

Research Article

Finite Mixture of the Hidden Markov Model for Driving Style Analysis

Lusa Ding , Ting Zhu, Yanli Wang , and Yajie Zou 

Key Laboratory of Road and Traffic Engineering of Ministry of Education, Tongji University, Shanghai 201804, China

Correspondence should be addressed to Yanli Wang; wangyanli@tongji.edu.cn and Yajie Zou; yajiezou@hotmail.com

Received 25 August 2021; Revised 7 November 2021; Accepted 27 December 2021; Published 31 January 2022

Academic Editor: Octavian Adrian Postolache

Copyright © 2022 Lusa Ding et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Analyzing driving style is useful for developing intelligent vehicles. Previous studies usually consider the statistical features (e.g., the means and standard deviations of brake pressure) of the measured driving data or manually define the number of patterns divided by behavior semantics to characterize driving styles. In this paper, we propose a driving style analysis to describe the personalized driving styles from time-series driving data without specifying the levels in advance but by estimating them from the data. First, range, range rate, and acceleration are selected as three feature variables to describe car-following scenarios. Then, the car-following data are normalized to reduce the scale influence of different variables on the segmentation results. The hidden Markov model (HMM) and the finite mixture of the hidden Markov model (MHMM) are adopted to extract behavior semantics. Compared with the HMM, the MHMM can identify the heterogeneity of data and then provide more reasonable primitive driving patterns. Based on the results, this study uses the K -means clustering to label all the driving patterns semantically and identifies a total of 75 different driving patterns. We use the normalized frequency distributions to describe personalized driving behavior characteristics, and similarity evaluations of driving styles are applied using the Kolmogorov–Smirnov test. The proposed approach in this paper is useful for exploring the characteristics of driving habits.

1. Introduction

The human factor plays an important role in affecting traffic safety, and about 80% of traffic accidents are mainly caused by human factors [1]. The human factors include psychology, physiology, and driving styles [2]. Driving styles are habitual manipulation behaviors of drivers [3], which also influence the traffic flow. In addition, driving styles are closely related to fuel economy [4, 5], ride comfort [6], and the risk of the vehicle [7, 8]. Some researchers focus on distinguishing and classifying driving styles. Most studies generally divided drivers into several categories based on the statistical metrics of vehicle operation data. For example, Murphey et al. [9] applied an online style classification system to analyze the rate of change of acceleration and deceleration and then classified the drivers into three categories: calm, normal, and aggressive. Meanwhile, Xu et al. [10] selected the steering wheel position to describe the variability of driving styles and built a driving style analysis model by improving neural networks. Traditional studies

often simply classify all driver styles into several classes, which may neglect the potential heterogeneous characteristics of driving behaviors. Therefore, it is important to conduct a detailed analysis of personalized driving style. Personalized driving style analysis can obtain the characteristics of drivers' driving habits and help drivers clearly understand the risks of their driving behaviors so that they can be targeted to improve driving behaviors. Meanwhile, personalized driving style analysis is also a core research component of the personification development of intelligent driving technology.

To obtain a personalized driving model, substantial studies focus on statistical indicators of driving data (e.g., the means and standard deviations of brake pressure and throttle position). Shi et al. [11] used the naturalistic driving data to train the parameters of radial basis function and obtained a personalized driver model by using the locally designed neural network and the real-world vehicle test data. Wang et al. [12] proposed a driving style recognition method based on conditional kernel density function and Euclidean

distance, which divided driving styles into seven classes from normal to aggressive, and the driving style of a driver at each moment can be inferred according to probability values. These studies adopt statistical indicators of driving data to describe driving styles, which can easily capture the static driving habits but cannot obtain the dynamic driving decision-making process.

Grouping driving behaviors into simple primitive driving patterns can improve the precision of personalized driving style analysis. Nilsson et al. [13] presented a low-complexity lane-change maneuver algorithm to calculate and determine if, when, and how to perform lane-change maneuvers. Many studies define driving behavior semantics based on predefined templates. For example, Taylor et al. [14] used a dynamic time warping model to estimate the parameters of the car-following model and then described the heterogeneity of driving styles. Meanwhile, some studies defined the driving behavior semantics by the changes in indicators (e.g., standard deviation) or derived signals (e.g., hidden states of hidden Markov models). Agamennoni et al. [15] proposed a method for automatically finding the boundaries between driving sequences, inferred the parameters of a multivariate linear driving model based on the maximum likelihood, and segmented driving behavior sequences using sticky hierarchical Dirichlet process HMMs. We introduce a specific mixture of HMM to analyze time-series data. It avoids the traditional behavioral semantic segmentation method and provides a better characterization of levels for the analysis of these data while adapting to the population. In addition, the use of several HMMs makes it take into account dependency over time and thus improves the traditional method based on cutoff points [16].

The rest of this paper is organized as follows: the second section introduces the hidden Markov model and the finite mixture of the hidden Markov model. In Section 3, the data description and preliminary data analysis are introduced. Section 4 illustrates the results of behavioral semantic segmentation and personalized driving style analysis. Finally, Section 5 provides conclusions and future work for this research.

2. Methodology

2.1. HMM Model. The hidden Markov model (HMM) is a dual stochastic process, and the state process is hidden. In general, the observation sequence is used to determine the unobservable state sequence [17].

As shown in Figure 1, node x_i is the hidden state of the observation and node y_j is the observation. a_{ij} is the state transition matrix of the hidden states; b_{ij} is the probability distribution between the state sequence and the observation sequence. The HMM model usually satisfies the following three requirements. First, the number of states in the HMM model is limited. Second, the hidden states evolve dynamically according to the state transition matrix. Third, the observation sequence only depends on the state sequence.

The initial state probability vector, the state transition probability matrix, and the observation probability matrix can be described as follows:

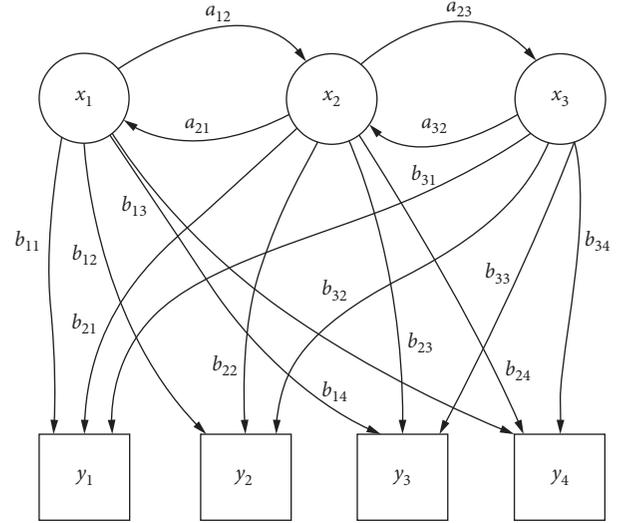


FIGURE 1: A graphical model of HMM.

$$\pi = (\pi_1, \pi_2, \dots, \pi_n),$$

$$A = [a_{ij}]_{n \times n}, \quad (1)$$

$$B = [b_{ij}]_{m \times n}$$

where π is the initial state probability vector, a_{ij} ($1 \leq i, j \leq n$) is the probability of state i which changes to state j ; b_{ij} ($1 \leq i \leq m, 1 \leq j \leq n$) is the observation probability at the time t ; n is the number of hidden states; m is the number of all possible values of the observed variable.

An HMM model is generally defined by the notation $\lambda = [A, B, \pi]$. The observation probability matrix B is the most important since it directly contributes to the observed states [18]. For this paper, the most important parts are computing the observation probability matrix and extracting the valid semantics, which can help segment drivers' behavior semantics in the car-following scene.

2.2. MHMM Model. For the driving behavior data analyzed in this study, the observation sequence is successive and contains a large amount of data. To address this issue, the finite mixture of hidden Markov model (MHMM) allows the heterogeneity of the population to be taken into account. And the probability of making an error in the partition estimation exponentially decreases with time, when the model parameters are known. The MHMM model is a combination of the hidden Markov model and the finite mixture model. Compared with other clustering methods, the finite mixture model (FMM) has the advantage that hidden groups in the data can be analyzed [19]. To split the data with different characteristics, the FMM model clusters the data into multiple homogeneous subdatasets. Meanwhile, each subdataset has its characteristic parameters and distribution type, which is especially suitable for continuous data analyzed in this paper. A FMM model with K subgroups is defined in the following equation:

$$h(y|x, \varphi) = \sum_{k=1}^K \delta_k f(y|x, \pi_k), \quad (2)$$

where x is the independent variable in a vector form; y is the dependent variable; h is the conditional density; δ_k is the mixing ratio which depends on the number of groups; $\varphi = (\delta_1, \dots, \delta_K, \bar{\pi}_1, \dots, \bar{\pi}_K)$ are all parameters in the vector form. If f is a member of the exponential families, as shown in Equation (2), we can combine the FMM model with the HMM Model to eventually obtain a MHMM model. The maximum likelihood estimation approach is adopted in this study to estimate the parameters of the MHMM model.

3. Data Description

3.1. Data Source. All car-following data in this study are from the Safety Pilot Model Deployment (SPMD) database. In 2012–2014, SPMD recorded driving behavior data from thousands of equipment vehicles in the USA. Equipment on the vehicle includes a real-time data collection system and electronic eye, and the data were all collected at a frequency of 10 Hz. This database can continuously observe driving behaviors over some time. The observation equipment is hidden from drivers, which can avoid disturbing their behaviors [20].

3.2. Data Extraction and Preprocess. On the one hand, a single variable does not characterize the driver's personalized driving style well. On the other hand, the number of variables should not be too large, which will lead to a complicated model with redundant information [21]. The variables selected from car-following events are demonstrated in [22], including the subject vehicle acceleration a_x ; the relative range between subject and lead vehicles d , $d = x_2 - x_1$, and the relative range rate v , $v = v_2 - v_1$.

For the frame dropping in the data, if the missing duration with the same lead vehicle is less than 1 second, we use linear interpolation to impute the missing data. The conditions for extracting the car-following events are as follows: (1) the same lane; (2) the relative range <120 m (the car-following events ended when the relative range less than 5 m); (3) the subject vehicle speed >18 km/h; (4) no overtaking; (5) the duration >50 s [23].

3.3. Variable Segmentation and Threshold Selection. In the process of training and testing a model, to avoid the value scale impact of the three feature variables on the behavioral semantic segmentation results [24], we normalized the feature variables in terms of drivers so that the mean was 0 and the standard deviation was 1. Then, we used the HMM model and the MHMM model presented in this section to segment drivers' behavior semantics in the car-following scenario, thus analyzing the personalized car-following behavior characteristics further.

According to the physiological and psychological perception thresholds of drivers, all characteristic variables are divided into different levels to facilitate the semantic

interpretation of the primitive driving patterns. In this way, the same driving behavior semantics extracted from different drivers can characterize the same driving patterns.

Based on the quantile characteristics of feature variables and the driver's comfortable thresholds, the relative distances are classified into three classes [25]: {LD, ND, CD}, and the relative speeds and accelerations are classified into five classes [26] {RFB, FB, KE, CI, RCI} and {AA, GA, NA, GD, AD}, respectively. More specifically, in order to clearly make a semantic explanation for primitive driving patterns, we classify each variable into different levels based on drivers' physical and psychological perception thresholds corresponding to their statistical feature. The variable segmentation is shown in Table 1.

4. Modeling Result Analysis

4.1. Segmentation Results and Comparisons. The behavioral semantic segmentation results are obtained based on the hidden state sequences of HMM and MHMM. Considering the length limitation of the paper, we only show the segmentation results from the representative trajectory of driver #1, as shown in Figure 2. The horizontal coordinate shows the time in 0.1 s, the background color blocks are the extracted driving behavior sequence units, and each color represents a different type of driving behavior semantics. Then, we make a further discussion and analysis on the segmentation results of different models as follows:

Figure 2 shows an example of the segmentation results using these two methods. As seen in Figure 2, it can be found that HMM can divide the driving data into different behavioral semantic units according to the data distribution characteristics, but this model is too sensitive to data fluctuation, which leads to many behavioral semantic units with a duration of less than 1.0 s, for example, at time 5.9 s, 26.2 s, 26.5 s, 45.2 s, 49.2 s, and 53.9 s with durations of 0.1 s, 0.3 s, 0.5 s, 0.5 s, 0.5 s, 0.4 s, and 0.1 s, respectively. These do not correspond to the real driving conditions. Drivers hardly adjust the driving maneuvers in such a short time (e.g., accelerating for 0.2 s and then taking the brake) [27].

The MHMM model can also divide the driving data into different behavioral semantic patterns. The behavioral semantic segmentation results can accurately extract effective behavioral semantic sequences based on the data distribution characteristics. At the same time, the MHMM model can effectively avoid the effect of random noise in data, resulting in very short behavioral semantic segments (i.e., durations are less than 1.0 s). It shows that the MHMM model not only divides time-series driving data into segments but also keeps the behavioral semantic pattern within a reasonable duration. Therefore, in the next section, the personalized driving styles are interpreted based on the MHMM results.

4.2. Labeling Behavior Semantics. In this paper, MHMM is used to segment the driving behavior semantics in terms of car-following events. Different car-following events

TABLE 1: Variable segmentation of feature variables.

Variable	Variable state	Threshold
Relative range (m)	Long distance (LD)	>59.33
	Normal distance (ND)	(27.04, 59.33]
	Close distance (CD)	(5, 27.04]
Relative range rate (m/s)	Rapidly closing in (RCI)	<-1.3
	Closing in (CI)	[-1.3, -0.2)
	Keeping (KE)	[-0.2, 0.2)
	Falling behind (FB)	[0.2, 1.0]
	Rapidly falling behind (RFB)	>1.0
Acceleration (m/s ²)	Aggressive acceleration (AA)	>0.19
	Gentle acceleration (GA)	(0.0522, 0.19]
	No acceleration (NA)	(-0.0642, 0.0522]
	Gentle deceleration (GD)	[-0.2, -0.0642]
	Aggressive deceleration (AD)	<-0.2

demonstrate distinct conditions. Thus, it is expected that different driving behavior semantics characterized by the same states may be extracted from multiple drivers. Therefore, we refer to behavior semantics defined in Table 1 and label them, which are based on the driving behavior semantic segmentation results of MHMM and the mean value of feature variables in each behavior semantic segment. To describe the semantic features concisely and capture the dynamic change regularity of behavior semantics, we use the following abbreviations to label the semantic category labels: $S_{\Delta v} \in \{RCI, CI, KE, FB, RFB\}$, $S_{a_x} \in \{AA, GA, NA, GD, AD\}$, $S_{\Delta d} \in \{LD, ND, CD\}$, where $S_{\Delta d}$, $S_{\Delta v}$, and S_{a_x} denote the semantic labels of relative range, relative range rate, and acceleration, respectively, and the meaning and value range of each label are the same as shown in Table 1.

Based on the results of MHMM, some car-following driving patterns are easily labeled with behavior semantics, while some others are not. To clearly label the states, we find common characteristics in the car-following driving patterns and then cluster the data of each segment into a single point, which can represent this car-following driving pattern. We choose the K -means clustering method to cluster the driving data, which is widely used [28–30]. As one of the methods to analyze and extract feature parameters from a large amount of raw data, clustering methods aim to classify the data objectively and stably. In this case, objective means that we can obtain the same results in each set of driving data by using the same methods. Stable means that the classification process remains constant across various drivers.

Based on the behavioral semantic segmentation results of the car-following event in Figure 2, Table 2 shows central values of the K -means clustering results for driver #1 and standard deviations of three variables for each segment. According to the clustering results, the labels of each semantic segment can be labeled. The results in Table 2 further demonstrate the advantages of MHMM; the reasons are as follows: (1) the standard deviation of each car-following segment is relatively small; (2) feature variables of the same behavioral semantic category take values close to each other; (3) there are significant differences in the values of feature variables taken by different behavioral semantic categories.

4.3. Normalized Frequency Distribution of Driving Style.

To analyze personalized driving styles, we interpret the driving behavior semantics as a driver unit. In this way, we can not only establish the mapping relation between the car-following driving states and driving patterns but also obtain the results of driver's behavior preference and the transform rules of driving patterns by aggregate analysis.

In this paper, we describe personalized driving behavior characteristics by using the normalized frequency distribution of driving behavior semantics, which is different from traditional statistical indicators such as the mean and standard deviation of feature variables, and normalized frequency distribution can help us intuitively analyze driving styles. Based on the labels of relative range, the semantic segments are divided into three datasets $\{LD, ND, CD\}$ and then compute the normalized probability of each distance pattern. And the normalized probability of each distance pattern is computed by the following equation:

$$p_{\blacksquare, * }^{(m)} = \frac{n_{\blacksquare, *}^{(m)}}{\sum_{\blacksquare} n_{\blacksquare, *}^{(m)}}, \quad (3)$$

$$\blacksquare = [S_{\Delta v}, S_{\Delta d}],$$

$$* = S_{\Delta d},$$

where $*$ is the relative distance labels $\{LD, ND, CD\}$; \blacksquare is the combination of 5 relative range rate labels and 5 acceleration labels, a total of 25 types; $n_{\blacksquare, *}^{(m)}$ is the number of points in one label, where the semantic label is \blacksquare and the relative distance label is $*$; $p_{\blacksquare, *}^{(m)}$ is the normalized probability of driving behavior semantics, where the semantic label is \blacksquare and the relative distance label is $*$. The sum of the normalized probability $f_{\blacksquare, *}^{(m)}$ of driving behavior semantic is equal to 1 in each driver.

Table 3 shows the normalized frequency distribution of driving behavior semantics for driver #1. The meaning of semantic labels is provided in Table 1. As shown in Table 3, we can find that the trend of relative range changes has a clear impact on driving patterns. When the relative range rapidly increases, drivers prefer to accelerate, and the frequency of acceleration is higher than deceleration; when the relative range rapidly decreases, drivers prefer to decelerate, and the frequency of deceleration is higher than acceleration.

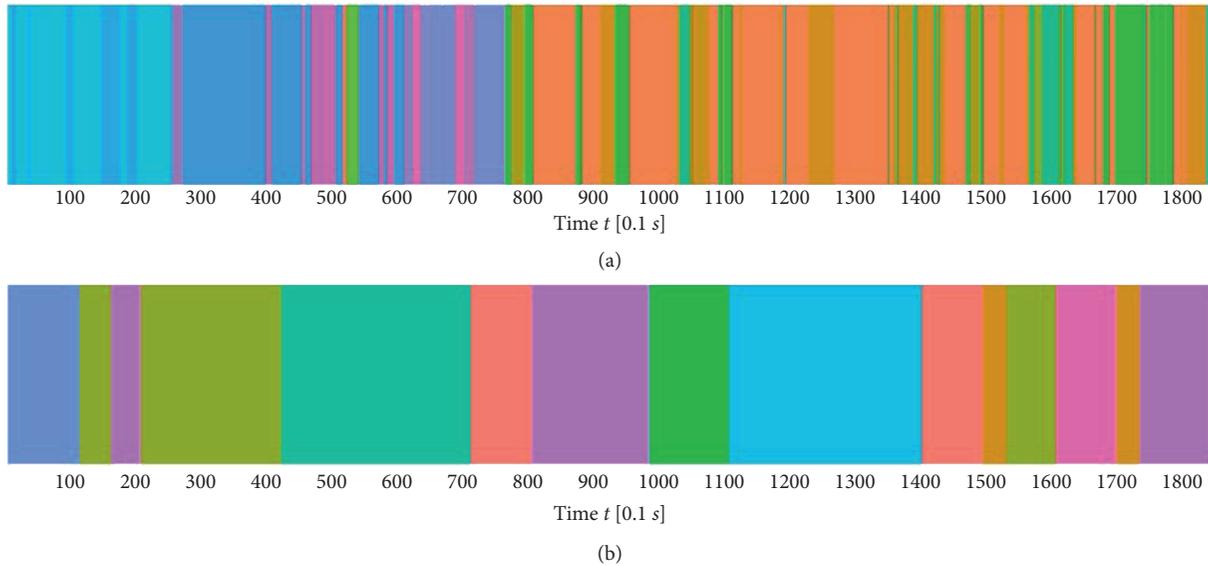


FIGURE 2: Example of segmentation results of one event for driver #1 using two different models: from top to bottom are HMM (a), MHMM (b).

TABLE 2: K -means centers of driver #1.

Label	Range	Speed	Accel	SD range ^a	SD speed ^b	SD accel ^c
ND-CI-AD	29.80185	-0.26538	-0.22313	0.15853	0.19988	0.15684
ND-KE-AA	29.52291	-0.10909	0.25414	0.09957	0.18234	0.15768
ND-CI-NA	28.78316	-0.51579	0.04505	0.15799	0.03746	0.17498
ND-CI-GA	28.20534	-0.63333	0.16375	0.18831	0.13707	0.13919
CD-CI-GD	26.56686	-0.62857	-0.18614	0.19853	0.12044	0.19391
CD-CI-NA	25.18637	-0.50556	0.03352	0.51508	0.09599	0.16900
CD-CI-GD	24.45029	-0.40000	-0.18060	0.18381	0.09608	0.21643
CD-KE-GD	22.61867	0.01111	-0.18276	0.16289	0.09634	0.24673
CD-KE-GA	22.47040	0.02667	0.07677	0.03245	0.07037	0.20655
CD-CI-NA	21.21973	-0.38500	0.01663	0.61539	0.06846	0.20971
CD-KE-AD	19.98133	-0.10833	-0.21778	0.03188	0.09003	0.18506
CD-FB-NA	21.05853	0.41053	-0.02336	0.99289	0.14292	0.16675
CD-FB-AA	25.50710	0.54194	0.24808	1.25977	0.16078	0.20842
CD-KE-AA	23.78611	-0.04474	0.26866	1.14383	0.14832	0.15873
CD-KE-AD	22.94667	-0.10000	-0.29215	0.02881	0.15374	0.10402

SD range^a denotes the standard deviation of relative range. SD speed^b denotes the standard deviation of relative range rate. SD accel^c denotes the standard deviation of acceleration.

TABLE 3: Normalized frequency distributions of driving patterns for driver #1.

Label		RFB	FB	KE	CI	RCI
Close distance	AA	0.00532	0.00159	0.01302	0.00391	0.00000
	GA	0.00000	0.01504	0.00104	0.01718	0.00000
	NA	0.00000	0.00838	0.00055	0.01369	0.00000
	GD	0.00073	0.00293	0.00819	0.00703	0.00141
	AD	0.00165	0.00336	0.00709	0.03711	0.00538
Normal distance	AA	0.00917	0.04322	0.00624	0.01217	0.00214
	GA	0.02256	0.11432	0.05575	0.03888	0.02146
	NA	0.01198	0.09158	0.05184	0.08332	0.00403
	GD	0.00226	0.01553	0.02237	0.02335	0.00960
	AD	0.01192	0.00844	0.00520	0.07244	0.02506
Long distance	AA	0.00098	0.00000	0.01761	0.00422	0.00000
	GA	0.00153	0.00006	0.00006	0.00281	0.00391
	NA	0.00196	0.00098	0.01296	0.02189	0.00153
	GD	0.00073	0.00006	0.00000	0.00507	0.00000
	AD	0.00043	0.00000	0.00141	0.00238	0.00000

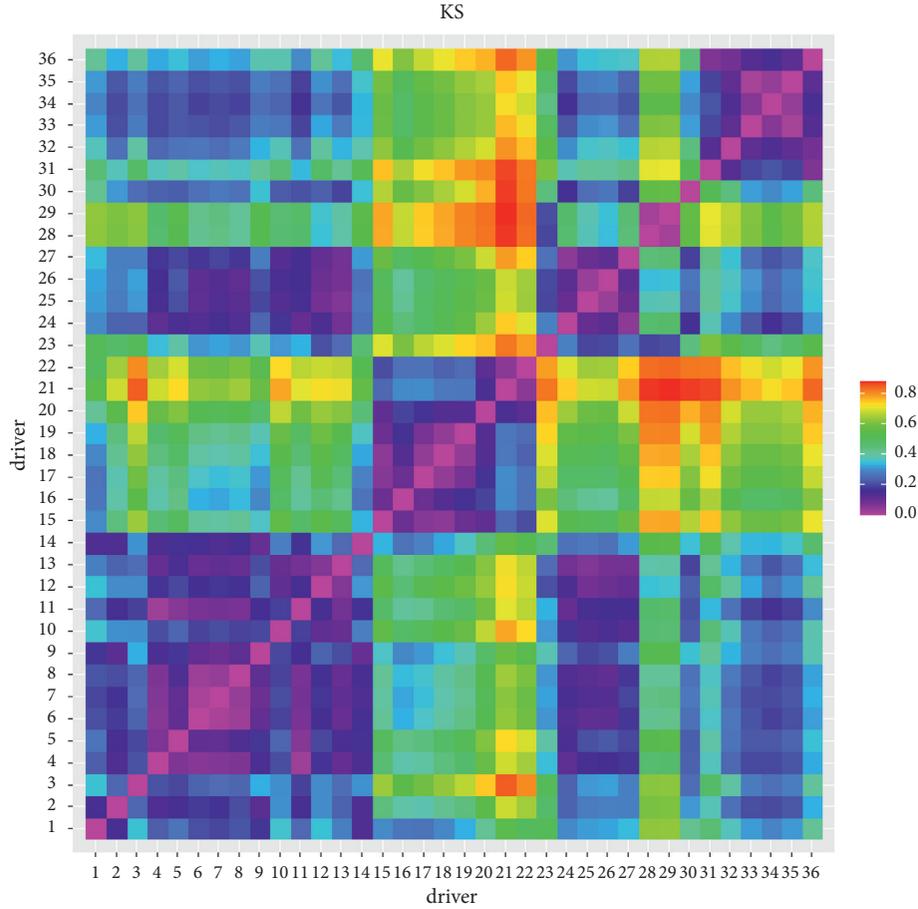


FIGURE 3: KS test between two normalized distributions of driving patterns among all drivers.

For example, when following a lead vehicle at a normal distance, driver #1 prefers to take a gentle acceleration operation and falls behind (ND-FB-GA), as shown in Table 3. The same is for other situations (long distance and close distance), which indicates that the dynamic changes of relative distance have a direct effect on the driver.

4.4. Similarity Evaluation of Driving Style. Similarity evaluation of driving style plays an important role in personalized driving style analysis. To reduce the difficulty of the personalized customization on the driving assistance system, we need to compute the similarity of different driving styles reasonably and allow the drivers with similar driving styles to share one class of the driving assistance pattern. However, the randomness and dynamic of driving styles are challenging for similarity evaluation. Furthermore, driving behavior is usually a highly nonlinear process; thus, it is not easy to directly evaluate the similarity between two drivers with continuous observation sequences. The common approaches can calculate statistical indicators directly [31], but they cannot evaluate the similarity of driving styles in terms of randomness and dynamic. Therefore, we use the Kolmogorov–Smirnov (KS) test [32, 33] to compute the similarity of two drivers in driving styles and illustrate the differences between drivers. The steps of the KS test are as follows:

Step 1. Assume that X_1, \dots, X_n and Y_1, \dots, Y_m are independent identically distributed samples from the population distribution function, $F(x)$ and $G(x)$, respectively, and all the samples are independent.

Step 2. The null and alternative hypotheses are defined as follows:

$$\begin{aligned} H_0: F(x) &= G(x), \quad \forall x \in (-\infty, +\infty), \\ H_1: F(x) &\neq G(x), \quad \exists x \in (-\infty, +\infty). \end{aligned} \quad (4)$$

Step 3. Construct test statistics: The empirical distribution functions of X and Y are $S_X(x)$ and $S_Y(y)$, respectively. Since the empirical distribution function of the sample is a good estimate of the population distribution, $|S_X(x) - S_Y(y)|$ should be small or tend to 0 for each value of x when the null hypothesis H_0 is well established as the test statistic is computed by

$$T = \sup |S_X(x) - S_Y(y).| \quad (5)$$

We can compare the difference between $F(x)$ and $G(x)$, and the rejection region is $\{T \geq c\}$.

Step 4. Set the significance level and determine the rejection region. The distribution of T can be divided into three cases: small samples with unequal capacities, small samples with equal capacities, and large samples.

In the first two cases of small samples, the critical value of the test can be obtained by Table 1. In the case of large samples, according to the sample sizes m and n and the significance level α , the quantile $T_{1-\alpha}$ can be obtained by

$$T_{1-\alpha} = \lambda \sqrt{\frac{m+n}{mn}}, \quad (6)$$

where λ is the coefficient related to α , normally α is 0.05 or 0.01, the corresponding λ is 1.32 or 1.63, and the rejection region of test is $W = \{T \geq T_{1-\alpha}\}$.

Step 5. Calculate the value of the statistic T . On the one hand, if $T \geq T_{1-\alpha}$, we can indicate that the sample data X_1, \dots, X_n and Y_1, \dots, Y_m come from different population distributions, and a significant difference can be found between X and Y . On the other hand, if $T < T_{1-\alpha}$, we have no idea to reject the null hypothesis, which means that the two population distributions are the same.

A heat map presents the KS test between two normalized distributions of driving patterns among all drivers, as shown in Figure 3. The red color represents a huge difference between the driving styles of the two drivers, and purple means that the two drivers are quite similar to each other. By the way, the values of the statistic T are equal to 0, which are drawn with dark purple color.

From Figure 3, on the one hand, we know that the occurrence probability of behavior semantics for driver #21 is significantly different from many other drivers, especially different from some drivers (i.e., drivers #3, #10, #23, #28, #29, #30, etc). On the other hand, driver #13 is similar to others, especially similar to drivers #4, #6, #7, #8, #10, etc. Last but not least, drivers #15 ~ #23 are significantly similar to each other, as they have blocks of similar color. In summary, the KS test can compare the degree of similarity or dissimilarity of driving styles for two drivers.

5. Conclusions

In this paper, we select the MHMM model to label behavior semantics. The normalized frequency distribution of behavior semantics is adopted to analyze driving style. Feature variables are divided into different behavioral semantic patterns, and the similarity of different driving styles is also compared through this model. The main findings can be summarized as follows:

- (1) Compared with the conventional HMM model, the MHMM model can better recognize the driving patterns. MHMM can identify the heterogeneity of data and divide the car-following data into different behavioral semantics. The MHMM model can also provide reasonable behavioral semantic segmentation results.
- (2) We can adopt the normalized frequency distribution of driving behavior semantics to analyze personalized driving styles, instead of statistical indicators,

which allows us to intuitively analyze driving styles. The normalized frequency distribution of driving behavior semantics is consistent with the results of previous studies and requires less computational time.

- (3) The KS test can quantitatively evaluate the similarity and dissimilarity of driving styles for different drivers. This method can be useful for identifying personalized driving styles.

For future work, we consider expanding the sample size to study the impact of different factors (e.g., gender, age, and vehicle type) on driving styles. In addition, we should consider more complicated traffic scenarios to further analyze the driving styles.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

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. 2120210009).

References

- [1] L. Aarts and I. van Schagen, "Driving speed and the risk of road crashes: a review," *Accident Analysis & Prevention*, vol. 38, no. 2, pp. 215–224, 2006.
- [2] J. Elander, R. West, and D. French, "Behavioral correlates of individual differences in road-traffic crash risk: an examination of methods and findings," *Psychological Bulletin*, vol. 113, no. 2, pp. 279–294, 1993.
- [3] T. Lajunen and T. Özkan, "Self-report instruments and methods," *Handbook of Traffic Psychology*, Elsevier, Amsterdam, Netherlands, pp. 43–59, 2011.
- [4] R. Langari and J. S. Won, "Intelligent energy management agent for a parallel hybrid vehicle - Part 1: system architecture and design of the driving situation identification process," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 3, pp. 925–934, 2005.
- [5] J. E. Meseguer, C. K. Toh, C. T. Calafate, J. C. Cano, and P. Manzoni, "DrivingStyles: a mobile platform for driving styles and fuel consumption characterization," *Journal of Communications and Networks*, vol. 19, no. 2, pp. 162–168, 2017.
- [6] H. Bellem, T. Schoenberg, J. F. Krems, and M. Schrauf, "Objective metrics of comfort: developing a driving style for highly automated vehicles," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 41, pp. 45–54, 2016.
- [7] T. H. Huang, V. Nikulin, and L. B. Chen, "Detection of abnormalities in driving style based on moving object

- trajectories without labels,” in *Proceedings of the 2016 fifth IEEE/IIAI International Congress on Advanced Applied Informatics*, Kumamoto, Japan, July 2016.
- [8] Y. Zou, B. Lin, X. Yang, L. Wu, M. Muneeb Abid, and J. Tang, “Application of the bayesian model averaging in analyzing freeway traffic incident clearance time for emergency management,” *Journal of Advanced Transportation*, vol. 2021, no. 4, 9 pages, Article ID 6671983, 2021.
 - [9] Y. L. Murphey, R. Milton, and L. Kiliaris, “Driver’s style classification using jerk analysis,” in *Proceedings of the IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems*, pp. 23–28, Nashville, TN, USA, April 2009.
 - [10] L. Xu, J. Hu, H. Jiang, and W. Meng, “Establishing style-oriented driver models by imitating human driving behaviors,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2522–2530, 2015.
 - [11] B. Shi, L. Xu, J. Hu et al., “Evaluating driving styles by normalizing driving behavior based on personalized driver modeling,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 45, no. 12, p. 1, 2015.
 - [12] W. Wang, D. Zhao, W. Han, and J. Xi, “A learning-based approach for lane departure warning systems with a personalized driver model,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 10, pp. 9145–9157, 2018.
 - [13] J. Nilsson, J. Silvlin, M. Brannstrom, E. Coelingh, and J. Fredriksson, “If, when, and how to perform lane change maneuvers on highways,” *IEEE Intelligent Transportation Systems Magazine*, vol. 8, no. 4, pp. 68–78, 2016.
 - [14] J. Taylor, X. Zhou, N. M. Roupail, and R. J. Porter, “Method for investigating intradriver heterogeneity using vehicle trajectory data: a Dynamic Time Warping approach,” *Transportation Research Part B: Methodological*, vol. 73, pp. 59–8, 2015.
 - [15] G. Agamennoni, S. Worrall, J. R. Ward, and M. N. Eduardo, “Automated extraction of driver behaviour primitives using bayesian agglomerative sequence segmentation,” in *Proceedings of the 2014 IEEE seventeenth International Conference on Intelligent Transportation Systems (ITSC)*, Qingdao, China, October 2014.
 - [16] W. Vitali, F. Ronja, P. Yannis, I. Pigeot, and N. Wirsik, “Using hidden Markov models to improve quantifying physical activity in accelerometer data—a simulation study,” *PLoS One*, vol. 9, no. 12, Article ID e114089, 2014.
 - [17] B. Schuster-Bockler and A. Bateman, “An introduction to hidden Markov models,” *Current protocols in bioinformatics*, vol. 3, 2007.
 - [18] S. Zhang, C. Cao, and A. Quinn, “Dynamic analysis on simultaneous iEEG-MEG data via hidden Markov model,” *NeuroImage*, vol. 233, 2021.
 - [19] L. E. Baum and T. Petrie, “Statistical inference for probabilistic functions of finite state Markov chains,” *The Annals of Mathematical Statistics*, vol. 37, no. 6, pp. 1554–1563, 1966.
 - [20] G. M. Fitch and R. J. Hanowski, *Using Naturalistic Driving Research to Design, Test and Evaluate Driver Assistance Systems*, Springer, London, UK, 2012.
 - [21] H. Al Reesi, A. Al Maniri, K. Plankermann et al., “Risky driving behavior among university students and staff in the Sultanate of Oman,” *Accident Analysis & Prevention*, vol. 58, pp. 1–9, 2013.
 - [22] W. Wang, J. Xi, and D. Zhao, “Driving style analysis using primitive driving patterns with bayesian nonparametric approaches,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 8, pp. 2986–2998, 2019.
 - [23] W. Wang, C. Liu, and D. Zhao, “How Much Data Are Enough? A Statistical Approach with Case Study on Longitudinal Driving Behavior,” *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 2, pp. 85–98, 2017.
 - [24] F. D. u, Y. Zhao, G. Li, and W Wang, “Evaluation of the autonomous parking system based on BP neural network,” in *Proceedings of the 2017 ninth International Conference on Intelligent Human-Machine Systems and Cybernetics*, Hangzhou, China, August 2017.
 - [25] K. I. Ahmed, “Modeling Drivers’ Acceleration and Lane Changing Behavior,” *Massachusetts Institute of Technology*, 1999.
 - [26] C. MacAdam, Z. Bareket, P. Fancher, and R. Ervin, “Using neural networks to identify driving style and headway control behavior of drivers,” *Vehicle System Dynamics*, vol. 29, pp. 143–160, 1998.
 - [27] H.-H. Yang and H. Peng, “Development of an errorable car-following driver model,” *Vehicle System Dynamics*, vol. 48, no. 6, pp. 751–773, 2010.
 - [28] W. Wang, J. Xi, and Ieee, “A Rapid Pattern-Recognition Method for Driving Styles Using Clustering-Based Support Vector Machines,” in *Proceedings of the 2016 American Control Conference*, pp. 5270–5275, Boston, MA, USA, July 2016.
 - [29] S. Wang, Y. Zhang, and C. Wu, “Online prediction of driver distraction based on brain activity patterns,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 136–150, 2015.
 - [30] 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.
 - [31] S. Lefevre, A. Carvalho, and F. Borrelli, “A learning-based framework for velocity control in autonomous driving,” *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 1, pp. 32–42, 2016.
 - [32] H. Hassani and E. Silva, “A Kolmogorov-smirnov based test for comparing the predictive accuracy of two sets of forecasts,” *Econometrics*, vol. 3, no. 3, pp. 590–609, 2015.
 - [33] L. Al-Labadi and M. Zarepour, “Two-sample Kolmogorov-smirnov test using a bayesian nonparametric approach,” *Mathematical Methods of Statistics*, vol. 26, no. 3, pp. 212–225, 2017.