IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 7, 2022, accepted April 24, 2022, date of publication April 29, 2022, date of current version May 9, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3171347

Unsupervised Driving Style Analysis Based on Driving Maneuver Intensity

XIAN-SHENG LI, XIAO-TONG CUI, YUAN-YUAN REN, AND XUE-LIAN ZHENG
Transportation College, Jilin University, Changchun 130022, China

Corresponding author: Yuan-Yuan Ren (renyy@jlu.edu.cn)

This work was supported by the National Key R&D Program of China (2021YFC3001500).

ABSTRACT This paper proposes a novel unsupervised clustering framework to identify driving style not in terms of the discrete features of driving behavior data, but rather the time-varying patterns of driving maneuver intensity. This framework can describe the dynamic decision-making process of driving behavior and the continuity of driving data, and driving maneuver intensity is the basic of this paper. Therefore, detection, feature analysis and clustered on intensity of driving maneuvers are carried out using a threshold-based approach, hierarchical feature extraction, and k-means clustering. Then, to analyze fine-grained driving style, dynamic time windows are determined according to road alignment. In dynamic time windows, this paper constructs time-varying patterns based on driving maneuver intensity, which consider the intensity and frequency of driving behavior and preserve the time-varying characteristics of time-series data. However, not all dynamic time windows are equal in maneuvers' duration and number, which means the time-varying patterns of driving maneuver intensity are curves with various lengths. So that, for clustering time-varying patterns, this paper proposes a novel curve clustering algorithm named Similarity-Based Clustering with Dynamic Time Warping (SBC-DTW) that can cluster curves with various lengths. The empirical results based on real driving data demonstrate that the proposed framework can classify driving style more accurately than the classical method. Moreover, according to this framework, we can have an in-depth understanding of dynamic driving behavior and the composition of drivers' long-term driving styles.

INDEX TERMS Driving behavior, unsupervised driving style analysis, dynamic decision-making process, driving maneuver, curve clustering.

I. INTRODUCTION

Driving style analysis has become the focus of public attention with the development of technology, which plays a vitally important role in road safety [1]–[3] eco-driving [4], [5], vehicle insurance [6], [7] and intelligent vehicle design [8], [9]. Driving style is usually defined as a habitual way of driving, which is characteristic of a driver or group of drivers [1]. Generally, it describes a series of dynamic actions taken when drivers are driving, according to their own conditions [10]. So far, both subjective and objective methods have been used in previous researches. Generally, subjective methods have mostly adopted the questionnaire [11], [12]. The former researchers designed questionnaires according to their experience and asked drivers to fill them out so that they could identify different drivers.

In recent years, objective methods have become more and more popular, as they're more objective and easier to implement than subjective analysis. Many methods have been used in objective driving style analysis. No matter which method is used, high-quality driving behavior characteristics are important prerequisites for the effective identification of driving style. High-risk drivers tend to have shorter time headways (THWs), harder braking and more frequent lane changes [1], [13]. Thus, scholars have utilized the statistical features of velocity, acceleration, and THW as driving behavior features. For example, Chen *et al.* [14] calculated the mean and standard deviation of the steering wheel angle. Shi *et al.* [15] considered more than 1,000 driving behavior features using a variety of functions. In addition to the basic

The associate editor coordinating the review of this manuscript and approving it for publication was Yichuan Jiang.

statistical features, scholars have also used driving behavior characteristics such as the zero-crossing rate [16] and P-feature [17]to express the differences in driving behaviors. Using these features, researchers have applied well-trained classifiers such as XGBoost [15], the Markov model [18], the Semi-Supervised Tri-CatBoost [16], structural equation models [19], the semi-supervised support vector [20], and neural networks [21], [22] to identify driving styles. The above methods have achieved ideal results, whereas they require labeled data.

Considering that the acquisition of data labels in real life consumes a lot of manpower and financial resources [23], unsupervised driving style analysis has attracted attention. We can effectively avoid the limitation of manually labeling data, and effectively analyze a large volume of driving data in the context of big data to identify driving behavioral semantics and driving style utilizing the unsupervised method.

Recently, some researchers [24] have used clustering algorithms to achieve the analysis of driving styles. Scholars have also used various methods to construct features using driving parameters. Based on these features, K-means [25], the Gaussian mixture model [14], [26], the Hierarchical cluster analysis [27], and the K-means clustering-based Support Vector Machine [28] have been used to directly identify driving characteristics.

As mentioned above, both supervised and unsupervised classification can only give one profile of driver's driving style. However, it is generally accepted that driving style is constantly changing with the external environment [29]. To describe dynamic driving style changing with surrounding environments, De Zepeda *et al.* [29] proposed a position-dependent dynamic clustering framework, and achieved dynamic clustering of driving style. Also, Murphey *et al.* [30], E Suzdaleva *et al.* [31] and other scholars [32] have achieved online recognition of driving behavior characteristics using supervised methods.

To date, all methods usually construct discrete features of continuous driving data, and use clustering algorithms or classifiers for those features to identify driving style. Considering the continuity of driving data and dynamic decision-making process of driving behavior, continuous fine-grained features should be proposed to estimate driving behavior characteristics as well as corresponding algorithms.

Given the gaps in the literature, this paper proposes a novel unsupervised driving style clustering framework to identify driving style. In our paper, we carry out driving style analysis using unsupervised methods. Sample groups with similar characteristics can be found using clustering methods, which overcomes the disadvantages of manually labeling data. Decomposing complex driving behaviors into basic driving maneuvers can facilitate driving style analysis and identification [33], so this paper takes driving maneuvers as the basic unit. And fine-grained driving styles in dynamic time windows are characterized by the time-varying patterns of driving maneuver intensity, which consider the intensity and frequency of driving maneuvers.

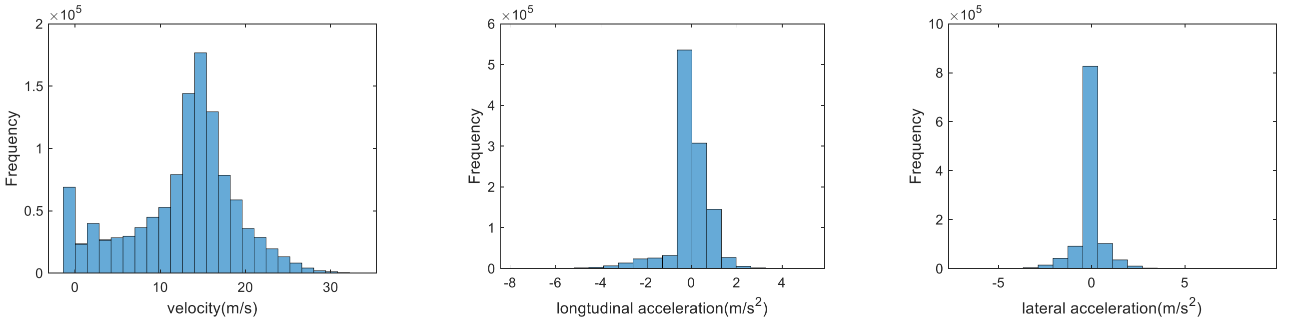
FIGURE 1. Testing device.

We also propose a new curve clustering algorithm for clustering curves with various lengths named Similarity-Based Clustering with Dynamic Time Warping (SBC-DTW). So that, clustering analysis of driving style in dynamic time windows is achieved according to time-varying patterns.

Based on this algorithm, this method can capture drivers' changing driving behaviors, achieve dynamic analysis of it, improve the result's interpretability and expand its range of applications. The main contributions of this paper can be summarized as follows:

(1) In this paper, dynamic time windows are determined according to road alignment. We achieve a clustering analysis of driving style based on dynamic time windows, which can effectively analyze drivers' changing fine-grained driving styles, and explain the composition of long-term driving style based on fine-grained driving style.
(2) We take driving maneuvers as the basic units. Considering both the intensity and frequency of driving maneuvers, we take the time-varying pattens of driving maneuver intensity as the feature to describe driving style. In this way, the dynamic decision-making process of driving behavior can be described, and the continuity of driving behavior data can be preserved. This improves the interpretability and range of applications of driving style. To the best of our knowledge, this is the first paper to use continuous features to solve the driving style analysis problem.
(3) Time-varying patterns are curves with various lengths due to dynamic time windows. Aiming at the problem of clustering curves with different lengths, we propose SBC-DTW to solve it, which avoids local optimization of hierarchical clustering of the classical algorithm.

The paper is organized as follows: In Section II, the data sources and tests are discussed in detail. In Section III, the four-part framework for unsupervised driving style analysis based on driving maneuvers is proposed. The results for each part are presented and discussed in Section IV. Finally, the paper is concluded in Section V.



(a) Statistical results for velocity **(b) Statistical results for longitudinal acceleration** **(c) Statistical results for lateral acceleration**

FIGURE 2. **Statistical results for processed variables.**

II. DATA DESCRIPTION

A. PARTICIPANTS

16 drivers (10 males, 6 females; age range 28∼50 years old, average age = 29.8, standard deviation (SD) = 2.7; driving experience 0∼12 years, average = 7.6 years, SD = 3.3) were paid for their participation in this study. We carried out naturalistic driving experiments using the same equipment throughout to avoid systematic errors. All drivers were required to drive the vehicle in similar conditions to minimize potential disturbance caused by external factors. These tests were conducted with the consent of all participants.

B. EXPERIMENT DESIGN

The experiments were carried out in the RADS 8 DOF Panoramic Driving Simulation (shown in Fig. 1). The test route contained 11 curves and the total round trip was about 10.35 km.

C. DATA PROCESSING

The data recorded by the RADS 8 DOF Panoramic Driving Simulation are recorded at a sampling rate of 60 Hz. Because raw data are volatile, we used a moving-average algorithm to smooth the data. The statistical results from the processed data are shown in Fig. 2. We can see that the velocity falls in the 0-34 m/s range, longitudinal acceleration falls in the -7 m/s^2-6 m/s^2 range, and lateral acceleration ranges from -7 m/s^2 to 9 m/s^2.

III. METHOD

Generally, the differences between drivers are analyzed using discrete features extracted from driving data in previous studies. However, we all know that different drivers will take kinds of driving maneuvers over time when they execute driving-tasks, and the driving data are long-term continuous time-series data. It's evident that these discrete features in the classical method cannot describe dynamic decision-making process of driving behavior over time and the continuity of driving data. So, it's necessary to find a continuous feature which can extract more information from driving data.

Driving data contain a large number of driving maneuvers, and previous studies have been proved decomposing complex driving behaviors into basic driving maneuvers can facilitate driving style analysis and identification [33]. So, we take the driving maneuvers as the basic unit to construct the continuous feature. To better understand the characteristics of driving behavior from maneuver level, we cluster driving maneuvers on intensity based on maneuver-level intensity features. Specifically, these intensity features are comprehensive features coupling multiple variables. Since driving behavior is embodied by intensity and frequency, we take the time-varying patterns of driving maneuver intensity as our continuous feature after obtaining driving maneuver intensity, which can compensate for the shortcomings of previous studies.

Also, in order to obtain fine-grained driving style, the data should be a period of time rather than the whole period of travel. Based on that, this paper analyzes time window data, namely driving data from a period of time containing a lot of maneuvers. Further, data from different road alignments would have different characteristics. We determine the time windows according to the road alignment so as to keep the data characteristics consistent within each time window.

We finally identify fine-grained driving style in dynamic time windows according to time-varying patterns of driving maneuver intensity. The time-vary patterns are curves with various lengths as a result of dynamic time windows. Therefore, we are supposed to cluster driving style through a clustering algorithm for curves with various lengths. Although this kind of clustering has been used successfully in other fields, the application for driving style analysis is still lacking. Given the gaps in the researches, we propose a novel curve clustering algorithm to cluster curves with various lengths.

The novel unsupervised driving-style clustering framework based on driving maneuver intensity is shown in Fig. 3. This framework is divided into four parts: driving maneuvers detection, feature analysis of driving maneuvers, driving maneuvers clustered on intensity, and driving style analysis. The first three parts are all used to get the driving maneuver intensity, which is prepared for the acquisition of time-varying patterns. Based on driving maneuver intensity, we get the time-varying patterns of driving maneuver intensity for dynamic time windows. Then, we identify driving style using

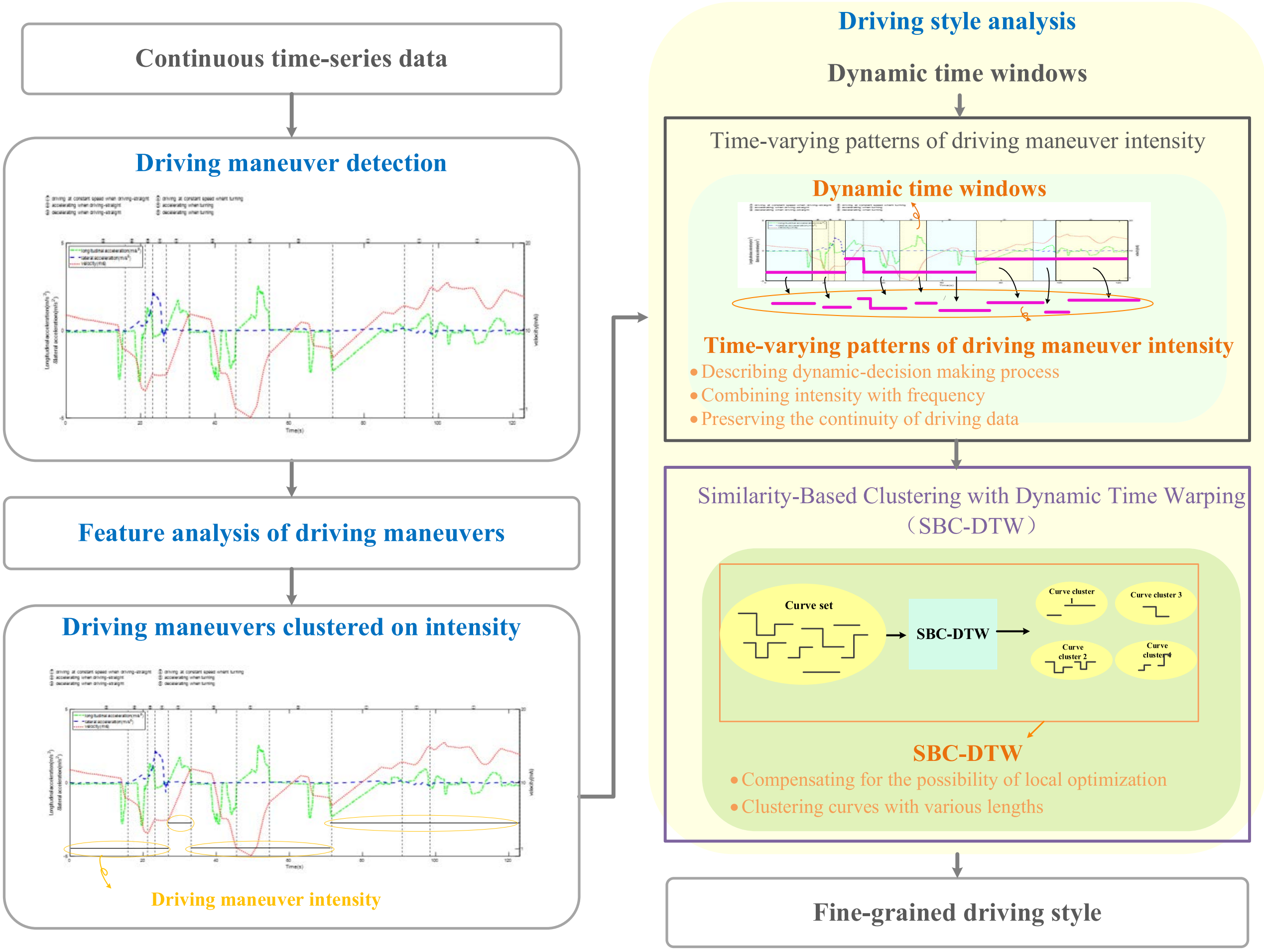


FIGURE 3. Flow diagram of the proposed framework.

our novel curve clustering algorithm for curves with various lengths.

This framework firstly avoids the limitation of manual labeling because it achieves the analysis of driving styles through the unsupervised method. Secondly, the fine-grained driving style analysis based on dynamic time windows can effectively capture dynamically changing driving styles. Last but not the least, this framework constructs the time-varying patterns of driving maneuver intensity that describes the dynamic decision-making process involved in driving behavior and the continuity of the data and uses a novel curve clustering algorithm to analyze driving style.

A. DRIVING MANEUVER DETECTION

We choose the driving maneuver as the basic unit to construct time-varying patterns. *Driving maneuvers*, in this paper, is defined as the basic building units of complex driving behavior. Longitudinal and lateral maneuvers are involved when drivers control vehicles. Longitudinal maneuvers include acceleration, deceleration, and driving at a constant speed. Lateral maneuvers include turning and driving-straight. Thus, we couple the longitudinal and lateral maneuvers to identify the driving maneuvers. We use a threshold-based approach to extract the driving maneuvers according to velocity, longitudinal acceleration, and lateral acceleration.

After processing the data, we calculate the driving trajectory radius (R) based on the velocity (vel) and lateral acceleration ($latacc$), as shown in(1).

$$R = \frac{vel^2}{latacc} \tag{1}$$

If $R > 1000$m, the vehicle is considered to be driving-straight, and if $R < 1000$m, it is considered to be turning. Further, we extract the complete wave of longitudinal acceleration, and then acceleration, deceleration, and driving at a constant speed are detected according to the peak (or trough) threshold. Finally, the driving maneuvers are detected by considering both longitudinal and lateral maneuvers. Specifically, the complete wave of longitudinal acceleration is the process of longitudinal acceleration increasing (decreasing) from zero to a certain value and then decreasing (increasing) to zero again. We determine the threshold as follows:

TABLE 1. Variables, operations, and descriptions used for feature construction.

Variable	Code	Description
Velocity	*vel*	Time-series data, velocity of vehicle
Longitudinal acceleration	*longacc*	Time-series data; positive values represent acceleration; negative values represent deceleration
Lateral acceleration	*latacc*	Time-series data; positive values represent turning left; negative values represent turning right
Operations	Code	Description
Basic center	mean, trim	Mean and trimmed mean
Basic dispersion	std, iqr	Standard deviation and quartile deviation
Mean absolute deviation	mad	Measure variability or dispersion
Percentile values	p05, p10, q1, q2, q3, p90, p95	The 5th, 10th, 25th, 50th, 75th, 90th and 95th percentiles, used to represent the data profile and distribution pattern
Extreme values	min, max	Minimum and maximum values
Information entropy	Shannon, ApEn	Shannon entropy and approximate entropy, used to describe the amount of information in the data and the complexity of the time series

- acceleration: $longacc > 1.0\text{m/s}^2$
- driving at a constant speed: $-1.0\text{m/s}^2 < longacc < 1.0\text{m/s}^2$
- deceleration: $longacc < -1.0\text{m/s}^2$

B. FEATURE ANALYSIS OF DRIVING MANEUVERS

1) FEATURE CONSTRUCTION

Features usually refer to data attributes that are beneficial to machine-learning algorithms. Therefore, before analyzing the maneuver intensity, we need to construct features, so as to comprehensively represent the data and improve the accuracy of the subsequent algorithms. In this paper, therefore, we select velocity, longitudinal acceleration, and lateral acceleration as the feature parameters to represent driving behavior. Firstly, statistical features are the most common ones and represent the data distribution. Secondly, entropy reflects the amount of information in complex time series. Therefore, 48 features are constructed to represent the data, considering statistics and entropy. The variables, operations, and descriptions used for the feature construction are shown in Table 1.

2) FEATURE EXTRACTION

In order to reduce the number of features so as to obtain more representative features and reduce the computational complexity, feature extraction is necessary. Feature extraction can be used to identify patterns and underlying structures in data sets and to combine original features to create new ones that better describe the data.

Because different variables have different weights when describing driving behavior, and due to different data distributions, we need to extract the variable-level intensity features and the maneuver-level intensity features successively.

Based on the above consideration, we use the hierarchical feature extraction to extract the maneuver-level intensity features. We calculate the feature weights twice, and use polynomials to combine the complex multi-dimensional features into comprehensive single-dimensional features. The maneuver-level intensity features comprehensively represent both the coupling effect of variables and different variables' different distributions. As shown in Fig. 4, the hierarchical

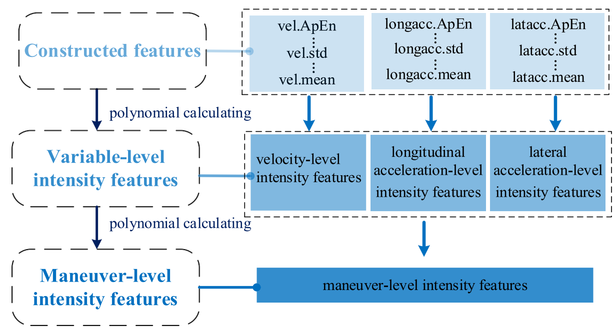


FIGURE 4. Hierarchical feature extraction system.

feature extraction system consists of three layers: the bottom layer is constructed features, the middle layer is variable-level intensity features, and the top layer is maneuver-level intensity features.

Specifically, we firstly calculate the weights of the constructed features using the entropy weight method [34], and obtain the variable-level features var_f_i (i.e., the variable of the i^{th} maneuver) using a polynomial. Then, we use principal component analysis (PCA) to calculate the weight of var_f_i, and extract the maneuver-level intensity features man_f_i (i.e., the i^{th} maneuver) using a polynomial.

After obtaining the weights of the constructed features $c\_f_{var,ik}$ (i.e., the k^{th} constructed feature of the variable for the i^{th} maneuver), we calculate the variable-level intensity features var_f_i, using (2).

$$var\_f_i = \sum_{k=1}^{K} c\_f_{var,ik} \cdot w_{var,ik} \tag{2}$$

where $var = [vel\ longacc,\ latacc]^T$, and $i = 1, 2, \ldots, I$. i is the number of maneuvers and k is the number of constructed features of the variable. Then, we use PCA [35] to calculate the weights of variable-level intensity features. PCA firstly converts correlation variables into linear unrelated variables by orthogonal transformation, and calls them principal components (PCs). Then, the number of PCs is selected according to the explained variance ratio. The explained variance ratio, which is the variance contribution percentage of the original feature parameters represented by the PC, is the proportion of the original information content represented by each PC in the total original information content represented by all the PCs.

The weight of each PC w_p_{im} (the m^{th} PC weight of the i^{th} event) is obtained by normalizing the explained variance ratio of the selected PCs. The weight of the PC containing more original information is larger. The PC must represent at least 90% of the variance of the original features, and then the maneuver-level intensity features man_f_i are extracted using (3) according to weight w_p_{im}, var_f_i, and $\alpha_{var,im}$ (the variable loading of the variable-level intensity features variable under the m^{th} PC for the i^{th} maneuver).

$$man\_f_i = \sum_{m=1}^{M} w\_p_{im} \left(\sum_{var} a_{var,im} \cdot var\_f_i \right) \tag{3}$$

where m is the number of PCs. In this paper, we use a multi-layer weight calculation system considering the data distribution differences of the variables and the differences in the weights of the variables in terms of how they affect driving behavior at the same time. We finally obtain the one-dimensional maneuver-level intensity features —the larger value is along with the greater maneuver intensity.

C. DRIVING MANEUVERS CLUSTERED ON INTENSITY

Driving behavior is embodied by intensity and frequency. Therefore, we need to use both intensity and frequency to represent driving behavior characteristics. However, the driving maneuvers extracted above only have a semantic definition and cannot quantitatively describe the intensity of maneuvers. Therefore, we labeled the intensity of the driving maneuvers.

More than 3,000 driving maneuvers were extracted from driving data. Although there were many maneuvers, the essential intensity clusters of the maneuvers were limited due to their similarity. For intuitive recognition of the operational characteristics of maneuvers and to understand the driving behavior characteristics from a maneuver level, maneuver intensity labels were extracted by clustering the various maneuvers into a smaller number of clusters according to maneuver-level intensity features.

The K-means algorithm is a classical clustering algorithm. Previous works have proved that K-means clustering is superior to other traditional algorithms [29], [36], [46]. This algorithm firstly randomly selects k cluster centers, then iteratively assigns each sample to the nearest cluster according to Euclidean distance until cluster membership stabilizes. Although there have been many variants of k-means, the k-means is still popular as its efficiency and empirical success [47], [48].

We use it to cluster the driving maneuvers. After obtaining the maneuver-level intensity features, we cluster the driving-straight and turning maneuvers separately because of the difference in their data distributions. Further, the intensity label of each maneuver is obtained according to the statistical features of the maneuver-level intensity features, and we use numerical values to represent the maneuvers of different clusters. Driving maneuver intensity clusters with larger values have greater intensity.

D. DRIVING STYLE ANALYSIS

1) DYNAMIC TIME WINDOW DETERMINATION

As mentioned above, we determine the time windows according to the road alignment, not only to ensure each time window contains enough maneuvers to reflect the dynamic driving behavior decisions, but also to avoid mixing data with different distributions, so that we can effectively analyze the fine-grained driving styles.

2) DRIVING STYLE CLUSTERING BASED ON THE TIME-VARYING PATTERNS OF DRIVING MANEUVER INTENSITY

After determining the dynamic time windows, we study driving style in the time windows. Considering both the intensity and frequency of the maneuvers, we use the time-varying patterns of driving maneuver intensity as the feature to describe driving styles. Specifically, the time-varying patterns of driving maneuver intensity are curves of the driving maneuvers changing over time. This feature preserves the continuity of the driving data and reflects the dynamic decision-making process of driving behavior. However, the time-varying patterns are curves with different lengths due to duration between dynamic time window is different.

For clustering curves with various lengths, Hierarchical Clustering based on DTW (HC-DTW) has been applied successfully in manufacturing [37], geology [38], transportation [39], and biology [40]. HC-DTW firstly calculates the distance matrix between any two curves in the curve set, then clusters curves of different lengths using a hierarchical clustering strategy based on the distance matrix. Hierarchical Clustering has many advantages, such as a visual clustering process, simple clustering strategy, and others. It also has some disadvantages, for example, that the greedy principle could easily lead to local optimization and that the clustering results depend on the selection of merging points.

To overcome the shortcomings of previous studies, we propose a novel curve clustering algorithm named Similarity-Based Clustering with Dynamic Time Warping. (SBC-DTW), which compensates for the possibility of local optimization. We use this algorithm to carry out the driving style analysis based on time-varying patterns of driving maneuver intensity, and thereby improve the accuracy of driving style analysis.

In this paper, we use the DTW to calculate the distance matrix between any two curves in the curve set, and cluster curves with various lengths by using a Similarity-Based Clustering [41].

Definition 1: Defining the curve set. $C\ (L_1, L_2, \ldots, L_n)$ The distance matrix D_C of curve set C is the distance between all pairs of curves in curve set C.

$$D_C = \begin{pmatrix} d_{11} & \cdots & d_{1n} \\ \vdots & \ddots & \vdots \\ d_{n1} & \cdots & d_{nn} \end{pmatrix}$$

$$d_{ij} = \underset{\substack{i=1,2,\ldots,n \\ j=1,2,\ldots,n}}{DTW} \left(L_i, L_j\right) \tag{4}$$

Definition 2: The intra-cluster similarity of C. is the maximum distance between any two curves in the curve set.

$$S\left(C\right) = \max\left(D_C\right) \tag{5}$$

Definition 3: the distance between curve L_i and curve set C is the average distance between curve L_i and any curve in curve set C.

$$d_{Ci} = \frac{1}{n}\sum_{j=1}^{n} DTW\left(L_j, L_i\right) \tag{6}$$

According to the clustering strategy of the Similarity-Based Clustering, this paper proposes SBC-DTW. The pseudo-code of SBC-DTW is shown in Algorithm.

Algorithm 1 Similarity-Based Clustering with Dynamic Time Warping (SBC-DTW)

Input: curve set $C = \{L_1, L_2, .., L_n\}$, threshold T
Output: curve clusters $\{C_1, C_2, C_3, \ldots, C_m\}$

```
1:  Initialize m = 0
2:  while C is not empty
3:      m = m+1
4:      calculate distance matrix D_C according to (4)
5:      find L_a and L_b, which d_ab = max D_C; initialize C1 =
        L_a, and update C = C − C1
6:      for L_1, L_2, .., L_e ∈ C, calculate D2 =
        {d_C11, d_C12, ..., d_C1e} according to (6)
7:      find L_z which d_C1z = min (D2), and initialize C1′ =
        C1 + L_z
8:      if S (C1′) < T
9:          update C1 = C1 + L_z and C = C − C1, then go to
        6
10:     else
11:         C_m = C1, then go to 2
12:     end if
13:     return: curve clusters {C_1, C_2, C_3, ..., C_m}
14: end while
```

This algorithm avoids the local optimization caused by hierarchical clustering, and calculates the intra-cluster similarity and the distance between a curve and a curve cluster based on the DTW. By clustering curves with different lengths, we efficiently cluster the driving styles in dynamic time windows based on the time-varying patterns of driving maneuver intensity. The result of the clustering maintains the continuity and temporal information of the driving data. In this way, we analyze driving styles more accurately and improve the interpretability and range of applications of the results of the driving style clustering.

3) DRIVING STYLE LABELING CONSIDERING DRIVING MANEUVER INTENSITY AND ITS TRANSITION CHARACTERISTICS

Curve clustering is an unsupervised data mining method, so the clusters obtained do not have semantic labels. In order to intuitively understand driving styles, we label the driving style of each cluster considering both the intensity and the transition characteristics of the maneuvers.

Driving maneuvers transfer into each other, and maneuvers can transfer into other maneuvers or achieve self-transfer. The transition probability of a maneuver represents the possibility of transition between maneuvers.

θ_t is the driving maneuver at time t. man_i and man_j are maneuver intensity clusters. The transition probability from man_i to man_j is defined as follows.

$$\begin{aligned} P_{ij} &= P\left(q_{t+1} = man\_j | q_t = man\_i\right) \\ &= \frac{n\left(q_t = man\_i, q_{t+1} = man\_j\right)}{n\left(q_{t+1} = man\_j\right)} \end{aligned} \tag{7}$$

where n is the number of maneuver intensity transitions from man_i to man_j.

After obtaining the transition probability of each driving style cluster, we label the driving style of each cluster based on the transition characteristics and intensity of maneuvers. Specifically, we label the driving style based on the following aspects:

- Diversity of maneuver intensity transition forms

The more diverse the transition forms are, the more random the driving behavior is, which means there is more information. Some drivers tend to transfer repeatedly between various maneuvers so as to reduce the time they take to reach their destination as much as possible or to obtain more driving pleasure. Such driving behavior is often considered to be highly aggressive. However, other drivers prefer to operate vehicles in a fixed mode where the transition forms are simpler. Such driving behavior is more conservative and less aggressive.

The amount of information in data can be quantified by information entropy. When data are more uncertain, they contain more information, so their information entropy are consequently higher. We calculate the information entropy of driving style cluster $Z(H(z))$ using the transition probability, and use the entropy to quantify the diversity of forms of transition intensity between maneuvers.

$$H\left(z\right) = -\sum_{i=1|j=1}^{K} p_{zij}\log\left[p_{zij}\right] \tag{8}$$

where K is the number of maneuver intensity clusters, and p_{zij} is the transition probability of transition from man_i to man_j in driving cluster Z.

- Maneuver intensity transition tendency

The transition tendency specifically describes the decision preferences in different driving behaviors. Because high-intensity maneuvers require more advanced driving skills,

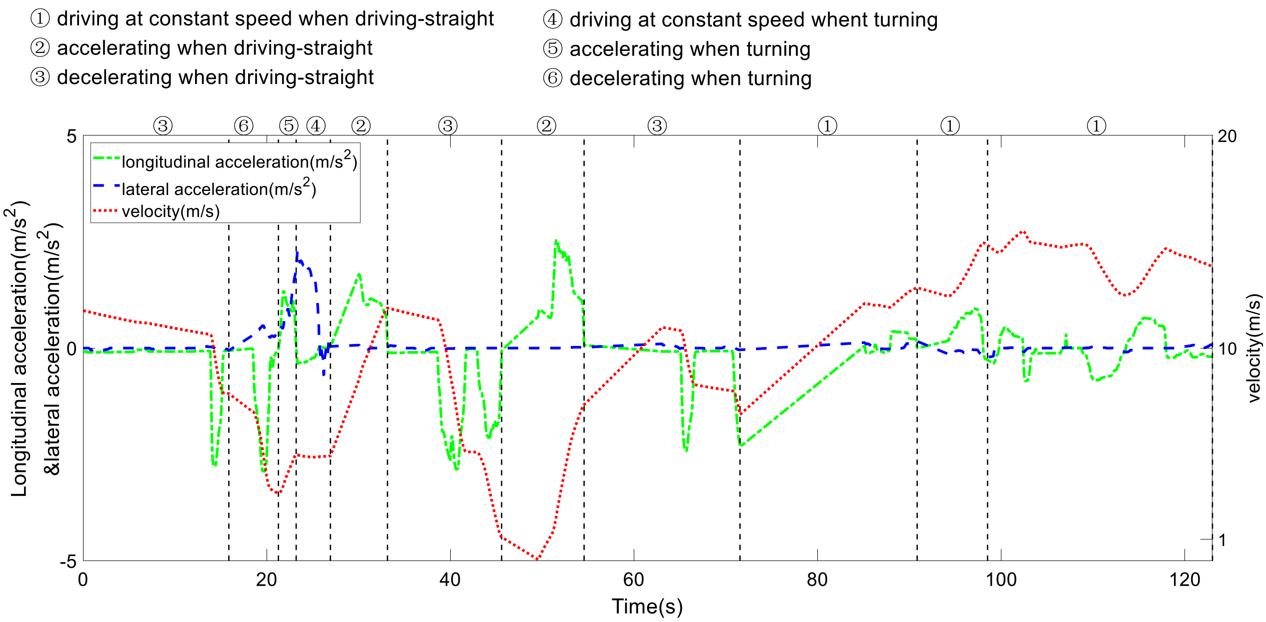


FIGURE 5. **Detection of driving maneuvers from time-series driving variables.**

when a driver transfers from the current maneuver to a higher-intensity maneuver, the driver will feel more excitement. However, the driver will feel less excited when transferring to a lower-intensity maneuver. From the above, we can describe driving style differences based on changes in the maneuver intensity clusters before and after maneuver transition. Further, existing research [42] has proved that differences in driving styles are mainly reflected in the intensity and difficulty of maneuvers. Thus, we assign different weights to different transition forms according to the maneuver intensity clusters after the transition. The greater is the intensity of the maneuver after the transition, the larger is the weight of the transition form.

As shown in (8), we take each event cluster as the current event in turn, and find the corresponding transition form with the maximum transition probability. Then, we use the weighted sum of difference between the maneuver intensity clusters after transition and the current maneuver intensity clusters for this transition form to obtain the transition tendency of this driving style cluster.

$$T(z) = \sum_{p=1}^{P} \left(w_{zp} \cdot tran_{zp} \right) \tag{9}$$

where P is the number of transition forms, w_{zp} is the weight of maneuver intensity transition form $p$in cluster Z, and $tran_{zp}$ is the range of maneuver intensity transition form p in cluster z. $w_{zp} = \frac{man\_j}{\sum_{k=1}^{K} man\_k}$, and$tran_{zp} = man_j - man_i$ where K is the number of maneuver intensity clusters, man_j is the next maneuver intensity cluster, and man_i is the current maneuver intensity cluster.

- Maneuver intensity

We calculate the mean of each maneuver intensity cluster to represent this driving style cluster's general intensity, as shown in (10).

$$Q(z) = \frac{1}{M} \left(\sum_{m=1}^{M} man\_i_m \right) \tag{10}$$

where M is the number of maneuvers in each driving style cluster, and man_i_m is the numerical values of the maneuver intensity cluster for the m^{th} maneuver.

Finally, we label each driving style cluster based on the above three indicators. The three indicators are all positive, with a smaller value indicating a more cautious driving style and a larger value a more aggressive driving style.

IV. RESULTS AND DISCUSSION

A. DRIVING MANEUVERS

We used a threshold-based approach to detect the driving maneuvers according to their velocity, longitudinal acceleration, and lateral acceleration. Fig. 5 shows the result.

Obviously, the lateral acceleration in turning is significantly greater than that in driving-straight. The acceleration profiles of the acceleration and deceleration maneuvers are distributed around the x axis respectively, while the acceleration in the maneuvers involving driving at a constant speed fluctuates around 0. In addition, for the longitudinal acceleration, the velocities of the different types of maneuver present trends of rising, horizontal change, and falling respectively. It can be seen from the variable profiles that the complete driving maneuvers are effectively and accurately detected in this paper.

TABLE 2. Constructed features.

Features	Code
Statistical features	{*vel, longacc, latacc*}.{mean, median, trim, std, iqr, mad, p05, p10, p25, p75, p90, p95, min, max}
Entropy features	{*vel, longacc, latacc*}.{Shannon, ApEn}

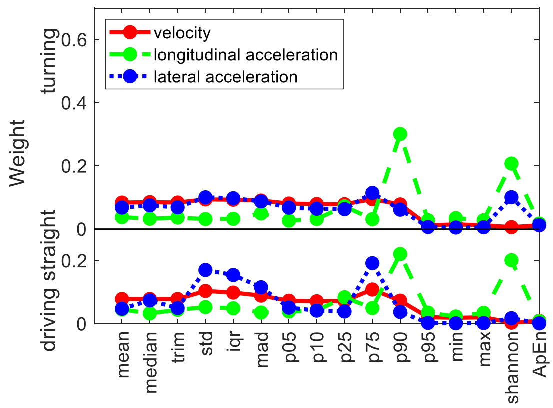


FIGURE 6. Weights of constructed features.

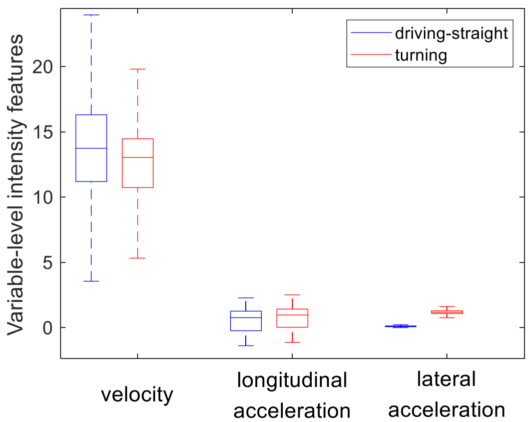


FIGURE 7. Statistical comparison between variable-level intensity features.

B. MANEUVER-LEVEL INTENSITY FEATURES

The constructed features are listed in Table 2. Because of the differences in the data distributions between turning and driving-straight, we analyze them separately.

First, we calculate the weights of each variable's constructed features based on the entropy weight method, for the driving-straight and turning, respectively. The results are shown in Fig. 6. Then, we calculate the variable-level intensity features using (2) for the driving-straight and turning. Fig. 7 shows the statistical comparison between the three variables-level intensity features.

In addition, Fig. 8 shows the distributions of the three variable-level intensity features, and Fig. 9 shows the statistical results. For the driving-straight, the velocity-level intensity features range from 0 to 28, the longitudinal acceleration-level intensity features range from −1.5 to 2.5, and the lateral acceleration-level intensity features range from 0 to 0.3. For the turning maneuvers, the velocity-level intensity features range from 0 to 23, the longitudinal acceleration-level intensity features range from −1.2 to 2.5, and the lateral acceleration-level intensity features range from 0 to 2.7.

TABLE 3. Explained variance ratio of each PC.

Principal component	Explained variance ratios	
	Driving-straight	Turning
PC 1	92.79	92.82
PC 2	6.24	6.01
PC 3	0.96	1.17

TABLE 4. Variable loadings on PC 1.

Variable	Variable loadings	
	Driving-straight	Turning
Velocity-level intensity features	0.59	0.60
Longitudinal acceleration-level intensity features	0.66	0.63
Lateral acceleration-level intensity features	0.46	0.49

Generally, velocity-level intensity features and longitudinal acceleration-level intensity features vary roughly in the same range. The velocity-level intensity features are normally distributed. However, the longitudinal acceleration-level intensity features have a bimodal distribution because they contain both positive and negative acceleration. Due to the influence of road alignment, the range of the lateral acceleration-level intensity features in driving-straight are greatly different from that in turning.

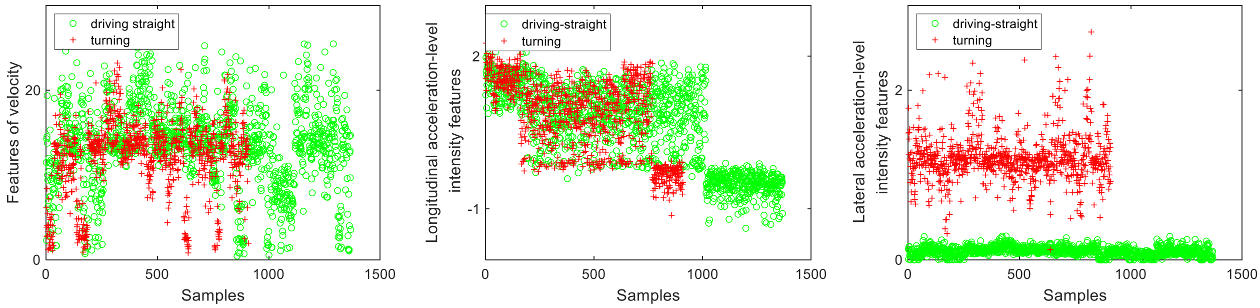
Based on the variable-level intensity features, the maneuver-level intensity features were obtained using the abovementioned PCA method. We also analyzed driving-straight and the turning respectively. Table 3 shows the explained variance ratios of the PCs. Since the explained variance ratios of the first PC of driving-straight and turning are both over 90%, we choose the first PC.

The variable loadings of the variable-level intensity features are shown in Table 4. They are consistent with the above results, being larger for the feature with the larger degree of dispersion. We calculated the maneuver-level intensity features for driving-straight and turning by (3). The results are shown in Fig. 10 and Fig. 11.

C. DRIVING MANEUVER INTENSITY

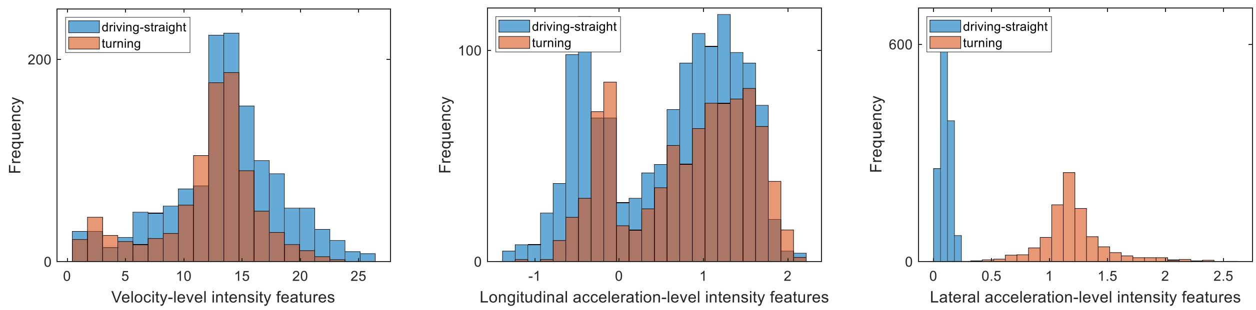
We clustered the driving maneuvers based on intensity after obtaining their intensity features, and labeled them. In this way, we amplified the similarity of similar data and the differences among different types of data to make the driving style analysis based on driving maneuvers easier.

The driving-straight and turning were still analyzed separately. We clustered the maneuvers using K-means based on their features. We selected the number of clusters to be 1, 2, 3, . . . 10 in turn, and used the Silhouette Coefficient (SC) [43]



(a) Velocity-level intensity features (b) Longitudinal acceleration-level intensity features (c) Lateral acceleration-level intensity features

FIGURE 8. Spatial distributions of variable-level intensity features.



(a) Velocity-level intensity features (b) Longitudinal acceleration-level intensity features (c) Lateral acceleration-level intensity features

FIGURE 9. Statistical results of variable-level intensity features.

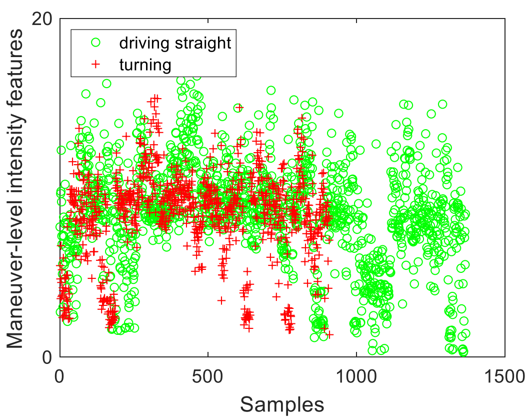


FIGURE 10. Spatial distributions of maneuver-level intensity features.

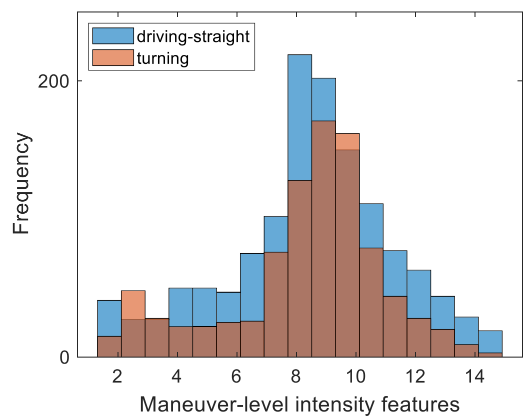


FIGURE 11. Statistical results for maneuver-level intensity features.

and Davies–Bouldin Index (DBI) [44] to measure the performance. The SC and DBI are common internal indicators used to measure clustering performance, which they evaluate by calculating the intra-cluster similarity and inter-cluster dissimilarity. The SC ranges from -1 to 1, with a larger value indicating a better result. On the contrary, with DBI, a smaller value indicates a better result.

The trends in the SC and DBI as we change the number of clusters are shown in Fig. 12. For driving-straight maneuvers, the clustering result is best when $k = 3$. Here, the SC is 0.75 and the DBI is 0.51. For turning maneuvers, the clustering result is best when $k = 2$, and the SC is 0.83 and DBI is 0.45. The above results also prove that the data distributions differ between driving-straight and turning. Fig. 13 shows the best results in detail.

As mentioned above, the maneuver-level intensity features describe the intensity of the maneuvers, so we used the statistics of maneuver-level intensity features to represent the intensity of the maneuver clusters. In addition, we used values with physical significance to represent different maneuver clusters, giving clusters with greater intensity larger values.

As shown in Table 5, for driving-straight, the overall trend is that the mean value is highest in Cluster 1, followed by Cluster 3, and lowest in Cluster 2. The maximum and

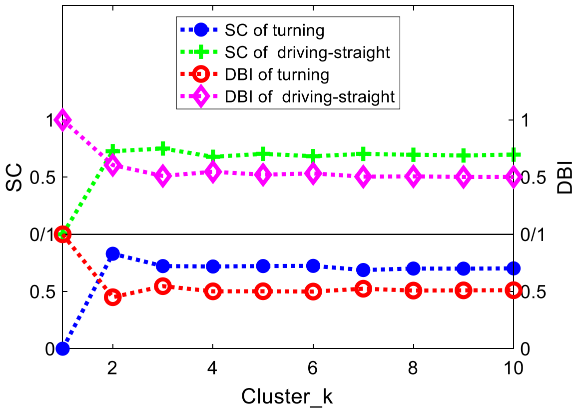


FIGURE 12. Relationship between clusters and evaluation metrics.

minimum values of the three clusters are also highest in Cluster 1, followed by Cluster 3, and lowest in Cluster 2, and the clusters all have small standard deviations. Therefore, Cluster 1 has the greatest intensity. Further, Cluster 1 has a value of 3, Cluster 2 has a value of 1, and Cluster 3 has a value of 2. For turning, the overall trend is that the mean value is highest in Cluster 2, and lowest in Cluster 1. The maximum and minimum values of the two clusters are arranged successively, and the clusters both have small standard deviations. Therefore, the intensity of Cluster 1 is weaker than that of Cluster 2. We set the value of Cluster 1 to 1 and the value of Cluster 2 to 2.

D. DRIVING STYLE

1) DYNAMIC TIME WINDOWS

We obtained 1,635 dynamic time windows based on the road alignment. The statistical results for the dynamic time window durations are shown in Fig. 14.

An example of dynamic time window determination is shown in Fig.15. The dot lines are change points of driving maneuvers, and the change of colors means different time windows. it's clear that not all the number of driving maneuvers in time windows are same and not all windows have same durations.

Two specific examples of dynamic time windows are shown in Fig. 16. The different colors represent different types of driving maneuvers. In Fig. 16(a), the longitudinal acceleration falls in the range of -2.5 m/s^2-1.1 m/s^2, which means the longitudinal maneuvers in this time window contain acceleration, deceleration, and driving at a constant speed. The radius is larger than 1,000 m, which means the lateral maneuver in this window is that of straight driving. To sum up, this window contains five maneuvers belonging to three types, namely, accelerating when driving- straight, driving at a constant speed when driving-straight, and decelerating when driving-straight. In Fig. 16(b), the longitudinal acceleration falls in the range of -3 m/s^2 - 0.6 m/s^2, and the radius is smaller than 1,000 m. Thus, this time window contains two driving maneuvers belonging to two types, namely, decelerating when turning and driving at a constant speed when turning.

2) DRIVING STYLE CLUSTERS

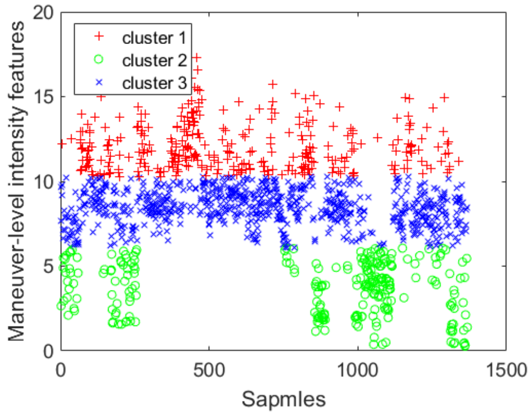
For each dynamic time window, we can obtain the time-varying pattens of driving maneuver intensity. Specifically, the time-varying patterns of driving maneuver intensity are represented by the curves of the driving maneuvers changing over time. Some examples for different time-varying patterns are shown in Fig. 17. The vertical dot lines are the change points of detected driving maneuvers, and the red lines are time-varying patterns of driving maneuver.

We then clustering driving styles in terms of time-varying patterns of driving maneuver intensity. Most studies classify driving styles into three clusters: aggressive, normal, and cautious [25], [42], [45]. We also set the number of driving clusters to three. Since the dynamic time windows were determined according to the road alignment, the differences in the data distributions between different dynamic time windows were also considered when analyzing the driving styles based on these time windows. We clustered the driving styles based on the dynamic time windows containing driving-straight maneuvers and those containing turning maneuvers separately.

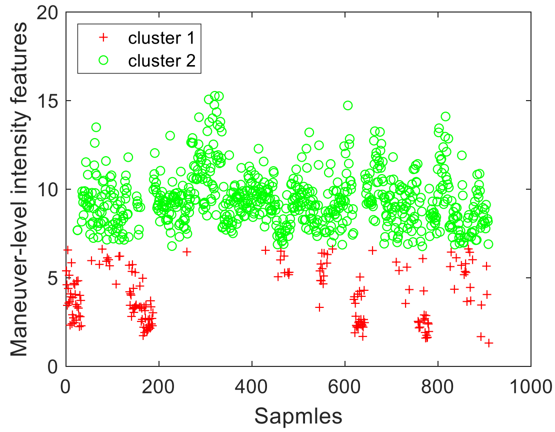
We obtained the distance matrix of the distances between pairs of curves using the DTW, and needed to choose a distance as the end threshold of SBC-DTW. However, it was difficult to determine this end threshold directly as the distances between curves are continuous.

In statistics, percentiles perfectly represent a data set, and can quantitatively reflect its distribution. Therefore, we chose a percentile of the distance matrix as the end value. A threshold selected in this way reflects the similarities and differences between the samples in a specific data set, so it had certain reliability. Further, to ensure the interpretability and acceptability of the clustering results, we selected the number of clusters to be three based on previous researches when we clustered the driving styles based on dynamic time windows. On this basis, and combined with experience, the 87th percentile was selected as the end threshold for the dynamic time windows containing driving-straight maneuvers, and the 50th percentile was selected for those containing turning maneuvers.

We also used the internal indicators to evaluate the results of the driving clustering. For the DBI, one needs to calculate the cluster centers. However, it is difficult to calculate the cluster centers of curves that have unequal lengths. As a result, we used the SC to evaluate the curve clustering algorithm, with the distances between curves obtained using the DTW. Table 6 implies that we have a good clustering result. Fig. 18 and Fig. 19 imply that we have good clustering results and statistical significances are found between the different clusters based on velocity ($p<0.01$), longitudinal acceleration ($p<0.01$), and lateral acceleration ($p<0.01$).



(a) Driving-straight with three clusters on maneuver intensity



(b) Turning with two clusters on maneuver intensity

FIGURE 13. **Results of driving maneuver clustered on maneuver intensity.**

TABLE 5. **Statistical features of driving maneuver clusters on intensity.**

Statistical features	Driving-straight			Turning	
	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2
mean	12.12	3.94	9.43	3.70	8.44
minimum	10.29	1.32	6.78	0.28	6.09
maximum	17.30	6.63	15.28	6.06	10.28
standard deviation	1.44	1.50	1.52	1.55	1.02

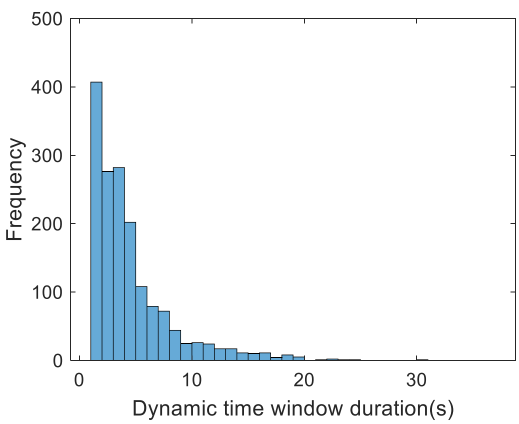


FIGURE 14. **Statistical results for dynamic time window durations.**

3) DESCRIPTIVENESS OF THE DRIVING STYLE CLUSTERS

We labeled the driving styles based on the diversity of maneuver intensity transition forms, maneuver intensity transition tendency, and maneuver intensity after clustering. The transition probabilities of the driving style clusters are shown in Fig. 20 and Fig. 21. Based on those, we used (8), (9), and (10) to obtain the diversity of maneuver intensity transition forms, maneuver intensity transition tendency, and maneuver intensity, the results of which are shown in Table 7.

For the dynamic time windows containing driving-straight maneuvers, Cluster 3 has the highest intensity and the minimum diversity of transition forms, which implies that

TABLE 6. **Results of driving style clustering.**

Types of dynamic time window	End threshold	SC
Driving-straight	87th percentile	0.72
Turning	50th percentile	0.97

Cluster 3 consists of more intense and more aggressive maneuvers. Although the maneuver intensity transition forms of Clusters 1 and 2 are more diverse, the intensity of these two clusters is lower. In other words, in Clusters 1 and 2, there is a preference for transferring between less intense maneuvers. Therefore, we label Cluster 3 as aggressive. Since the diversity of transition forms of Cluster 1 is higher than that of Cluster 2 and they have similar intensity, we label Cluster 1 as normal and Cluster 2 as cautious.

Similarly, for the dynamic time windows containing turning maneuvers, Cluster 1 is obviously aggressive, with relative high values for the three indicators. Cluster 2 is normal because it has greater intensity than Cluster 3. Cluster 3 is cautious due to the transferring between low-intensity maneuvers. To further verify the reliability of the results, we chose samples of different driving styles and compared their variable profiles. The results are shown in Fig. 22 and Fig. 23.

For dynamic time windows containing driving-straight, there is no significant difference in lateral acceleration, but the velocities of the samples with different driving styles do differ. Further, the longitudinal acceleration in this cautious sample remains almost unchanged. The normal sample

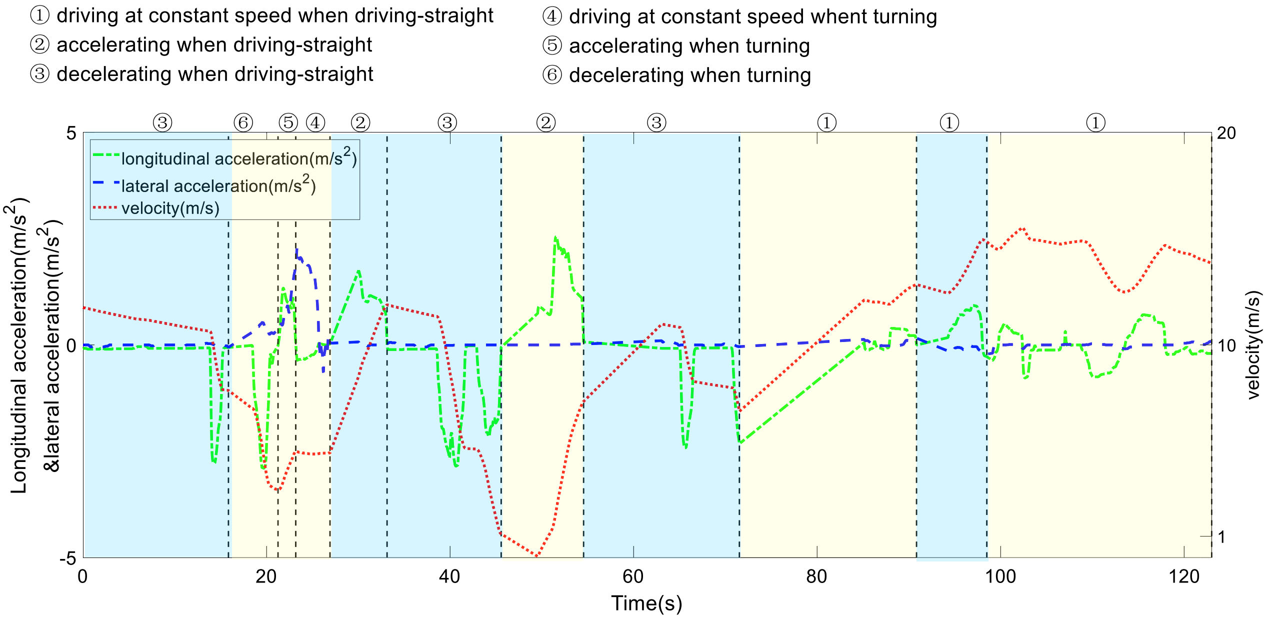
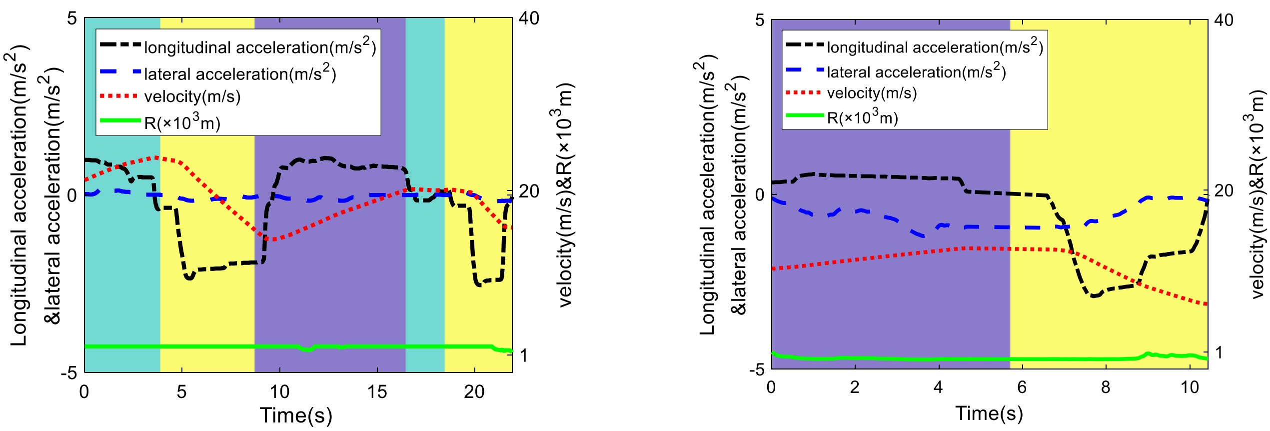


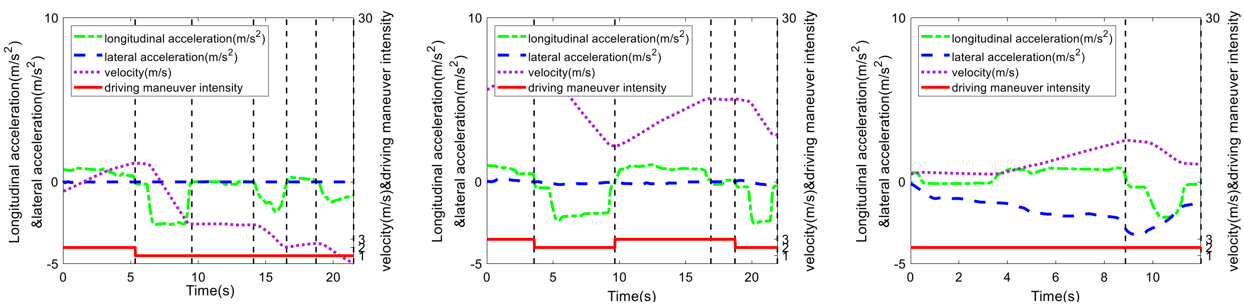
FIGURE 15. An example of dynamic time window determination: the dot lines are change points of driving maneuvers and the change of colors means different time windows.



(a) A dynamic time window with five driving maneuvers

(b) A dynamic time window with two driving maneuvers

FIGURE 16. Examples of dynamic time windows.



(a) example 1

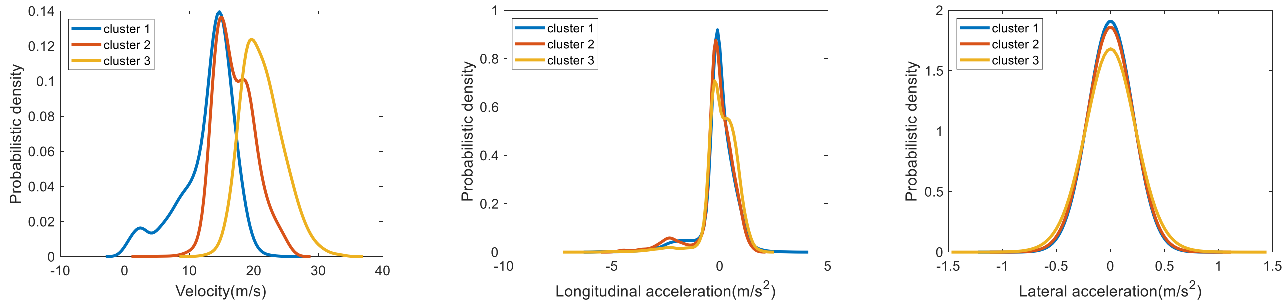
(b) example 2

(c) example 3

FIGURE 17. Examples of time-varying patterns of driving maneuver intensity.

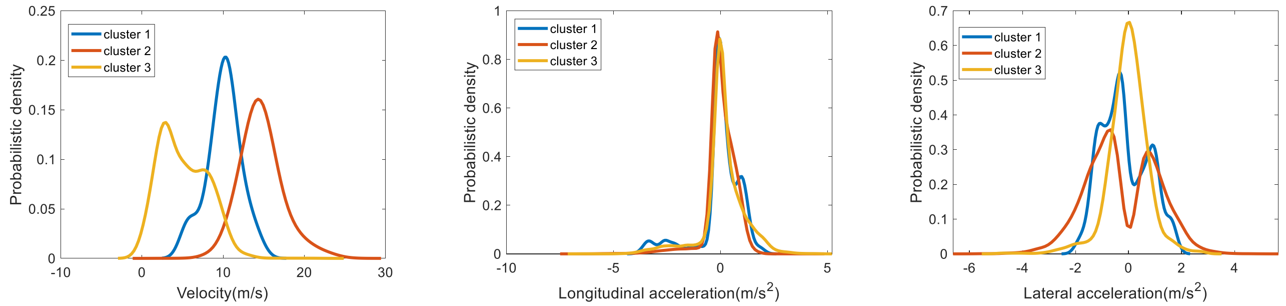
changes longitudinal acceleration clearly. The aggressive sample changes longitudinal acceleration significantly at high speeds.

For the dynamic time windows containing turning, there are differences between variables as well. We chose samples with similar lateral acceleration to keep the road



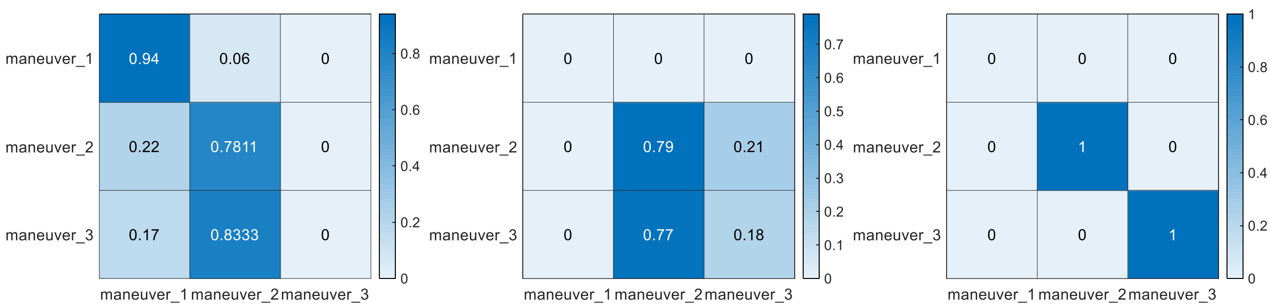
(a) Probabilistic density of velocity (b) Probabilistic density of longitudinal acceleration (c) Probabilistic density of lateral acceleration

FIGURE 18. Probabilistic density of three driving style clusters under the three variables for driving-straight.



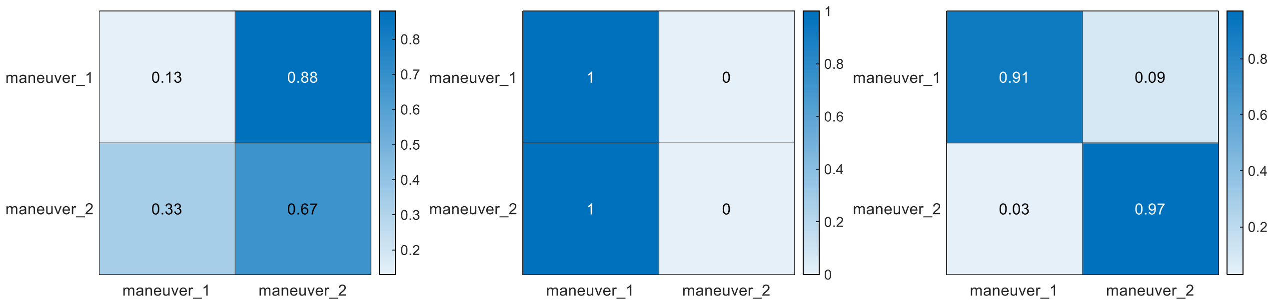
(a) Probabilistic density of velocity (b) Probabilistic density of longitudinal acceleration (c) Probabilistic density of lateral acceleration

FIGURE 19. Probabilistic density of three driving style clusters under the three variables for turning.



(a) Cluster 1 (b) Cluster 2 (c) Cluster 3

FIGURE 20. Transition probabilities of driving maneuver intensity under the three driving style clusters for driving-straight.



(a) Cluster 1 (b) Cluster 2 (c) Cluster 3

FIGURE 21. Transition probabilities of driving maneuver intensity under the three driving style clusters for turning.

alignment consistent. The velocities of the cautious samples were the lowest. The cautious samples decelerated when entering curved roads, and accelerated over time until the turning maneuver was complete. The curve of the cautious sample conforms to the truth that cautious samples firstly turn at the low velocity to reduce driving risk and slowly accelerate

TABLE 7. The three indicators for driving style labeling.

Driving style clusters	Driving-straight			Turning		
	Diversity of maneuver intensity transition forms	Maneuver intensity transition tendency	Maneuver intensity	Diversity of maneuver intensity transition forms	Maneuver intensity transition tendency	Maneuver intensity
Cluster 1	2.13	-0.33	1.73	1.46	0.67	1.60
Cluster 2	1.5	-0.33	2.22	0	-0.67	2.00
Cluster 3	0	0	2.99	0.64	0	1.01

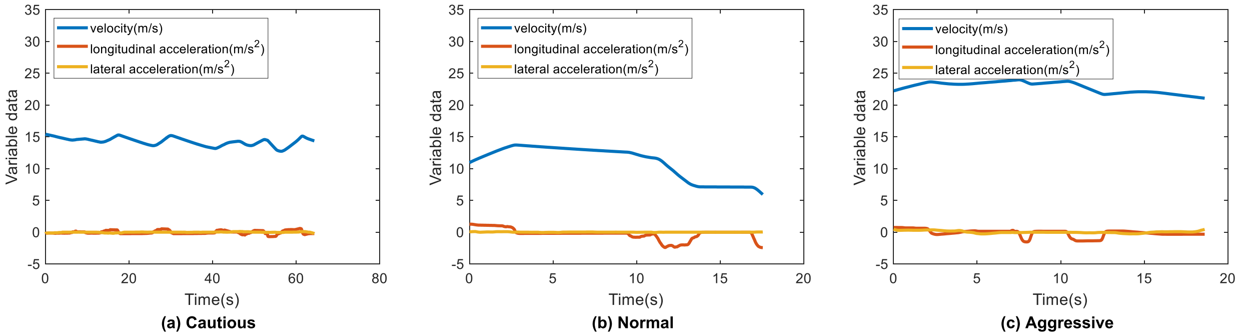


FIGURE 22. Comparison of three variable profiles under the three driving styles for driving-straight.

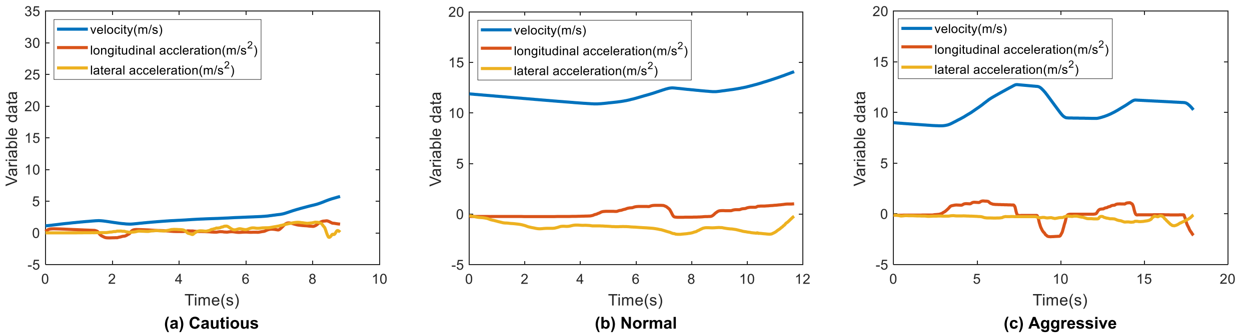


FIGURE 23. Comparison of three variable profiles under the three driving styles for turning.

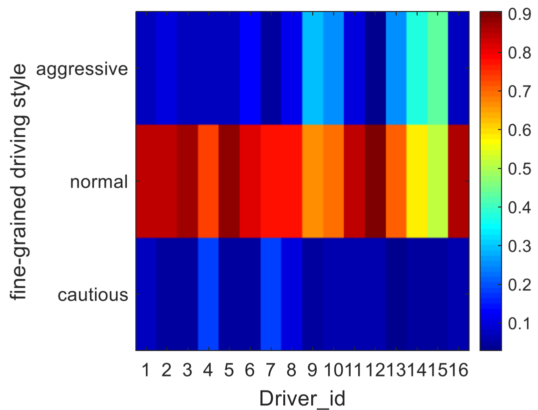


FIGURE 24. Percentages of fine-grained-driving styles for different drivers.

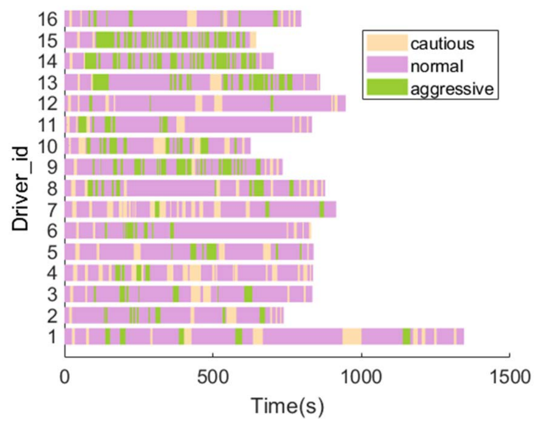


FIGURE 25. Changes in fine-grained driving styles over time for different drivers.

over time to complete the turning. However, the aggressive samples preferred to turn at a high velocity, with significant acceleration and deceleration, implying that they pursue a feeling of excitement when driving and that their driving behaviors are greatly influenced by the external environment. The normal samples passed through curves in roads using relatively modest and efficient driving behaviors.

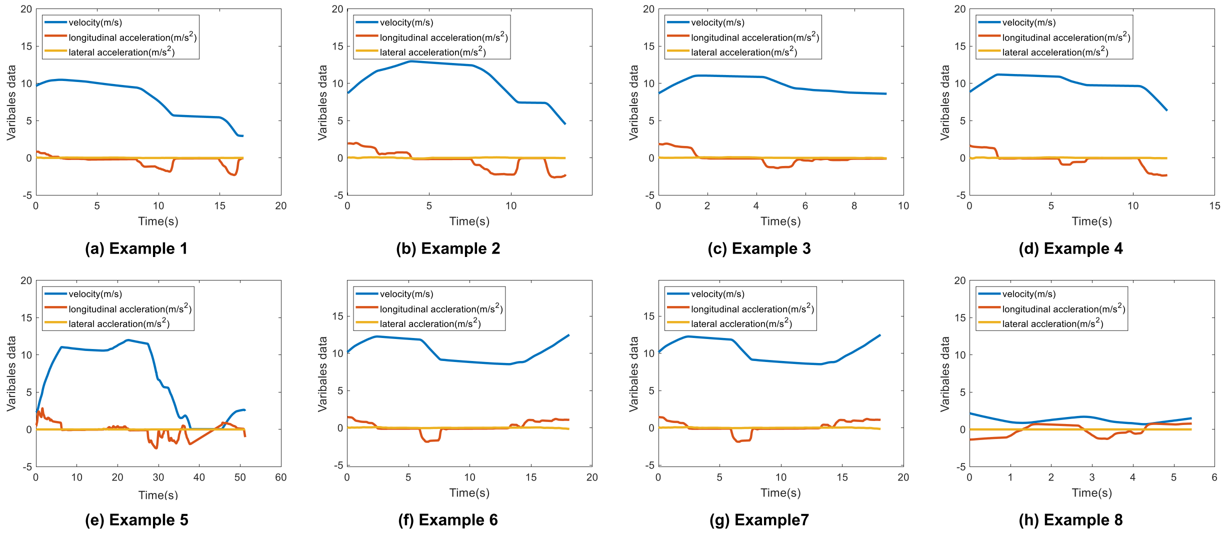


FIGURE 26. Comparison of the three variable profiles for driving-straight, obtained using different methods: (a)-(d): examples using the framework proposed in this paper; (e)-(h): examples using the classical method (K-means clustering based on discrete statistical features).

The above analysis further proves the accuracy of our method. In this paper, driving style analysis based on dynamic time windows has been carried out to describe fine-grained driving styles. For each driver, we can describe the composition of long-term driving styles they exhibit using the fine-grained driving styles, as shown in Table 8. Fig. 24 shows this too, with warmer colors implying higher proportions. We can see that, for most drivers, the normal driving style makes up the largest proportion. Several drivers (such as drivers 14 and 15) also have high proportions of the aggressive driving style, which implies that these drivers are more aggressive than most. However, the driving styles of some other drivers (such as drivers 4 and 7) are, while mostly normal, also cautious to a significantly higher degree than the other drivers, indicating that they are cautious drivers. Fig. 25 depicts the time-varying patterns of the fine-grained-driving styles of the different drivers. It can be seen that, even if their driving style proportions are similar (such as those of drivers 4 and 7), the changes over time are different. Therefore, fine-grained driving styles can be used to represent the differences in drivers' long-term driving styles, and the time-varying patterns of fine-grained driving styles can also be obtained.

E. COMPARISON WITH DRIVING STYLE ANALYSIS USING PREVIOUS METHODS

1) COMPARISON WITH DRIVING STYLE ANALYSIS USING HC-DTW

Here, we compared the results from using our curve clustering algorithm with the results from using HC-DTW. We also evaluated the clustering results using the SC. Using the control variable method, we set the number of clusters to be three for HC-DTW. The results are shown in Table 9. Compared to the results of using HC-DTW, the results of using our method have a larger SC and more balanced sizes of clusters.

TABLE 8. Composition of drivers' long-term driving styles.

Driver id	Percentage of fine-grained driving styles		
	Cautious	Normal	Aggressive
1	0.08	0.84	0.08
2	0.05	0.84	0.11
3	0.05	0.87	0.08
4	0.19	0.74	0.07
5	0.05	0.88	0.07
6	0.05	0.82	0.13
7	0.19	0.77	0.04
8	0.10	0.78	0.12
9	0.04	0.66	0.30
10	0.06	0.69	0.25
11	0.06	0.84	0.10
12	0.07	0.90	0.03
13	0.04	0.71	0.25
14	0.04	0.58	0.38
15	0.05	0.51	0.44
16	0.06	0.86	0.08

TABLE 9. Comparison of driving style analysis using different curve clustering algorithms.

Types of dynamic time window	SC		Number of samples in each cluster	
	SBC-DTW	HC-DTW	SBC-DTW	HC-DTW
Driving-straight	0.72	0.22	633；32；238	1；930；1
Turning	0.97	0.44	477；7；105	3；698；1

2) COMPARISON WITH DRIVING STYLE ANALYSIS USING THE CLASSICAL METHOD

Here, we compared the results of using our proposed framework to the results obtained using the classical method to verify our framework's superiority. In the classical method,

TABLE 10. Cluster centers of three driving styles using the classical method.

Cluster centers	Driving-straight			Turning		
	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
vel. mean	21.266	15.083	7.987	18.153	4.697	13.449
vel.std	1.366	1.016	2.025	0.808	1.169	0.921
vel.max	23.156	16.629	10.703	19.849	7.188	15.127
vel.min	18.440	13.133	4.352	17.041	2.921	12.094
longacc. mean	0.025	-0.008	-0.194	-0.023	0.167	0.066
longacc std	0.594	0.496	0.861	0.436	0.856	0.515
longacc.max	0.692	0.666	1.058	0.518	1.74	0.740
longacc.min	-1.435	-1.120	-2.067	-1.028	-1.887	-1.055
latacc. mean	0.003	-0.001	0.003	-0.283	-0.071	-0.023
latacc.std	0.134	0.068	0.024	0.520	0.453	0.355
latacc.max	0.351	0.187	0.075	0.769	0.807	0.683
latacc.min	-0.373	-0.186	-0.071	-1.308	-0.951	-0.749

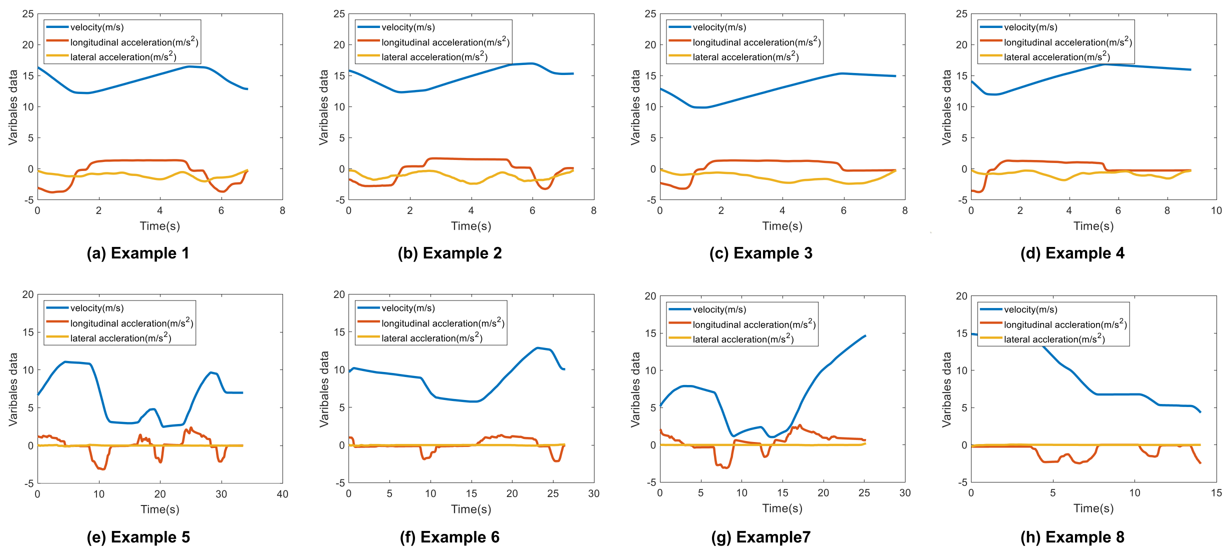


FIGURE 27. Comparison of three variable profiles for turning, obtained using different methods: (a)-(d): examples using the framework proposed in this paper; (e)-(h): examples using the classical method (K-means clustering based on discrete statistical features).

discrete statistical features were chosen to represent the driving behavior. The most used statistical features are maximum, minimum, mean value and standard deviation of variables, and the most popular and effective clustering algorithm is K-means clustering [25], [29], [32], [45]. So, we constructed the discrete statistical features, and used K-means to cluster driving styles with setting the number of clusters to three. The SC of the dynamic time windows containing driving-straight maneuvers was 0.67, and the SC of the dynamic time windows containing turning maneuvers was 0.63. The statistical features of the cluster centers are shown in Table 10.

The results of our method are better than those of the classical method (driving-straight: $0.67<0.72$; turning: $0.63<0.97$), although the advantage for the driving-straight maneuvers was not obvious. Further, we compared the results according to the variable curves of samples in the same cluster. The results are shown in Fig. 26 and Fig. 27. In Fig. 26(a), (b), (c) and (d) and Fig. 27(a), (b), (c), and (d), the profiles of the variables are similar. However, the variable profiles of Fig. 26(e), (f), (g) and (h) and Fig. 27(e), (f), (g) and (h) have significant differences and the samples in the same cluster have individual maneuvers. From above we can see that our framework proposed in this paper is superior to the classical method in terms of both clustering indicators and raw data.

V. CONCLUSION

This paper proposed a novel unsupervised driving style clustering framework based on driving maneuvers. To analyze driving style, we first presented a framework of using continuous curve features considering both the intensity and frequency of driving maneuvers. This framework is mainly consisted of driving maneuver extraction, driving maneuver feature analysis, driving maneuver clustering, and driving style analysis. The analysis results suggested that our framework could reflect the dynamic decisions of different drivers' behaviors more effectively and retain the continuity of driving data. In addition, we considered a new clustering algorithm for curves with various lengths, and compared it to the classical curve clustering algorithm. In short, the new framework we proposed is able to identify driving behavior characteristics in different environments and capture dynamic behavior decisions. The accuracy of this framework was also verified on a real data set.

This framework is particularly important in the new mixed-traffic environment in which human drivers interact with intelligent vehicles. Specifically, for intelligent vehicles, this framework can capture driving behavior's dynamic decision-making process, and help improve the effectiveness and acceptability of intelligent vehicles' decision systems. For manual vehicles, this framework can be used to analyze driving styles more accurately, which can help traffic management and insurance companies to formulate strategies.

However, some limitations still exist in this paper. There are certain subjective limitations in extracting driving maneuvers using the threshold-based algorithm. One idea for the future will be to carry out extraction maneuvers using unsupervised methods in terms of data attributes. This paper focuses on unsupervised driving style analysis but, in reality, driving style recognition is required. We will design a classifier that can recognize driving style based on the data already having driving style labels, which will provide an in-depth understanding of the potential relationship between driving style and driving behavior. We will also consider adding more environmental factors (such as traffic flow and road type) to the classifier to improve its accuracy. In this paper, we only collected driving data for 16 drivers, which means that our data are not fully representative in terms of age, gender, and experience. We plan to recruit more participants in future work.

REFERENCES

[1] F. Sagberg, G. F. B. Piccinini, and J. Engström, "A review of research on driving styles and road safety," *Hum. Factors*, vol. 57, no. 7, pp. 1248–1275, 2015, doi: 10.1177/0018720815591313.

[2] F. Guo and Y. Fang, "Individual driver risk assessment using naturalistic driving data," *Accident Anal. Prevention*, vol. 61, no. 6, pp. 3–9, Dec. 2013, doi: 10.1016/j.aap.2012.06.014.

[3] A. Kashevnik, I. Lashkov, and A. Gurtov, "Methodology and mobile application for driver behavior analysis and accident prevention," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 6, pp. 2427–2436, Jun. 2019, doi: 10.1109/tits.2019.2918328.

[4] S. Di Cairano, D. Bernardini, A. Bemporad, and I. V. Kolmanovsky, "Stochastic MPC with learning for driver-predictive vehicle control and its application to HEV energy management," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 3, pp. 1018–1031, May 2014, doi: 10.1109/TCST.2013.2272179.

[5] J. Xu, M. Saleh, and M. Hatzopoulou, "A machine learning approach capturing the effects of driving behaviour and driver characteristics on trip-level emissions," *Atmos. Environ.*, vol. 224, Mar. 2020, Art. no. 117311, doi: 10.1016/j.atmosenv.2020.117311.

[6] W. Nai, Y. Chen, Y. Yu, F. Zhang, D. Dong, and W. Zheng, "Fuzzy risk mode and effect analaysis based on raw driving data for pay-how-you-drive vehicle insurance," presented at the ICBDA, 2016. [Online]. Available: https://ieeexplore.ieee.org/document/7509806/citations#citations

[7] P. Handel and J. Ohlsson, "Smartphone-based measurement systems for road vehicle traffic monitoring and usage-based insurance," *IEEE Syst. J.*, vol. 8, no. 4, pp. 1238–1248, Dec. 2014, doi: 10.1109/JSYST.2013.2292721.

[8] C. M. Martinez, M. Heucke, F.-Y. Wang, B. Gao, and D. Cao, "Driving style recognition for intelligent vehicle control and advanced driver assistance: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 13, pp. 666–676, Mar. 2016, doi: 10.1109/TITS.2017.2706978.

[9] W. Wang, J. Xi, C. Liu, and X. Li, "Human-centered feed-forward control of a vehicle steering system based on a driver's path-following characteristics," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 6, pp. 1440–1453, Jun. 2017, doi: 10.1109/TITS.2016.2606347.

[10] C. M. Martínez, *Ihorizon-Enabled Energy Management for Electrified Vehicles || The Driver in the Loop*. Oxford, U.K.: Butterworth-Heinemann, 2019, pp. 78–79, doi: 10.1016/B978-0-12-815010-8.00003-X.

[11] T. B. A. Orit, M. Mario, and G. Omri, "The multidimensional driving style inventory-scale construct and validation," *Accident Anal. Prevention*, vol. 36, no. 3, pp. 323–332, May 2004, doi: 10.1016/S00014575(03)00010-1.

[12] M. Ishibashi, M. Okuwa, S. Doi, and M. Akamatsu, "Indices for characterizing driving style and their relevance to car following behavior," in *Proc. SICE Annu. Conf.*, Sep. 2007, pp. 1132–1137. [Online]. Available: https://ieeexplore.ieee.org/document/4421155

[13] H. Xiong, L. N. Boyle, J. Moeckli, B. R. Dow, and T. L. Brown, "Use patterns among early adopters of adaptive cruise control," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 54, no. 5, pp. 722–733, Oct. 2012, doi: 10.1177/0018720811434512.

[14] C. Guoying, "Study on identification of driver steering behavior characteristics based on pattern recognition," *Int. Robot. Autom. J.*, vol. 1, no. 1, pp. 22–28, Oct. 2016, doi: 10.15406/iratj.2016.01.00005.

[15] X. Shi, Y. D. Wong, M. Z.-F. Li, C. Palanisamy, and C. Chai, "A feature learning approach based on XGBoost for driving assessment and risk prediction," *Accident Anal. Prevention*, vol. 129, pp. 170–179, Aug. 2019, doi: 10.1016/j.aap.2019.05.005.

[16] W. R. Liu, K. Y. Deng, and X. Y. Zhang, "A semi-supervised tri-CatBoost method for driving style recognition," *Symmetry-Basel*, vol. 12, no. 3, p. 336, 2020, doi: 10.3390/sym12030336.

[17] M. M. Bejani and M. Ghatee, "A context aware system for driving style evaluation by an ensemble learning on smartphone sensors data," *Transp. Res. C, Emerg. Technol.*, vol. 89, pp. 303–320, Apr. 2018, doi: 10.1016/j.trc.2018.02.009.

[18] C. Guardiola, B. Pla, D. Blanco-Rodríguez, and A. Reig, "Modelling driving behaviour and its impact on the energy management problem in hybrid electric vehicles," *Int. J. Comput. Math.*, vol. 91, no. 1, pp. 147–156, Jan. 2014, doi: 10.1080/00207160.2013.829567.

[19] X. Zhao, W. Xu, J. Ma, H. Li, and Y. Chen, "An analysis of the relationship between driver characteristics and driving safety using structural equation models," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 62, pp. 529–545, Apr. 2019, doi: 10.1016/j.trf.2019.02.004.

[20] W. Wang, J. Xi, A. Chong, and L. Li, "Driving style classification using a semisupervised support vector machine," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 5, pp. 650–660, Oct. 2017, doi: 10.1109/THMS.2017.2736948.

[21] C. Chen, Q. Liu, X. Wang, C. Liao, and D. Zhang, "Semi-Traj2Graph: Identifying fine-grained driving style with GPS trajectory data via multi-task learning," *IEEE Trans. Big Data*, early access, Mar. 2, 2021, doi: 10.1109/TBDATA.2021.3063048.

[22] J. Chen, Z. Wu, and J. Zhang, "Driver identification based on hidden feature extraction by using adaptive nonnegativity-constrained autoencoder," *Appl. Soft Comput.*, vol. 74, pp. 1–9, Jan. 2019, doi: 10.1016/j.asoc.2018.09.030.

[23] E. Ackerman. (Oct. 20, 2019). *How Drive. AI is Mastering Autonomous Driving With Deep Learning*. [Online]. Available: https://spectrum.ieee.org/cars-thatthink/transportation/self-driving/how-driveai-is-mastering-autonomous-driv-ing-with-deep-learning

[24] A. Gacem, E. Monacelli, T. Wang, O. Rabreau, and T. Al-ani, "Assessment of wheelchair skills based on analysis of driving style," *Cognition, Technol. Work*, vol. 22, no. 1, pp. 193–207, Feb. 2020, doi: 10.1007/s10111-019-00563-6.

[25] Y. Ma, W. Li, K. Tang, Z. Zhang, and S. Chen, "Driving style recognition and comparisons among driving tasks based on driver behavior in the online car-hailing industry," *Accident Anal. Prevention*, vol. 154, May 2021, Art. no. 106096, doi: 10.1016/j.aap.2021.106096.

[26] C. Miyajima, Y. Nishiwaki, K. Ozawa, T. Wakita, K. Itou, K. Takeda, and F. Itakura, "Driver modeling based on driving behavior and its evaluation in driver identification," *Proc. IEEE*, vol. 95, no. 2, pp. 427–437, Feb. 2007, doi: 10.1109/JPROC.2006.888405.

[27] Z. Constantinescu, C. Marinoiu, and M. Vladoiu, "Driving style analysis using data mining techniques," *Int. J. Comput. Commun. Control*, vol. 5, no. 5, pp. 654–663, 2010, doi: 10.15837/ijccc.2010.5.2221.

[28] W. S. Wang and J. Q. Xi, "A rapid pattern-recognition method for driving styles using clustering-based support vector machines," presented at the ACC, Boston, MA, USA, 2016. [Online]. Available: https://ieeexplore.ieee.org/document/7526495

[29] M. V. N. de Zepeda, F. Meng, J. Su, X.-J. Zeng, and Q. Wang, "Dynamic clustering analysis for driving styles identification," *Eng. Appl. Artif. Intell.*, vol. 97, Jan. 2021, Art. no. 104096, doi: 10.1016/j.engappai.2020.104096.

[30] Y. L. Murphey, R. Milton, and L. Kiliaris, "Driver's style classification using jerk analysis," presented at the CIVVS, 2009. [Online]. Available: https://ieeexplore.ieee.org/document/4938719

[31] E. Suzdaleva and I. Nagy, "An online estimation of driving style using data-dependent pointer model," *Transp. Res. C, Emerg. Technol.*, vol. 86, pp. 22–36, Jan. 2018, doi: 10.1016/j.trc.2017.11.001.

[32] Q. Shangguan, T. Fu, J. Wang, T. Luo, and S. Fang, "An integrated methodology for real-time driving risk status prediction using naturalistic driving data," *Accident Anal. Prevention*, vol. 156, Jun. 2021, Art. no. 106122, doi: 10.1016/j.aap.2021.106122.

[33] B. Higgs and M. Abbas, "Segmentation and clustering of car-following behavior: Recognition of driving patterns," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 81–90, Feb. 2015, doi: 10.1109/TITS.2014.2326082.

[34] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jun. 1948.

[35] T. H. Itkonen and E. Lehtonen, "Characterisation of motorway driving style using naturalistic driving data," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 69, pp. 72–79, Feb. 2020, doi: 10.1016/j.trf.2020.01.003.

[36] L. Yang, R. Ma, H. M. Zhang, W. Guan, and S. Jiang, "Driving behavior recognition using EEG data from a simulated car-following experiment," *Accident Anal. Prevention*, vol. 116, pp. 30–40, Jul. 2018, doi: 10.1016/j.aap.2017.11.010.

[37] M. Subramaniyan, A. Skoogh, A. S. Muhammad, J. Bokrantz, B. Johansson, and C. Roser, "A generic hierarchical clustering approach for detecting bottlenecks in manufacturing," *J. Manuf. Syst.*, vol. 55, pp. 143–158, Apr. 2020, doi: 10.1016/j.jmsy.2020.02.011.

[38] Z. Hong, K.-H. Li, M.-J. Su, G.-M. Hu, J. Yang, G. Gao, and B. Hao, "A DTW distance-based seismic waveform clustering method for layers of varying thickness," *Appl. Geophys.*, vol. 17, no. 2, pp. 171–181, Jun. 2020, doi: 10.1007/s11770-020-0819-5.

[39] H. M. A. Aziz and S. V. Ukkusuri, "A novel approach to estimate emissions from large transportation networks: Hierarchical clustering-based link-driving-schedules for EPA-MOVES using dynamic time warping measures," *Int. J. Sustain. Transp.*, vol. 12, no. 3, pp. 192–204, Mar. 2018, doi: 10.1080/15568318.2017.1346732.

[40] P. Jancovic, M. Kokuer, and M. Zakeri, "Unsupervised discovery of acoustic patterns in bird vocalisations employing DTW and clustering," presented at the EUSIPCO, 2013. [Online]. Available: https://ieeexplore.ieee.org/document/6811728

[41] M. Ye, "Automatic classification method of large-scale curves and application," *J. Syst. Manag.*, vol. 19, no. 6, pp. 640–644, 2010.

[42] G. F. Li, "Driving style assessment using maneuver transition probabilities and driver operation aggressiveness," Ph.D. dissertation, Dept. Mech. Eng., THU, Beijing, China, 2016.

[43] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, no. 1, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.

[44] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979, doi: 10.1109/TPAMI.1979.4766909.

[45] A. Mohammadnazar, R. Arvin, and A. J. Khattak, "Classifying travelers' driving style using basic safety messages generated by connected vehicles: Application of unsupervised machine learning," *Transp. Res. C, Emerg. Technol.*, vol. 122, Jan. 2021, Art. no. 102917, doi: 10.1016/j.trc.2020.102917.

[46] O. F. Ozgul, M. U. Cakir, M. Tan, M. F. Amasyali, and H. T. Hayvaci, "A fully unsupervised framework for scoring driving style," presented at the IS, 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8710574

[47] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2010, doi: 10.1016/j.patrec.2009.09.011.

[48] K. P. Sinaga and M.-S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/access.2020.2988796.

XIAN-SHENG LI received the bachelor's degree in automobile application engineering from Jilin University, Changchun, China, in 1982, and the Ph.D. degree in vehicle operation engineering from the School of Transportation, Jilin University. His research interests include driving safety and reliability and transportation system resources optimization.

XIAO-TONG CUI was born in Zibo, Shandong, China, in 1996. She received the bachelor's degree in traffic engineering from the Harbin Institute of Technology, Weihai, Shandong, in 2018. She is currently pursuing the Ph.D. degree in vehicle operation engineering with the School of Transportation, Jilin University, Changchun, China. Her research interests include the understanding and utilizing driving behavioral data, and the driving behavior characteristics recognition.

YUAN-YUAN REN was born in Jilin City, Jilin, in 1982. She received the B.S. and M.S. degrees in traffic information engineering and control and the Ph.D. degree in vehicle operation engineering from Jilin University, in 2006, 2008, and 2011, respectively. Her research interests include driving stability and safety technology, intelligent analysis of driving behavior, and intelligent vehicle planning and control research.

XUE-LIAN ZHENG was born in Shandong, in 1987. She received the B.S. degree in automobile engineering from Northeast Forestry University, in 2009, and the Ph.D. degree in vehicle operation engineering from Jilin University, in 2014. She currently works with the School of Transportation, Jilin University, as a Vice President. Her research interests include vehicle dynamics and control, driving behavior, and autonomous vehicle design and control.