in manual and ACC driving*, 1999.

[48] S. Hetrick, "Examination of driver lane change behavior and the potential effectiveness of warning onset rules for lane change or" side" crash avoidance systems," Virginia Tech, 1997.

[49] W. Wang, J. Xi, and D. Zhao, "Learning and Inferring a Driver's Braking Action in Car-Following Scenarios," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 3887-3899, 2018.

Yue Zhang received the B.S. degree in school of economics and management from Chang'An University, Xi'an, China in 2018. She is currently working toward the Ph.D. degree in transportation engineering at the Tongji University, Shanghai, China. Her research interests include automated vehicles, traffic safety and driving behaviors.



Yajie Zou is associate professor at Tongji University, Shanghai, China. He holds the M.S. and Ph.D. in Transportation Engineering from Texas A&M University, and B.S. in Engineering Mechanics from Tongji University, Shanghai, China. Dr. Zou's main research interests are traffic operations, traffic safety and microscopic traffic simulation models.

Selpi received a PhD degree in computing from the Robert Gordon University in the UK in 2008, an MSc degree in bioinformatics from Chalmers University of Technology in Sweden in 2004, and a BSc degree in computer science from the University of Indonesia in 2000. She currently works at Chalmers University of Technology. Her current research interests include applications of machine learning and data science for transport and traffic safety domain, and understanding how mixed traffic, with vehicles with different driving styles and automation levels sharing the same roads, affects traffic safety and efficiency. She has served as a reviewer for several IEEE conferences and is an editor for IEEE Transactions on Intelligent Transportation Systems.

Yunlong Zhang received the B.S. degree in civil engineering and the M.S. degree in highway and traffic engineering from the Southeast University of China, Nanjing, China, in 1984 and 1987, respectively, and the Ph.D. degree in transportation engineering from Virginia Polytechnic Institute and State University, Blacksburg, VA, USA, in 1996. He is currently a Professor with the Zachry Department of Civil Engineering, Dwight Look College of Engineering, Texas A&M University, College Station, TX, USA. His research interests include transportation modeling and simulation, traffic operations, and computer and operations research applications in transportation.

Lingtao Wu is an assistant research scientist with Texas A&M Transportation Institute, College Station, TX, USA. He received the B.S. degree in traffic engineering from Beijing Jiatong University in 2005, Beijing China, and the Ph.D. degree in civil engineering from Texas A&M University, College Station, TX, USA, in 2016. His primary research interests include traffic crash data modeling, safety effectiveness evaluation, and risk analysis.