Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 3024101, 13 pages
https://doi.org/10.1155/2020/3024101

Research Article

Road Traffic Safety Risk Estimation Method Based on Vehicle Onboard Diagnostic Data

Xiaoyu Cai,[1,2] **Cailin Lei**,[1,2] **Bo Peng**,[1,2] **Xiaoyong Tang**,[3] **and Zhigang Gao**[3]

[1]*Chongqing Jiaotong University, College of Traffic and Transportation, Chongqing 400074, China*
[2]*Key Laboratory of Traffic System & Safety in Mountain Cities of Chongqing, Chongqing 400074, China*
[3]*Urban Transportation Big Data Engineering Technology Research Center of Chongqing, Chongqing 400020, China*

Correspondence should be addressed to Xiaoyu Cai; caixiaoyu@cqjtu.edu.cn

Received 7 September 2019; Revised 24 December 2019; Accepted 17 January 2020; Published 26 February 2020

Guest Editor: Bilal Farooq

Currently, research on road traffic safety is mostly focused on traffic safety evaluations based on statistical indices for accidents. There is still a need for in-depth investigation on preaccident identification of safety risks. In this study, the correlations between high-incidence locations for aberrant driving behaviors and locations of road traffic accidents are analyzed based on vehicle OBD data. A road traffic safety risk estimation index system with road traffic safety entropy (RTSE) as the primary index and rapid acceleration frequency, rapid deceleration frequency, rapid turning frequency, speeding frequency, and high-speed neutral coasting frequency as secondary indices is established. A calculation method of RTSE is proposed based on an improved entropy weight method. This method involves three aspects, namely, optimization of the base of the logarithm, processing of zero-value secondary indices, and piecewise calculation of the weight of each index. Additionally, a safety risk level determination method based on two-step clustering (density and k-means clustering) is also proposed, which prevents isolated data points from affecting safety risk classification. A risk classification threshold calculation method is formulated based on k-mean clustering. The results show that high-incidence locations for aberrant driving behaviors are consistent with the locations of traffic accidents. The proposed methods are validated through a case study on four roads in Chongqing with a total length of approximately 38 km. The results show that the road traffic safety trends characterized by road safety entropy and traffic accidents are consistent.

1. Introduction

With the rapid development of urban road traffic systems, traffic accidents have become a serious social problem that poses a grave threat to the safety of human lives and property. In the period from 2011 to 2017, the number of traffic accident casualties in China decreased each year but was still very high. On average, approximately 60,000 people died from traffic accidents each year. Research has shown that more than 95% of traffic accidents are caused by driver cognitive and behavior decision errors [1]. Therefore, studying road traffic accidents and safety risks from a driving behavior perspective can effectively support prevention and early warning for traffic accidents and improve road passing efficiencies and service levels.

Currently, road traffic safety risk is extensively studied. In general, the relevant research can be divided into three categories, including research that evaluates road traffic safety based on statistical indices for traffic accidents utilizing methods such as Bayesian networks (BNs) and accident rate methods; research that establishes an evaluation index system considering the different characteristics of people, vehicles, roads, and environment and evaluates road traffic safety using methods like analytic hierarchy processes (AHPs) and fuzzy evaluation; and research that evaluates road traffic safety based on driving behavior and traffic accident data.

Regarding the evaluation of road traffic safety based on statistical indices for traffic accidents, by analyzing methods for identifying accident-prone locations in China and elsewhere, Fang et al. proposed a level-based identification algorithm applicable to road traffic in China and a new

microevaluation method (the cumulative frequency curve method) for identifying accident-prone locations [2]. Xin comprehensively evaluated the road traffic safety state using the entropy weight-technique for order preference by similarity to ideal solution (TOPSIS) method based on five evaluation indices, namely, the number of traffic accident deaths, average number of deaths per accident, fatality rate, number of deaths per 10,000 vehicles, and number of deaths per 100,000 people [3]. Mbakwe et al. evaluated national highway traffic safety using the Delphi technique in conjunction with a BN model based on highway traffic accident data [4]. Mohan et al. and Wang et al. studied urban traffic safety evaluation methods based on accident rates [5, 6]. Sandhu et al. evaluated road traffic accident black spots using the kernel density estimation method based on road traffic accident data [7]. Dang et al. established a regional road traffic safety evaluation index system by multiple correlation analysis of traffic accident data [8]; similarly, these researchers evaluated urban road traffic safety based on accident rates. Wang et al. and Elvik et al. evaluated urban road traffic safety using BNs based on traffic accident data [9, 10]. From a traffic management perspective, Eusofe et al. and Gomes et al. evaluated road traffic safety based on traffic accident data [11, 12]. Zhang et al. established an equivalent accident frequency model based on the absolute accident frequency, accident consequences, and impact on traffic [13]; additionally, these researchers used this model in a combined location safety evaluation method for urban expressways.

Regarding establishment of road traffic safety evaluation indices based on the different characteristics of people, vehicles, road, and environment, Wang et al. used an eight-degree-of-freedom driving simulator to replicate the full range of combined alignments used on a mountainous freeway in China [14]; additionally, multiple linear regression models were developed to estimate the effects of the combined alignments on the lateral acceleration. Li et al. examined the effects of subjective and objective safety indices on road safety and analyzed the relationships between an objective safety index, which comprises road linearity, pavement, traffic facilities, and natural environment and road safety [15]. Sun et al. evaluated the traffic safety state of interwoven areas using indices such as the number of traffic conflicts, traffic count, and the length of the interwoven area [16]. Niu et al. evaluated the road traffic safety state using two indices, namely, the road conditions and traffic accidents [17]. Cheng et al. evaluated road traffic safety with road conditions as evaluation index [18]. Luo et al. established an urban traffic safety state evaluation model using a fuzzy algorithm with people, vehicles, roads, and environment as evaluation indices [19]. Li created a multilevel safety evaluation index system based on expressway linearity and established a comprehensive linear safety evaluation model for expressways based on the extension theory; additionally, Li determined the weights of indices using the entropy weight method and classified safety levels [20].

Regarding research on road traffic safety risks based on driving behavior, traffic flow, and traffic accident data, Gao et al. studied and analyzed a road traffic accident risk prediction model for the technical environment of continuous urban traffic observation and dynamic control (continuous data environment for short) based on logistic regression and random forests [21]. Chen et al. proposed a new hotspot identification method based on quantitative risk assessment and used this method to identify potential accident-prone locations on highways [22]. Yu et al. proposed a hybrid latent class analysis modeling approach to consider the heterogeneous effects of geometric features in accident risk analysis; additionally, these researchers established traffic accident risk analysis models using a Bayesian random parameter logistic regression algorithm [23]. Sun and Sun conducted modeling analysis on the real-time traffic flow parameters of expressways in Shanghai and the accident risk based on coil detector and accident data in combination with a BN model [24]. Xu and Shao established a dynamic whole-vehicle model for a certain microvehicle and a road model using the multibody dynamic software Automated Dynamic Analysis of Mechanical Systems (ADAMS); subsequently, these researchers used the models to conduct virtual simulations to quantitatively analyze the effects of driver behaviors on brake safety [25]. Based on driving behavior data, Min determined the road traffic safety state using an AHP and a comprehensive fuzzy evaluation method [26]. Li et al. and Qu et al. formulated road traffic safety evaluation methods based on traffic accident and driving behavior data [27, 28].

As mentioned above, the relevant research on road traffic safety evaluation has accumulated rich results but is still deficient to a certain extent due to the different evaluation methods and data involved. (1) The evaluation based on statistical indices for traffic accidents is performed after the occurrence of traffic accidents. It does not consider the fundamental causes of traffic accidents, including the aberrant driving behaviors, road, weather, and traffic conditions. Therefore, it is important to estimate road traffic safety risk in advance for accidents prevention; however, existing research is insufficient for preassessment of road traffic safety risk. (2) The research that establishes an evaluation index system considering the different characteristics of people, vehicles, roads, and environment is short of intermediate feature data for describing driving behaviors; thus, it is difficult to accurately predict road traffic safety risk. (3) Driving behavior data provides support for exploring the intrinsic causes of accidents; unfortunately, up to now most researches employ a small amount of driving behavior data which covers a few behavior patterns. As a result, it is difficult to depict traffic safety risk under various road conditions when driving behavior data is inadequate.

Hence, based mainly on onboard diagnostic (OBD) driving behavior data and the information entropy theory, this study establishes an urban road traffic safety risk evaluation index system and a relevant calculation method, investigates a traffic safety risk estimation method, and classifies road traffic safety risks.

2. Data Preprocessing

2.1. Brief Introduction to Vehicle OBD Data. In the main urban area of Chongqing, there are approximately 100,000

private vehicles with OBD devices installed. An OBD device updates and records 13 types of vehicle data (including global positioning system (GPS), driving behavior, and security alarm data) every 2–10 s. Based on a preliminary analysis of original data, two types of vehicle data, namely, GPS and driving behavior data, are primarily used in this study for analysis. Vehicle GPS data consist of 27 fields, including data type, vehicle identification number (ID), time, longitude, latitude, and speed. Driving behavior data consist of four fields, namely, data type, vehicle ID, time, and driving behavior type. Because an OBD device transmits data independently based on the type of data, it is necessary to match a vehicle's GPS and driving behavior data to obtain driving behavior and relevant information.

2.2. Driving Behavior and GPS Data Matching. Based on the vehicle ID and time fields in the GPS and driving behavior data collected by the OBD device onboard a vehicle, the vehicle's GPS and driving behavior data are matched to obtain aberrant driving behavior and relevant longitude and latitude information. Table 1 summarizes the driving behavior data obtained after data matching.

2.3. Classification of Road Sections. Vehicle driving behaviors are, to a relatively large extent, affected by road conditions. To accurately evaluate road traffic safety risk under different road conditions based on aberrant driving behaviors, road sections are classified into eight categories, according to three characteristic parameters (the slope gradient, radii of turns, and presence of openings). The classification standard of road sections is shown in Table 2.

3. Establishment of a Road Traffic Safety Risk Evaluation Index System Based on Aberrant Driving Behaviors

3.1. Correlation Analysis of Aberrant Driving Behaviors and Traffic Accidents. In actual traffic, aberrant vehicle driving behaviors, such as rapid acceleration, rapid deceleration, and rapid turning, can easily occur as a result of road, climate, and traffic conditions. When a vehicle exhibits an aberrant driving behavior, this behavior alone may result in an accident or may have a relatively significant impact on the surrounding vehicles, causing a multivehicle traffic accident. Therefore, there may be a relatively high probability of traffic accidents at high-incidence locations for aberrant driving behaviors.

To verify the above inference, a case study of Xuefu Avenue in Chongqing (sections between Si Gongli and Liu Gongli) was performed. Based on the aberrant driving behavior data of 8,486 vehicles in a 6-consecutive-day period (231 rapid acceleration data items, 416 rapid deceleration data items, 99 rapid turning data items, 12 speeding data items, and 87 high-speed neutral coasting data items), an aberrant driving behavior distribution heat map was produced using ArcGIS, as shown in Figure 1(a). Additionally, 1-month traffic accident data (23 accidents) for Xuefu Avenue were obtained from Chongqing municipal traffic management authorities. Figure 1(b) shows the location of each accident. As demonstrated in Figure 1, spatially, the locations of traffic accidents agreed well with the sections with high incidence of aberrant driving behaviors.

Additionally, 6-day aberrant driving behavior data (10,715 aberrant driving behavior data items for 42,558 vehicles) and traffic accident data in a month (302 traffic accidents) for two other roads, including Longteng Avenue and Shi Xiaolu Avenue, were gathered and processed. A matching analysis of these data, similar to that shown in Figure 1, was performed in Figures 2 and 3.

Accidents number and aberrant driving behavior frequency of each road section on the three avenues mentioned above were calculated, and their distribution curves were as shown in Figure 4.

Results show that, in most cases for the three avenues, when aberrant driving behavior frequency rises, accidents number increases, which infers that trends of aberrant driving behavior frequency and accident frequency are basically consistent with each other. There may be a few exceptions, where drivers can perceive a potential high safety risk and take corresponding precautions to avoid accidents as field survey implies. Nevertheless, aberrant driving behavior data can represent the risk of road traffic safety in general.

3.2. Selection of Road Traffic Safety Risk Evaluation Indices. For any road sections, the lower the aberrant driving behavior frequencies are, the more orderly the traffic flow is and the lower the probability of traffic accidents is, and vice versa. This phenomenon is very similar to the disorderliness of a system described by information entropy. In 1865, German physicist Rudolf Clausius proposed the concept of entropy. In 1948, Shannon quantified entropy to reflect the orderliness of a system [29]. The more orderly a system is, the lower the information entropy of the system is, and vice versa.

Therefore, a road traffic safety risk evaluation index system is established with the road traffic safety entropy (RTSE) as the primary index and the frequencies of various aberrant driving behaviors affecting the road traffic safety as the secondary indices (Table 3).

4. RTSE Calculation Method Based on an Improved Entropy Weight Method

4.1. Calculation Process for RTSE. Overall, the RTSE calculation method involves two steps, namely, calculating the values of the secondary evaluation indices and calculating the weights of the secondary indices and the value of RTSE.

The value of a secondary index (aberrant driving behavior frequency), P_{ij}^k, is calculated as follows:

$$P_{ij}^k = \frac{A_{ij}^k}{Z_{ij}^k}, \tag{1}$$

where i is the sections number, k is the time, j is the index number, A_{ij}^k is the aberrant driving behavior frequency for the road sections i corresponding to the index j within time

Table 1: Samples of driving behavior data.

Vehicle ID	Time	Direction angle	Speed	Altitude	Latitude	Longitude	Driving behavior type
fdc14ca4. . .bb4b130	2018/5/16 07:37:20	306	28	72	29.622507	106.522736	Rapid acceleration
fdc14ca4. . .bb4b130	2018/5/16 07:37:30	313	26	70	29.622723	106.52349	Rapid deceleration
fdc14ca4. . .bb4b130	2018/5/16 07:38:00	337	0	69	29.622875	106.52382	Rapid turning

Table 2: Classifications of example road sections.

Road condition	Bend		Straight line	
	With openings	Without openings	With openings	Without openings
Flat	Type I	Type II	Type III	Type IV
Sloped	Type V	Type VI	Type VII	Type VIII

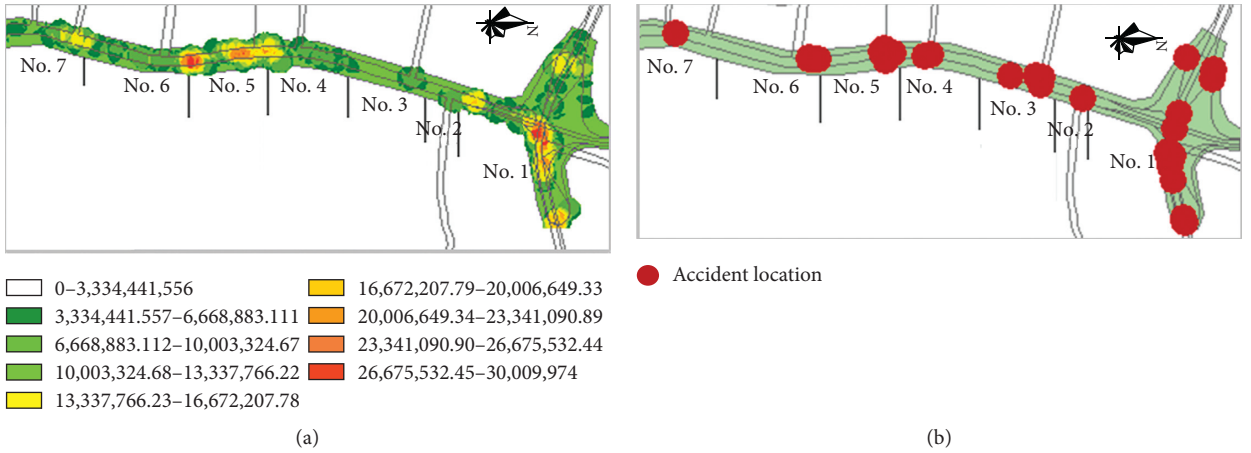
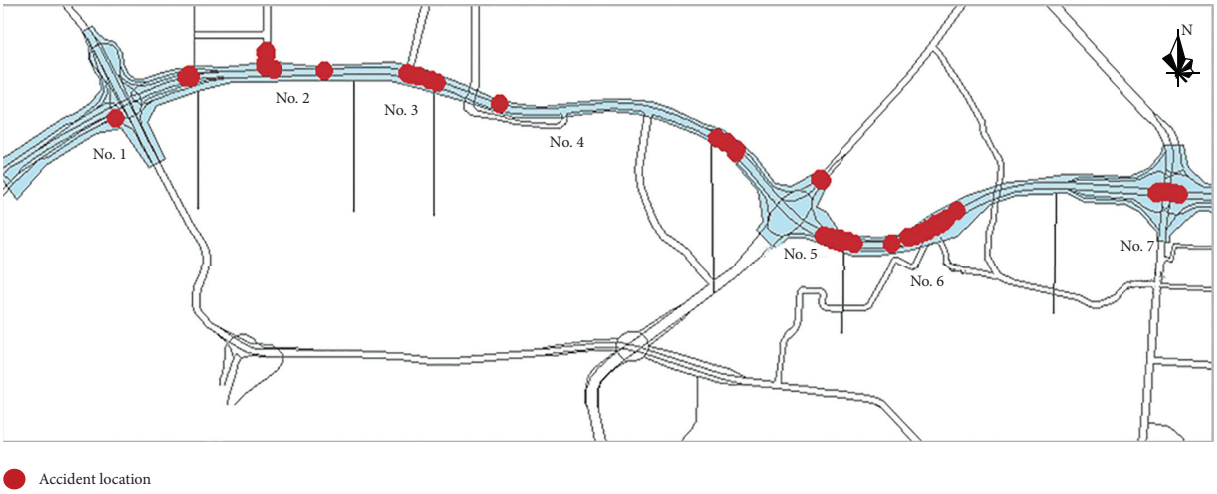


Figure 1: Aberrant driving behaviors and traffic accidents on Xuefu Avenue.

(a)

Figure 2: Continued.



(b)

Figure 2: Aberrant driving behavior and traffic accidents on Longteng Avenue.

(a) (b)

Figure 3: Aberrant driving behavior and traffic accidents on Shi Xiaolu Avenue.

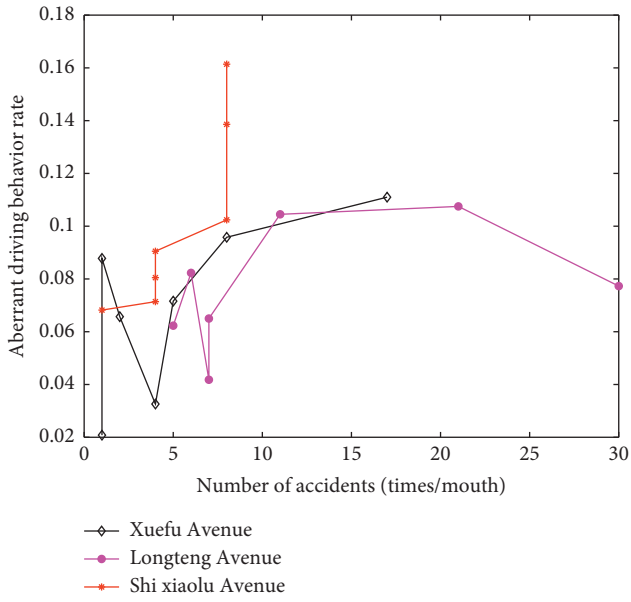


Figure 4: Distribution of frequencies of aberrant driving behaviors and traffic accidents.

Table 3: Evaluation index system.

Primary evaluation index	Secondary evaluation index
Road traffic safety entropy (RTSE)	Rapid acceleration frequency, rapid deceleration frequency, rapid turning frequency, speeding frequency, and high-speed neutral coasting frequency

k, and Z_{ij}^{k} is the number of OBD-equipped vehicles that travel through the road sections i within time k.

Several methods are available for calculating the weight of an index, including the entropy weight method, AHP, and principal component analysis. The entropy weight method determines the weight of an index based on the difference of the index from the other indices. The more significantly an index differs from other indices, the greater the weight of the index is. The entropy weight method is relatively applicable to description of the effects of aberrant driving behaviors on the road traffic safety risk level. For example, for several road sections differing in traffic accident frequency, if there is a relatively significant change in the frequency of a certain aberrant driving behavior and the frequencies of other aberrant driving behaviors remain basically unchanged, then the frequency of the aberrant driving behavior in question results in a difference in the accident frequency. Therefore, the aberrant driving behavior in question can be assigned a relatively large weight. The entropy weight method calculates the weight of an index in the following process:

(1) Data standardization:

$$\lambda_{ij}^{k} = \frac{P_{ij}^{k} - P_0}{P_1 - P_0}, \tag{2}$$

where $P_0 = \mathrm{Min}(P_{ij}^{1}, P_{ij}^{2}, \ldots, P_{ij}^{q})$ and $P_1 = \mathrm{Max}(P_{ij}^{1}, P_{ij}^{2}, \ldots, P_{ij}^{q})$.

(2) Calculation of the entropy value of the index h_j:

$$h_j = \sum_{i=1}^{n}\sum_{k=1}^{q}\left(-\lambda_{ij}^{k}\right)\log_a \lambda_{ij}^{k}, \tag{3}$$

where n is the total number of road sections, q is the number of time periods, and a, the base of the logarithm, is set to 2.

(3) Calculation of the weight of the index w_j:

$$w_j = \frac{1 - h_j}{\sum_{j=1}^{m}\left(1 - h_j\right)}, \tag{4}$$

where m is the total number of indices.

The entropy weight method can objectively calculate the weight of an index. However, when using this method in practice, optimization and demonstration are required. For example, setting the base a of the logarithm to an unsuitable value may result in a negative weight. It is impossible to calculate the entropy value of a zero-value index. Additionally, when the entropy values of all the indices are close to 1, the difference between the indices may be greater.

4.2. Improvement of the Entropy Weight Method

4.2.1. Optimization of the Base of the Logarithm a. When calculating the entropy value of an index, a used in the original entropy weight method is set to 2. In certain studies, a is set to 10 or the number of objects evaluated. This assignment may lead to an unreasonable weight for the index. Thus, it is proposed that a be set to the number of secondary evaluation indices. The reason is discussed below.

The information entropy proposed by Shannon primarily solves communication problems. There are only two basic computer storage units (binary), 0 and 1. When an event may have two consequences, each of which has a probability of 50%, the system results are the most random; that is, the level of disorderliness is the highest. Under this condition, the entropy value of the system is 1 when the logarithm of a is 2. Based on (4), when calculating the entropy values of indices, a needs to ensure that the maximum entropy is 1 and the weights of the indices are reasonably allocated when the indices have the same probability of occurrence.

Here, 2,000 groups of random numbers (including data groups in which each index has a value of 0.2) are generated under the following conditions: number of indices, 5; sum of indices, 1. A plot is created with the variance of each group of numbers as the x-axis and the product of each group of numbers as the y-axis, as shown in Figure 5. Evidently, the smaller the variance is, the greater the product is. The product reaches the maximum value of 3.2×10^{-4} at a variance of 0 (i.e., all the indices have a value of 0.2).

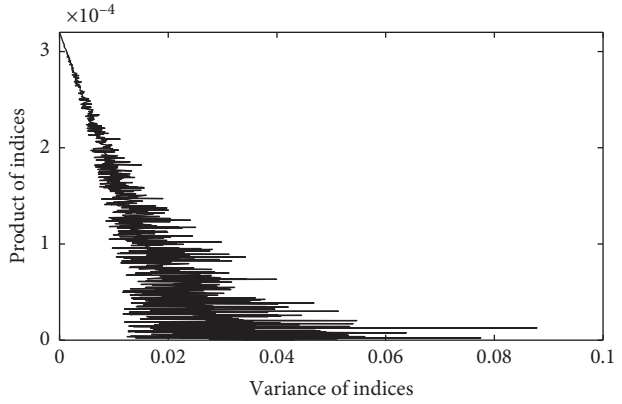


Figure 5: Relationship between the variance and product of data groups.

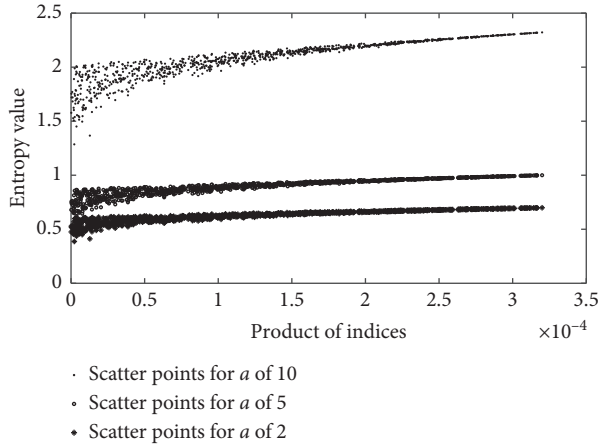


Figure 6: Relationship between a and the product of indices.

Then, the entropy value of the system is calculated with a of 2, 5, and 10 and the product of each group of data as the input. The relationship between the product of the indices and the entropy value of the system is shown in Figure 6.

The entropy value increases with the product of the indices. The entropy reaches the maximum value of 1 at a product of indices of 3.2×10^{-4} (i.e., all the indices have the same probability of occurrence) and a of 5. This result agrees with the information entropy theory. Therefore, when calculating the entropy value of an index, it is recommended that a be equal to the number of evaluation indices.

4.2.2. Zero-Value Processing of Secondary Indices. The original entropy weight method is unable to calculate the entropy value for zero-value data. The available studies mainly use two methods for processing zero-value data, namely, by directly discarding the group of zero-value data and adding an increment of 1 to the zero-value data. Here, an alternative method for processing zero-value data is proposed. When a certain evaluation index in a group has a zero value, 0.00001 is added to each index in the group. The reason is discussed below.

The aberrant driving behavior frequency varies relatively significantly between different road conditions. For upslope or long straight road sections, the probability of high-speed neutral coasting is almost 0. This result is an objectively existent phenomenon. Discarding the group of data in question leads to a deficient description of the objective phenomenon. Adding an increment of 1 to all the data accounts for this limitation. However, the slope of the logarithmic function for calculating the entropy value of an index continuously decreases as the value of the independent variable increases. If an increment of 1 is added to all the data, the difference in the entropy values between the other nonzero-value indices decreases, which relatively significantly affects the allocation of weights to the indices. Therefore, when a secondary index has a value of 0, it is recommended that a minor increment be added to all the index data and that the addition of this increment have a nonsignificant impact on the difference between indices.

Here, an example is given. There is a group of five indices with values of 0.23, 0.27, 0.21, 0.10, and 0.19. The weight of each index is calculated. Then, an increment ΔP ranging from 0.00000001 to 1 is added to each index. Subsequently, the change in the weight of each index is calculated. The relationship between the increment added to each index and the change in its weight is shown in Figure 7.

As demonstrated in Figure 4, when the increment ΔP is less than 0.00001, there is almost no change in the weight of each index. Therefore, when a secondary index has a value of 0, an increment of 0.00001 could be uniformly added to this group of data.

4.2.3. Weight Calculation Method. When the entropy values of all the indices are close to 1 and have a very slight difference, the weights calculated may differ multifold. Thus, a piecewise calculation method for the weights of the indices based on their entropy value distribution is proposed, as shown in

$$w_j'' = \begin{cases} \dfrac{\left(1 + \overline{h} - h_j\right)}{\left[\sum_{j=1, h_j \neq 1}^{m} \left(1 + \overline{h} - h_j\right)\right]}, & \text{Case1,} \\ \dfrac{\left(1 - h_j\right)}{\left[\sum_{j=1}^{m} \left(1 - h_j\right)\right]}, & \text{Case2,} \end{cases} \tag{5}$$

where case 1 describes a situation where there is a relatively small difference in the entropy values or indices and all entropy values are distributed in the range of (0.8, 0.91) or (0.95, 1); case 2 describes other situations.

When using (4) to calculate weights, if the entropy values of all the indices are close to 1 and differ nonsignificantly, then the weights calculated may differ multifold [30, 31]. Here, an example is given. Five indices are selected. Correspondingly, a group of data is selected as the entropy values for the indices. The maximum difference between the data in this group does not exceed 0.04. Additionally, the data in this group vary in the range close to [0.6, 1]. The weight of each index is calculated using (4). Moreover, the product of the weights of the indices is calculated. A plot is

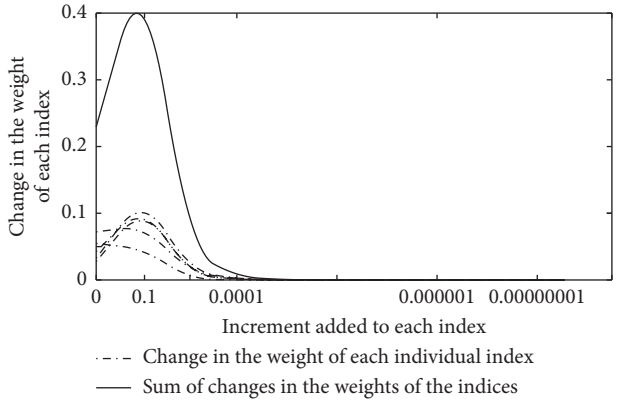


Figure 7: Relationship between the increment added to the indices and the changes in their weights.

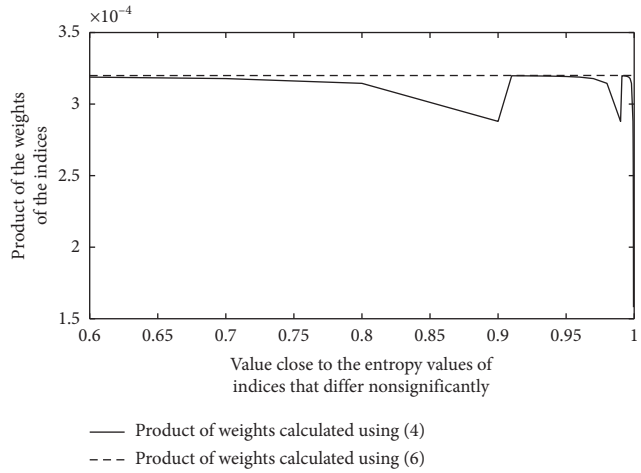


Figure 8: Product of the weights of the indices.

created with the values close to the entropy values of the indices as the x-axis and the product of the weights of the indices as the y-axis, as shown in Figure 8.

As mentioned previously, the product of a group of data increases as its variance decreases. As demonstrated in Figure 5, when the entropy values of all the indices are distributed in the range of (0.8, 0.91) or (0.95, 1), the variance of the weights of the indices is relatively large; that is, the difference between the weights of the indices is relatively large. Ouyang proposed an improved weight calculation method [32], as shown in

$$w_j' = \frac{1 + \overline{h} - h_j}{\sum_{j=1, h_j \neq 1}^{m} \left(1 + \overline{h} - h_j\right)}. \quad (6)$$

In this study, the weights of the indices are calculated using (6). The relationship between the values close to the entropy values of the indices and the product of the weights of the indices is shown in Figure 8. This method effectively addresses the problem of the entropy values of the indices differing nonsignificantly.

4.3. Calculation of RTSE. Based on the calculation of the weights of aberrant driving behavior frequencies on various types of road sections, the aberrant driving behavior frequencies p_{ij} for any road section and the RTSE value SH_i of the section i are calculated. One has

$$p_{ij} = \sum_{k=1}^{q} \frac{P_{ij}^k}{q}, \quad (7)$$

$$\mathrm{SH}_i = \sum_{j=1}^{m} w_j'' \times \left(-p_{ij}\right) \log_m\left(p_{ij}\right), \quad (8)$$

where the value of P_{ij} is set to 0.00001 when $P_{ij} = 0$.

4.4. Road Traffic Safety Risk Classification Based on Cluster Analysis

4.4.1. Road Traffic Safety Risk Level Determination Based on Two-Step Clustering. A high traffic safety risk does not necessarily translate to a large number of traffic accidents. The RTSE values of different type sections are not absolutely correlated with the number of traffic accidents.

Density-based spatial clustering of applications with noise (DBSCAN) is able to identify data points distributed in a relatively isolated manner based on the data distribution density, thereby preventing isolated data points from affecting classification. Then, k-means clustering is conducted based on various numbers of clusters to calculate the silhouette coefficients for various numbers of clusters (the higher the coefficient is, the better the cluster separation is). The optimum number of clusters is selected as the number of road safety risk classification levels.

4.4.2. Road Traffic Safety Risk Level Threshold Optimization Algorithm Based on k-Means Clustering. Level thresholds are calculated based on optimum k-means clustering results. It is assumed that there is a number of levels r. The number of classification level thresholds $(r-1)$ is calculated. The pseudocode of the algorithm (Algorithm 1) is as follows:

5. Validation Case Study

5.1. Selection of Example Roads. Chongqing, a typical mountainous city in China, has multiple centers and cluster-typed urban space. The clusters in Chongqing are connected only by expressways and arterial roads. In this study, the road traffic safety risks of Longteng Avenue (Road A, approximately 4.4 km long), Hongshi Avenue (Road B, approximately 2.5 km long), the Inner Ring Expressway and Airport Expressway (Road C, approximately 27 km long), and Xuefu Avenue (Road D, approximately 4.5 km long) in Chongqing are evaluated, as shown in Figure 9.

5.2. Data Preprocessing. By the OBD data processing method, the GPS and driving behavior data for 13,004

(1) int $r \longleftarrow k$-means number of clusters
(2) For int $t = 1$ to $r - 1$
(3) int $a_t \longleftarrow$ Safety entropy value of the center of the t^{th} cluster
(4) int $b_t \longleftarrow$ Safety entropy value of the center of the $(t+1)$th cluster
(5) int $s_t \longleftarrow$ Sum of the data in the tth and $(t+1)^{th}$ clusters
(6) int $f = 1$
(7) For float $e_{tf} = a_t$ to b_t
(8) int $c_{tf} \longleftarrow$ Volume of data in the tth cluster that is misclassified
(9) int $d_{tf} \longleftarrow$ Volume of data in the $(t+1)$th cluster that is misclassified
(10) $g_{tf} = 1 - ((c_{tf} + d_{tf})/s_t)$//Calculation of accuracy
(11) $e_{tf} = \text{etf} + 0.01$
(12) $B_t\ (f, 1{:}\ 2) = [e_{tf}, g_{tf}]$//The threshold and accuracy are stored in the matrix B_t
(13) $f = f + 1$
(14) End for
(15) $C_t \longleftarrow$ Generation of the threshold corresponding to the highest accuracy
(16) End for
(17) $C = [C_1, C_2, \ldots, C_{r-1}]$//A number of thresholds $(r-1)$ is successively stored in the vector C

Algorithm 1: Level threshold optimization algorithm ("⟵" represents value assignment).

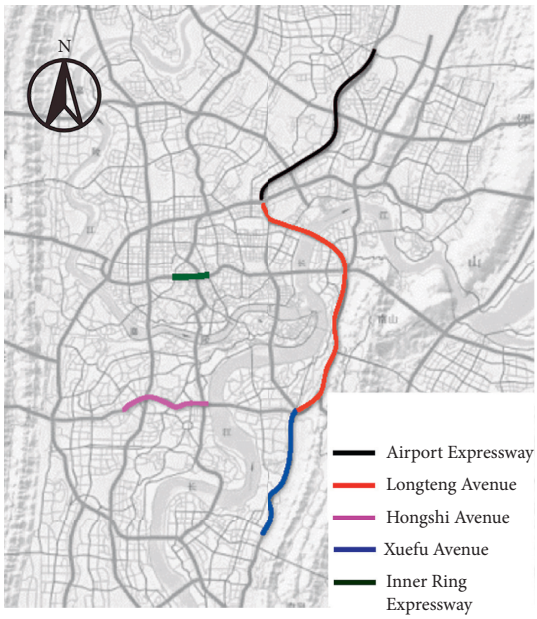


Figure 9: Example roads.

vehicles on Road A, 8,474 vehicles on Section B, 21,080 vehicles on Road C, and 8,486 vehicles on Road D, extracted from the OBD data, were matched each other.

By the classification standard of road section, the four road sections were divided into a total of 46 sections, each of which was 0.2–0.8 km long, as shown in Figure 10.

5.3. Calculation of RTSE Value

5.3.1. Calculation of the Weights of Secondary Indices for the Road Sections. By the improved entropy weight method, the weights of aberrant driving behavior frequencies for various types of road sections were calculated. Table 4 summarizes the results.

5.3.2. Calculation of the RTSE Values of the Road Sections. Based on the calculated weights for various types of road sections, the safety entropy values of the 46 road sections were calculated using (7) and (8). For example, the eight sections of road A have safety entropy values of 0.0436, 0.0278, 0.0318, 0.0385, 0.0439, 0.0358, 0.0277, and 0.0204.

5.3.3. Comparative Analysis of RTSE and Number of Traffic Accidents. There are obvious differences in traffic safety between signal-controlled urban arterial roads and expressways with and without openings. Twelve road sections of three types were selected from the example roads. The RTSE value of each road section was calculated and then compared with the number of traffic accidents during one month. Figure 11 shows the results.

As demonstrated in Figure 11, road safety entropy values are consistent with the change trend of traffic accidents, indicating that road safety entropy values can effectively represent road traffic safety risks.

5.4. Classification of Road Traffic Safety Risk

5.4.1. Determination of the Number of Risk Levels. The RTSE values and traffic accident data for 12 sections of signal-controlled arterial roads, 12 sections of expressways with openings, and 12 sections of expressways without openings (a total of 36 road sections) were selected. These data were then subjected to a DBSCAN analysis to remove the data points distributed in a relatively isolated manner. The remaining data points were subsequently subjected to k-means clustering analysis.

In MATLAB, the numbers of clusters obtained from 2-(Figure 12), 3-, and 4-means clustering were analyzed. Additionally, the silhouette coefficients for various numbers of clusters were calculated. The silhouette coefficients for k of 2, 3, and 4 were found to be 0.44, 0.37,

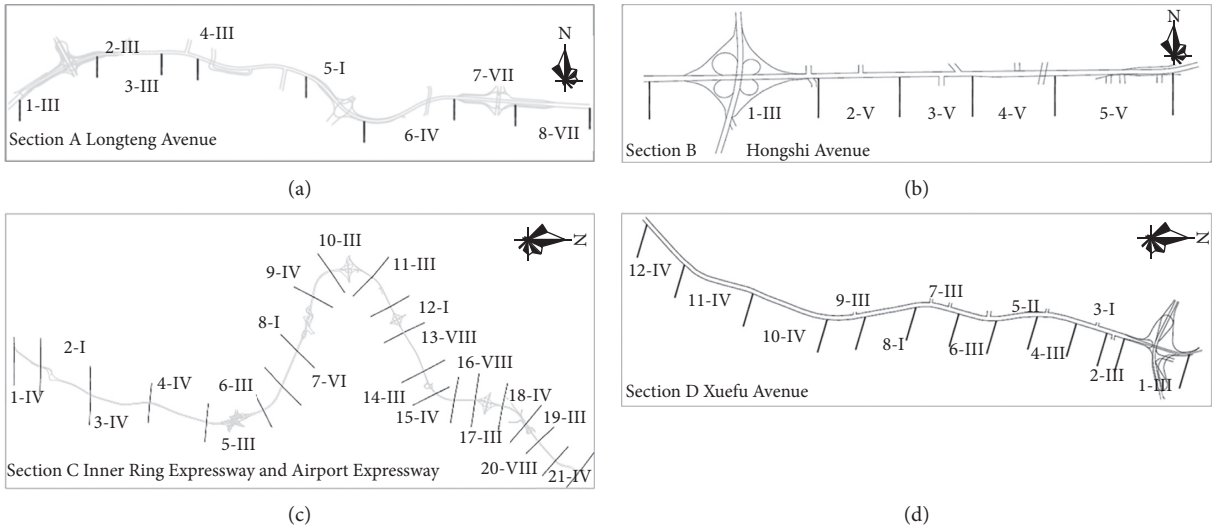


Figure 10: Classification of road section.

Table 4: Weights of aberrant driving behavior frequencies.

Road section types	Rapid acceleration frequency	Rapid deceleration frequency	Rapid turning frequency	Speeding frequency	High-speed neutral coasting frequency
I	0.19	0.19	0.16	0.30	0.18
II	0.18	0.23	0.18	0.26	0.21
III	0.16	0.21	0.18	0.33	0.16
IV	0.21	0.11	0.28	0.25	0.18
V	0.20	0.16	0.32	0.12	0.17
VI	0.25	0.14	0.27	0.24	0.14
VII	0.15	0.09	0.20	0.45	0.15
VIII	0.24	0.16	0.34	0.12	0.07

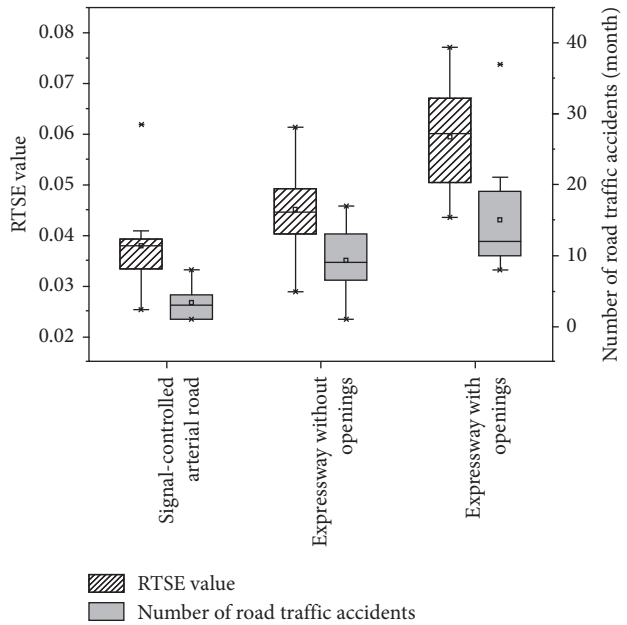


Figure 11: Comparative analysis of RTSE value and number of road traffic accidents.

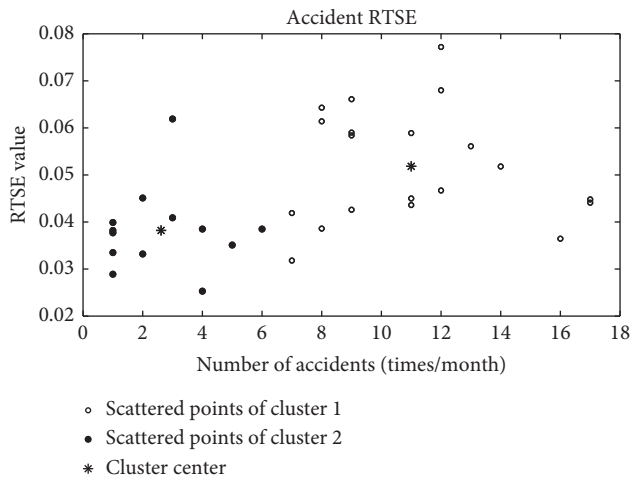


Figure 12: 2-means clustering results.

and 0.39, respectively. Evidently, 2-means clustering produced the best results. Thus, in this study, the road traffic safety risks are classified into two levels, namely, high and low risk.

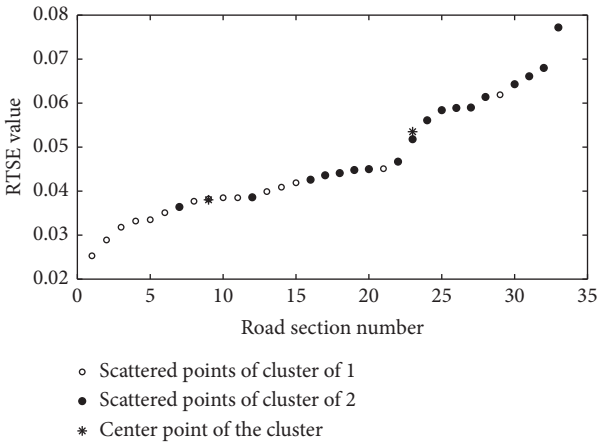


Figure 13: Sorted RTSE scatter points.

5.4.2. Calculation of Road Traffic Safety Risk Level Classification Threshold. In this part, *k*-means clustering, fuzzy clustering, and support vector machine were used to calculate risk classification thresholds and corresponding accuracies, on the basis of the 36 road sections' data as described in 4.4.1.

(1) k-Means Clustering. The RTSE values of all the road sections in class 1 and class 2 obtained from *k*-means clustering were sorted in an ascending order. Figure 13 shows the sorted data.

The classification accuracies of different RTSE threshold values for road traffic safety grading were calculated. The potential thresholds range from RTSE of clustering center of class 1 to that of clustering center in class 2. Figure 14 shows the threshold calculation result.

Evidently, the accuracy is the highest (87.88%) when the traffic safety risk level classification threshold for the road sections is 0.042.

(2) Fuzzy Clustering. Fuzzy clustering was conducted to separate data points of RTSE values and accident numbers into 2 classes, and the result was presented in Figure 15 and Table 5.

The RTSE values of all the road sections in class 1 and 2 obtained from Fuzzy clustering were sorted in an ascending order, and classification accuracies of different RTSE threshold values for road traffic safety grading were calculated. The potential thresholds range from RTSE of clustering center of class 1 to that of clustering center in class 2. The threshold calculation result was shown in Figure 16.

As the result of fuzzy clustering shows, traffic safety risk classification accuracy achieves the best (87.88%) when RTSE threshold is 0.041 or 0.042.

(3) Support Vector Machine. 15 road sections' RTSE values and accident numbers were selected for training support vector machine, and then it was used to classify traffic safety risk levels of all the road sections into two classes. The result shows that when the accuracy reaches the best (87.88%), the RTSE threshold is 0.041 or 0.042.

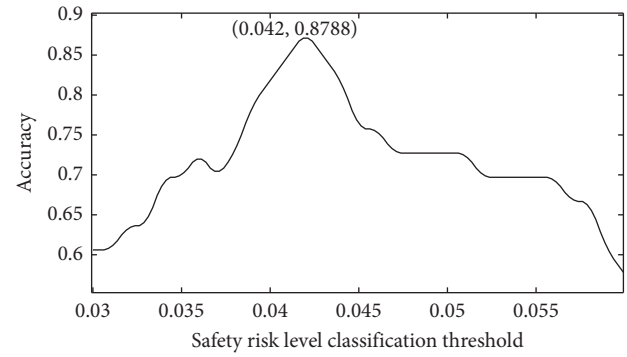


Figure 14: Road traffic safety risk level classification threshold and accuracy.

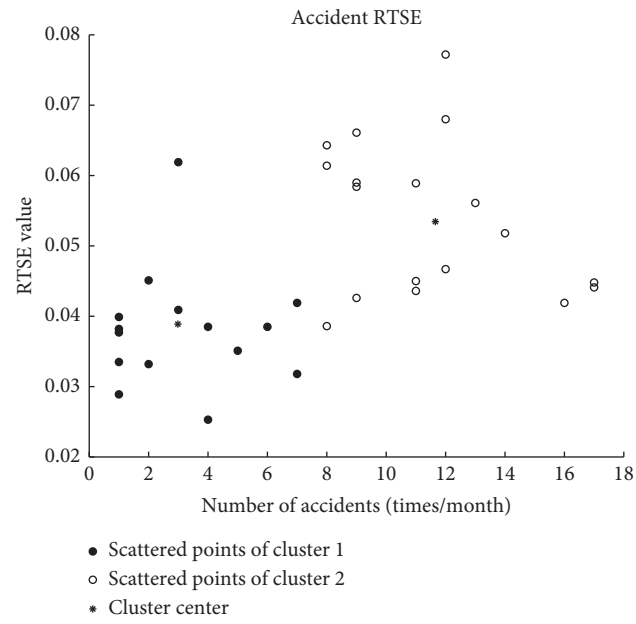


Figure 15: Fuzzy clustering result.

Table 5: Clustering centers of fuzzy clustering.

Class	Number of accidents (month)	RTSE values
Class 1	3.0	0.039
Class 2	11.6	0.053

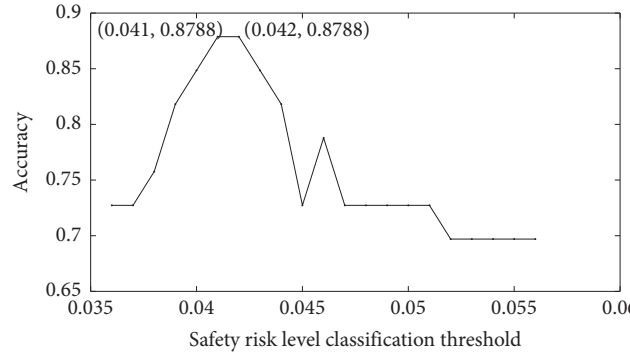


Figure 16: Fuzzy clustering calculation of road traffic safety entropy threshold and accuracy.

As we can see, classification accuracy achieves the highest as 87.88% when RTSE threshold is 0.042 for each of the three methods. Therefore, RTSE threshold is recommended to be 0.042 to identify road traffic safety risk level; that is to say, the road traffic safety is at a low level if RTSE is less than 0.042; otherwise, it is of high safety risk if RTSE is greater than 0.042.

6. Concluding Remarks

In this study, based on OBD vehicle driving behavior data, the correlation between aberrant driving behaviors and traffic accidents is analyzed. On this basis, a road traffic safety risk evaluation index system and an index calculation method are established based on information entropy theory. Additionally, based on traffic accident data, a road traffic safety risk estimation method is established through cluster analysis.

The validation case study demonstrates that the road traffic safety condition depicted by the RTSE value exhibits the same trend as that depicted by the number of traffic accidents. The road traffic safety risk prediction method established based on driving behavior data is able to effectively and objectively evaluate road traffic safety risk. The results derived from this study can effectively support identification of high road traffic safety risk locations, prevention, and early warning of traffic accidents. Additionally, these results can provide decision-making reference for traffic operation control in the collaborative vehicle-road environment.

Road traffic safety is affected by a multitude of factors, including the characteristics of road, driver, weather, and traffic conditions. This study is conducted primarily from the perspectives of driving behaviors and road conditions. As data continue to accumulate, it is necessary to conduct a classification study on the road traffic safety risk while considering more influencing factors.

Data Availability

The vehicle OBD data used to support the findings of this study were supplied by Chongqing Urban Transportation Big Data Engineering Technology Research Center under license and so cannot be made freely available. Requests for access to these data should be made to Zhigang Gao, 3585680376@qq.com.

Conflicts of Interest

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

Acknowledgments

This research is supported in part by Chongqing University Outstanding Talents Support Program, Chongqing Municipal Key Research and Development Project of Technology Innovation and Application Demonstration, Research Project of Chongqing Urban Traffic Big Data Engineering Technology Research Center, National Natural Science Foundation of China, Chongqing Research Program of Basic Research and Frontier Technology Innovation, and Scientific Research Project of Key Laboratory of Traffic System & Safety in Mountain Cities.

References

[1] L. Duan, *Study on the Model of Drivers' Cognitive Behavior Errors and Reliability Evaluation*, Central South University, Changsha, China, 2012.

[2] S.-E. Fang, Z.-Y. Guo, and W. Yang, "A new method for multiple location identification of highway traffic accidents," *Journal of Traffic and Transportation Engineering*, vol. 1, pp. 90–94, 2001.

[3] D.-Q. Xin, "Evaluation of road traffic safety in various urban areas of Shaanxi Province based on entropy weight-TOPSIS," *China Safety Science and Technology*, vol. 11, no. 10, pp. 118–122, 2015.

[4] A. C. Mbakwe, A. A. Saka, K. Choi, and Y.-J. Lee, "Alternative method of highway traffic safety analysis for developing countries using delphi technique and bayesian network," *Accident Analysis & Prevention*, vol. 93, pp. 135–146, 2016.

[5] D. Mohan, G. Tiwari, and S. Mukherjee, "Urban traffic safety assessment: a case study of six Indian cities," *IATSS Research*, vol. 39, no. 2, pp. 95–101, 2016.

[6] C. Wang, J.-X. Xia, Z.-B. Lu et al., "Safety evaluation method for urban intersection based on microscopic simulation and extreme value theory," *China Journal of Highway and Transport*, vol. 31, no. 176, pp. 292–299, 2018.

[7] H. A. S. Sandhu, G. Singh, M. S. Sisodia, and R. Chauhan, "Identification of black spots on highway with kernel density estimation method," *Journal of the Indian Society of Remote Sensing*, vol. 44, no. 3, pp. 457–464, 2016.

[8] X.-X. Dang, Y.-Q. Wang, Z.-H. Wu et al., "Improved generalized DEA evaluation model for regional road traffic safety," *Journal of Transportation Systems Engineering and Information Technology*, vol. 16, no. 4, pp. 11–16, 2016.

[9] J. Wang and H. Huang, "Road network safety evaluation using bayesian hierarchical joint model," *Accident Analysis & Prevention*, vol. 90, pp. 152–158, 2016.

[10] R. Elvik, H. Ulstein, K. Wifstad et al., "An empirical bayes before-after evaluation of road safety effects of a new motorway in Norway," *Accident Analysis & Prevention*, vol. 108, pp. 285–296, 2017.

[11] Z. Eusofe and H. Evdorides, "Assessment of road safety management at institutional level in Malaysia: a case study," *IATSS Research*, vol. 41, no. 4, pp. 172–181, 2017.

[12] S. V. Gomes, J. L. Cardoso, and C. L. Azevedo, "Portuguese mainland road network safety performance indicator," *Case Studies on Transport Policy*, vol. 6, no. 3, pp. 416–422, 2018.

[13] X.-Q. Zhang, X. Liu, Y. Zhu et al., "Location security analysis method for urban expressway based on internet data," *Journal of Transportation Systems Engineering and Information*, vol. 18, no. 5, pp. 57–63, 2018.

[14] X. Wang, T. Wang, A. Tarko, and P. J. Tremont, "The influence of combined alignments on lateral acceleration on mountainous freeways: a driving simulator study," *Accident Analysis & Prevention*, vol. 76, pp. 110–117, 2015.

[15] Z. Li, X. Zhou, X. Wang, and Z. Guo, "Study on subjective and objective safety and application of expressway," *Procedia—Social and Behavioral Sciences*, vol. 96, pp. 1622–1630, 2013.

[16] L. Sun, Y.-P. Li, J. Qian et al., "Traffic safety evaluation of interwoven area based on traffic conflict technology," *China Safety Science Journal*, vol. 23, no. 1, pp. 55–60, 2013.
[17] S.-F. Niu, Y.-X. Zheng, S.-D. Feng et al., "Traffic safety evaluation method for highway section based on accident tree," *Journal of Chongqing Jiaotong University*, vol. 32, no. 1, pp. 87–90, 2013.
[18] J.-Z. Cheng, Z.-F. Li, L.-C. Ren et al., "Research on fuzzy comprehensive evaluation of traffic safety based on road factors," *Journal of Taiyuan University of Science and Technology*, vol. 37, no. 4, pp. 296–301, 2016.
[19] Q. Luo, S.-G. Hu, H.-W. Gong et al., "Research and model construction of urban road traffic safety evaluation system," *Journal of Guangxi University (Natural Science)*, vol. 42, no. 2, pp. 587–592, 2017.
[20] B. Li, "Research on expressway linear safety evaluation based on extension theory," *Highway*, vol. 6, pp. 186–190, 2017.
[21] Z. Gao, W. Gao, R.-J. Yu et al., "Road traffic accident risk prediction model under continuous data environment," *China Journal of Highway and Transport*, vol. 31, no. 176, pp. 284–291, 2018.
[22] C. Chen, T.-N. Li, J. Sun et al., "Hotspot identification for Shanghai expressways using the quantitative risk assessment method," *International Journal of Environmental Research & Public Health*, vol. 14, no. 1, p. 20, 2017.
[23] R. Yu, X. Wang, and M. Abdel-Aty, "A hybrid latent class analysis modeling approach to analyze urban expressway crash risk," *Accident Analysis & Prevention*, vol. 101, pp. 37–43, 2017.
[24] J. Sun and J. Sun, "Active risk assessment of real-time traffic flow operation in urban expressway," *Journal of Tongji University (Natural Science)*, vol. 42, no. 6, pp. 0873–0879, 2013.
[25] J. Xu and Y.-M. Shao, "Quantitative analysis of the impact of driver driving behavior on brake safety," *Ergonomics*, vol. 4, pp. 29–32, 2007.
[26] Q. Min, *Road Traffic Safety Evaluation Method and Application Based on Simulated Driving*, University of Technology, Wuhan, China, 2014.
[27] X. Li, N. Zhao, and W. Zheng, "Evaluation of road traffic safety level based on cloud model," *Journal of Beijing University of Technology*, vol. 8, pp. 1219–1224, 2015.
[28] Z.-W. Qu, X. Qi, Y.-H. Chen et al., "Reverse variable lane release characteristics and its safety evaluation," *Journal of Transportation Systems Engineering and Engineering*, vol. 4, pp. 76–82, 2018.
[29] C. E. Shannon, "A mathematical theory of communication," *Bell Labs Technical Journal*, vol. 27, no. 4, pp. 379–423, 1948.
[30] H.-C. Zhou, G.-H. Zhang, and G.-L. Wang, "Multi-objective decision-making method for reservoir flood control operation based on entropy weight and its application," *Journal of Hydraulic Engineering*, vol. 38, no. 1, pp. 100–106, 2007.
[31] Y.-H. Li and J.-Z. Zhou, "Multi-objective flood control scheduling decision method based on improved entropy weight and vague set," *Hydroelectric Energy Science*, vol. 28, no. 6, pp. 38–41, 2010.
[32] S. Ouyang and Y.-L. Shi, "Improved entropy weight method and its application in power quality assessment," *Automation of Electric Power Systems*, vol. 37, no. 21, pp. 156–159, 2013.