Evaluation of Bus Driver's Driving Behavior based on Entropy Weight TOPSIS Model

Yaxuan Wang

School of Dalian Maritime University, Dalian 116000, China

w1084703036@163.com

Abstract. In recent years, due to frequent traffic accidents caused by bus drivers' irregular driving behavior, the safe operation of buses has become the focus of social attention. How to reduce traffic safety hazards and build a smart transportation system is an urgent problem to be solved. This paper uses bus GPS data, considers seven driving behavior characteristics from the perspectives of economy, safety, and drivers themselves, and uses K-means clustering to preliminarily divide driving behavior types, and divides driving behavior into radical, general, and There are three types of caution, and a comprehensive evaluation model of driver's driving behavior based on the entropy weight TOPSIS method is proposed. By scoring the driving behavior of bus drivers, the drivers with bad driving behavior can be warned and trained in time to ensure the safety of passengers. The research results of this paper can provide a certain reference value for the establishment of a comprehensive evaluation system for bus drivers' driving behavior and the development of driving behavior training.

Keywords: Bus GPS; K-means Clustering; Entropy Weight TOPSIS Method; Comprehensive Evaluation Model; Behavior Analysis.

1. Introduction

With the development of economy and industry and people's pursuit of green and environmental protection life, buses have become the main means of transportation for people to go out. In recent years, a number of traffic accidents have been directly caused by problems such as the irregular driving behavior of bus drivers. According to the 2014 Road Traffic Accident Statistical Annual Report [1], the total number of road traffic accidents involving casualties nationwide was 196,812, resulting in 58,523 deaths, 211,882 injuries, and direct property losses of 1.0754 billion yuan. A large number of research data show that in the human-vehicle-road environment, people are the main factor affecting traffic accidents, accounting for 55% to 90% of all accident factors [2]. Shen Xiaoyan [3] and others found that 67% of the accidents were caused by personnel factors through the statistical analysis of 886 road transportation accidents. Grabarek et al. [4] analyzed the accident data and concluded that exceeding the speed limit and failing to obey the traffic rules lead to the reduction of the driver's psychological efficiency as an important reason for the accident. It can be seen that the driving behavior of drivers is a key factor affecting road safety. How to regulate the operation of public transport vehicles and reduce the occurrence of accidents from the source is an urgent problem to be solved.

Existing research on the driving behavior of bus drivers generally adopts the form of questionnaire survey. Although its accuracy is high, it consumes a lot of manpower, material resources and time, and the efficiency is low. To solve this problem, this paper uses vehicle GPS data to comprehensively evaluate the driving behavior of bus drivers through data processing and modeling analysis, which saves manpower and material resources and improves the efficiency of evaluation, which is conducive to the construction of intelligent transportation systems.

This paper collects the driving data of bus No. 1 in Rizhao City with the help of GPS and denoises the data, then uses K-means clustering to classify the types of driving behavior, and then uses the entropy weight TOPSIS method to build a comprehensive evaluation model. The main contributions are as follows:

- Considering the comprehensive impact of various driving behaviors and driver's driving age

on driving safety, the data is analyzed through the driving behavior recognition algorithm, and then the entropy weight TOPSIS method is used to evaluate the driver's driving behavior. Finally, a comprehensive score of the driving behavior of each driver is obtained, in order to help the bus company to supervise and improve in a targeted manner.

- Determine the characteristic indicators that have a large proportion of driving safety through research, and focus on monitoring this indicator in the future. By comparing the original data of drivers with high scores and low scores, the general law between driving behavior, driving age and comprehensive score is obtained, which provides theoretical support for regulating the driving behavior of bus drivers.

The structure of the rest of this paper is as follows: In Section 2, the selection of feature indicators, K-means clustering model and entropy weight TOPSIS model are introduced in detail, and an example analysis is carried out by taking the GPS data of bus No. 1 in Rizhao City, Shandong Province as an example , to obtain the comprehensive score of each driver's driving behavior, and explain the case results; Section 3 introduces the literature review related to the evaluation of driving behavior; Section 4 summarizes the entire study.

2. Organization of the Text

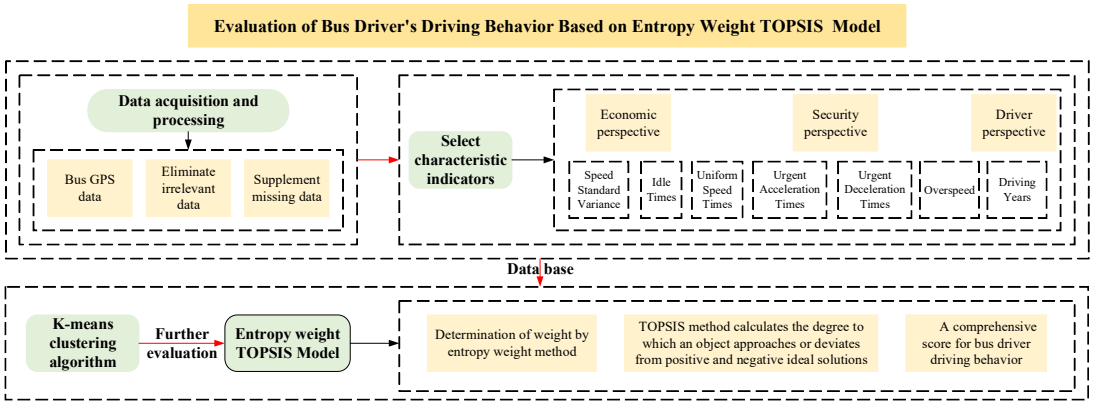


Fig 1. Framework of this research

2.1 Dataset

The data in this paper comes from the Shandong Data Open Network. The GPS positioning system is used to read the real-time location, time, vehicle NO, and the serial number of the last passing station of bus No. 1. The sampling frequency is 10 seconds. Part of the original data is shown in Figure 2.

GPS time stamp	Bus NO	Circuit NO	Driver NO	Serial Number Of The Last Station	Longitude	Latitude
20220427 06:30:00	76365	1	174	2	119.449618	35.42742
20220427 06:30:09	76365	1	174	2	119.536235	35.396825
20220427 06:30:19	76365	1	174	2	119.517342	35.408547
20220427 06:30:29	76365	1	174	2	119.53632	35.395232
20220427 06:30:39	76365	1	174	3	119.515843	35.408788
20220427 06:30:49	76365	1	174	3	119.536278	35.394673
20220427 06:30:59	76365	1	174	3	119.452675	35.427412

Fig 2. Part of the original data

2.2 Data Cleaning

2.2.1 Eliminate Irrelevant Data

Query the operating hours of the downlink section of No. 1 bus in Rizhao City, and exclude GPS data that is not between 06:00-21:00.

2.2.2 Supplement Missing Data

In this paper, equidistant points before and after the interval point are used as filling. Assume that within the time interval of missing data, the trajectory of the bus is approximately a straight line, assuming that the latitude and longitude of the previous data is (x1, y1), the latitude and longitude of the next data is (x2, y2), and the missing interval is n times 10 seconds, then It is necessary to supplement n−1 data, and the longitude and latitude of the new data inserted between the data with sampling loss are:

$$x_m = x_1 + \frac{x_2 - x_1}{n-1} * m \tag{1}$$

$$y_m = y_1 + \frac{y_2 - y_1}{n-1} * m \tag{2}$$

In the formula, m is the mth data to be filled. The interpolation filling method can effectively supplement the missing data points in the data.

2.3 Select Characteristic Indicators

From the perspective of economy, this paper selects two driving behavior characteristic indicators, speed standard variance and idle times; from the perspective of safety, selects four driving behavior characteristic indicators: uniform speed times, urgent acceleration times,urgent deceleration times and overspeed; Starting from oneself, and selecting driving years as the characteristic index at the same time, to ensure the scientificity and practicability of the driving behavior evaluation system.

2.3.1 Speed Standard Variance

Speed stability refers to the degree of stability of the speed of the vehicle during driving. In accident statistics, it is found that the higher the average speed, the greater the dispersion of speed, and the greater the probability of accidents [5].In this paper, we use the standard variance of the vehicle speed to represent the degree of dispersion of the vehicle speed, that is, the larger the standard deviation, the greater the dispersion, the more frequent the change of the vehicle speed, and the worse the stability. The formula for calculating the standard deviation of vehicle speed is as follows:

$$\bar{v} = \frac{1}{N}\sum_{i=1}^{N} v_i \tag{3}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(v_i - \bar{v})^2} \tag{4}$$

Among them, $\bar{v}$ represents the average speed of the car during this period of driving, N represents the number of speed data recorded in total, and σ represents the standard deviation of the vehicle speed.

2.3.2 Idle Times

If the vehicle speed v=0 and the maintenance time is greater than 10 seconds, it is determined that the driving behavior belongs to idling driving, and the number n of idling driving and the total duration I_t are recorded.

2.3.3 Uniform Speed Times

If the acceleration a=0 and the speed v> 0, and the maintenance time is more than 10 seconds, the driving behavior is considered to be uniform driving, and the uniform driving times n and the total duration I_t are recorded.

2.3.4 Number of Rapid Shifts

The calculation expression of the acceleration (deceleration) speed of the bus from kI to (k+1)I can be obtained as

$$a = \frac{v_{(k+1)I} - v_{kI}}{I} \tag{5}$$

Select the data of No. 1 bus whose vehicle NO is 76365 on April 27, 2022, and draw the speed-acceleration (decrease) speed change diagram, as shown in Figure 3.

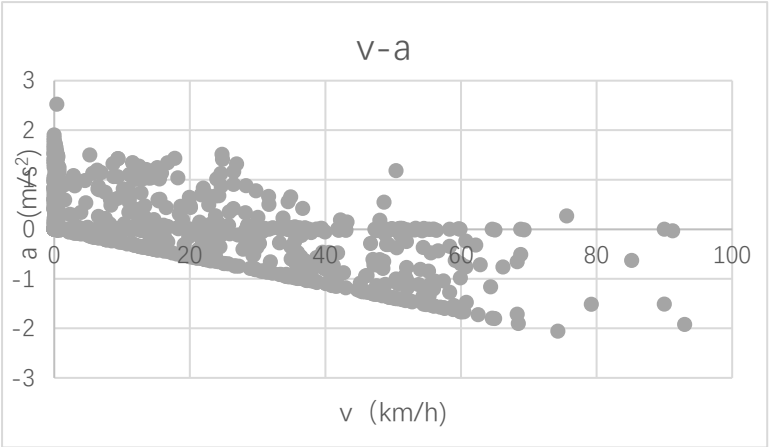


Fig 3. v-a change graph

It can be seen from the figure that the acceleration (deceleration) speed is inversely proportional to the change of speed. Relevant studies have pointed out that there are certain changes in the driving behavior of drivers at different speeds. Generally speaking, when the bus is at a low speed, the driver believes that the environment is safe, and the speed change will not have a great impact on traffic safety. On the contrary, when the bus is at a high speed, the driver is highly concentrated and more cautious in speed control.

According to statistics, the probability that the driver's acceleration falls within the interval (-1.5, 1.5) during driving is 95.4%, that is, the probability that the absolute value of acceleration is greater than 1.5 during normal driving is less than 5%, which is a small probability event. Therefore, the variable speed threshold is set at 1.5m/s^2 in this paper, that is, a_1=1.5m/s^2.

If $|a| \geq a_1$, and the speed change occurs within 10 seconds, the speed increase is a rapid acceleration behavior, and the speed decrease is a rapid deceleration behavior. Record the number n of rapid acceleration/deceleration driving and the total duration I_t.

2.3.5 Overspeed

According to relevant laws and regulations, the four departments of Rizhao City Public Security Bureau, Transportation Bureau, Highway Administration Bureau, and Environmental Protection Bureau have a speed limit of 60 kilometers per hour for buses. Therefore, the maximum speed threshold in this paper is set to 60km/h.

If the vehicle speed exceeds the maximum vehicle speed threshold and the time remains greater than 10 seconds, then the driving behavior is determined to be speeding, and the number n of speeding and the total duration I_t are recorded.

2.3.6 Driving Years

By inquiring the driving age of bus drivers of No. 1 bus in Rizhao City, it was found that the occurrence of abnormal driving behaviors of bus drivers is closely related to their own driving years.

2.3.7 Characteristic Indicators' Value

Finally, the values of each driver's characteristic indicators are obtained, as shown in Table 1.

Table 1. The values of each driver's characteristic indicators

NO	Uniform Speed Times	Urgent Acceleration Times	Urgent Deceleration Times	Idle Times	Overspeed	Speed Standard Variance	Driving Years
174	28	21	24	68	29	5.719175823	5
467	44	13	19	90	43	5.622839272	25
566	39	21	21	61	37	5.558893399	18

569	42	25	19	68	38	5.82827897	8
707	37	21	18	46	43	5.713887329	12
775	36	21	22	76	46	5.637927103	15
6722	40	12	18	69	35	5.613380528	20
6285	28	23	26	56	38	5.306697478	10
208	41	27	22	79	59	5.867439443	8

2.4 Model

2.4.1 K-means Clustering Algorithm

In this paper, K-means clustering is used to classify the types of driving behavior, to explore the distribution of samples, and to refine the characteristics of driving behavior.

The K-means algorithm is a distance-based clustering algorithm, which uses distance as the similarity evaluation index[6]. Its basic principle is to judge the similarity of the samples in the sample set according to the Euclidean distance and divide them into different clusters. The phase difference between clusters is the largest, and the basic steps of the algorithm are as follows.

(1) Randomly select k samples from the input data point set as the initial cluster center C = $\{c_1, c_2, \ldots, c_k\}$.

(2) For each point xi in the data set, calculate its distance to k cluster centers and classify it into the class corresponding to the cluster center with the smallest distance.

(3) For each category c_i, recalculate the cluster center:

$$C_i = \frac{1}{|C_i|}\sum_{x \in C_i} x \tag{6}$$

(4) Perform step (2) ~ step (3) in a loop, if the new cluster center does not move anymore, then terminate the calculation; otherwise, go to step (2).

2.4.2 Entropy Weight TOPSIS Model

The entropy weight method is an objective weighting method, which uses information entropy to determine the index weight according to the degree of dispersion of each index value. The index weight determined by the entropy weight method can make full use of the original data, avoid errors caused by human factors, and has higher credibility and accuracy than other subjective weighting methods [7-8]. TOPSIS method, which can perform final comprehensive ranking according to the closeness between a certain number of evaluation objects and corresponding targets.

In this paper, the entropy weight method is combined with TOPSIS. First, the entropy weight method is used to determine the weight of each driver's characteristic index, and then the degree to which the evaluation object is close to or deviates from the positive and negative ideal solutions is calculated to realize the objective quantification of the driving behavior scores of 9 bus drivers. The process is shown in Table 2.

Table 2. Entropy weight TOPSIS method solution process

Entropy weight TOPSIS method solution process
Step 1 Index positive and standardized processing
Step 2 Calculate the weight of each indicator
Step 3 Get the weighted normalized judgment matrix
Step 4 Calculate the proximity of each evaluation object to the positive and negative ideal solution
Step 5 Calculate the importance of each evaluation object and convert it into a percentile score, and sort according to the score
Output:TOPSIS evaluation results for each driver

2.5 Result Analysis

2.5.1 Analysis of Clustering Results

From the driver's driving behavior and driving age factors, use K-means clustering to classify driving behavior types, set the number of clusters to 3, and obtain the approximate distribution of each feature index as shown in Table 3.

Table 3. The approximate distribution of each feature index

	Clustering		
	I	II	III
Number Of Uniform Speed	35.8	42.0	34.7
Urgent Acceleration Times	25.0	12.5	21.7
Urgent Deceleration Times	23.8	18.5	21.7
Idle Times	73.6	79.5	54.3
Overspeed	43	39	39
Driving Years	8	23	13
Speed Standard Variance	5.765908375	5.618109900	5.526492735

According to the clustering results, the characteristics of driving behavior are obtained initially, and the driving behavior is divided into three categories. The four drivers whose driver NOs are 174, 569, 775, and 208 belong to the first category, the two drivers whose driver NOs are 467 and 6722 are grouped into the second category, and the three drivers whose driver NOs are 566, 707, and 6285 are grouped into third category. The first type of drivers had more times of rapid acceleration, rapid deceleration, and overspeeding, but less uniform speed and driving experience, showing aggressive driving behavior; the second type of drivers had less rapid acceleration, rapid deceleration, and overspeeding, and less constant speed 1. Older drivers, more cautious driving behavior; the third category of drivers whose index values are between the first two categories, and belong to general driving behavior. Through K-means clustering, the driver's driving behavior is divided into three types: aggressive type, general type and cautious type.

According to the classification results, it can be seen that K-means clustering can only preliminarily divide driving behavior, it is difficult to accurately classify according to different characteristics of multiple indicators, and it cannot accurately describe the specific characteristics of each type of behavior.

2.5.2 Analysis of the Comprehensive Evaluation Model

Apply the above entropy weight TOPSIS model to the trip data of 9 No. 1 buses in Rizhao City, and the weights of the driving behavior characteristic indicators of each driver are shown in Table 4.

Table 4. The weight of each driving behavior characteristic index

Evaluating Indicator	Weight
UNIFORM SPEED TIMES	0.1698
URGENT ACCELERATION TIMES	0.1320
URGENT DECELERATION TIMES	0.0811
IDLE TIMES	0.1288
OVERSPEED	0.0741
DRIVING YEARS	0.1748
SPEED STANDARD VARIANCE	0.2394

As far as each characteristic index is concerned, driving years has the largest weight, followed by speed standard variance and uniform speed times, indicating that the driving years of drivers has a

greater impact on their driving behavior, and there are large differences in the driving stability of different drivers. There are large fluctuations. The weight of overspeed is the smallest, indicating that the proportion of the number of speeding times of drivers in the driving process is relatively small, and the overall speeding behavior of the fleet is relatively small.

Through the calculation of the above model, the importance of the driving behavior of the drivers of the 9 No. 1 buses in Rizhao City is obtained, and the scores and rankings are shown in Table 5.

Table 5. Score and ranking of each driver's driving behavior

NO	SCORE	RANKING
174	24.4286	8
467	100.0000	1
566	86.0466	3
569	41.7843	7
707	66.7439	4
775	57.7241	6
6722	98.0775	2
6285	63.4997	5
208	18.0087	9

The average score of driving behavior of No. 1 bus drivers is 55.63134 points, and the drivers with a score greater than 60 only account for 55.56%, and some drivers have low scores. Through the analysis of the original data, it is found that the top-ranking drivers have higher driving experience, smaller speed fluctuations, more constant speed times, less rapid speed changes and idling times, good driving stability, and better driving behavior; the lower-ranking drivers The driving experience is relatively young and the speed fluctuates greatly, and there are obvious sudden shifts and idling behaviors, which usually manifest as aggressive driving behaviors.

The above results show that using the entropy weight TOPSIS method to analyze the driving behavior characteristic indicators can accurately identify drivers with poor driving behavior. In order to ensure the safety of passengers, drivers should reduce the number of sudden shifts and other bad driving behaviors. At the same time, for aggressive drivers at the bottom of the ranking, the bus company can take measures such as education and training, key monitoring, learning and rectification, so as to improve the safety of driving behavior in an all-round way and reduce the probability of traffic accidents caused by human factors.

3. Literature References

Yang Man et al. [9] found that the average deceleration, the type and cause of driving safety incidents, and weather have a significant impact on driving risk. Lei Cailin et al. [10] used the rough set theory in the field of artificial intelligence to objectively evaluate the state of road safety. Shi Kun et al. [11] used AHP to study the reliability of driving behavior from six aspects such as vehicle speed and temperature; Zhang Xiaobo et al. [12] used entropy weight method to give weights from three aspects of driving intensity, driving efficiency and driving speed to establish Evaluation system; Xu Zhiqi et al. [13] constructed an evaluation index system from the aspects of driving mileage, overspeed time, dangerous operation, etc., and used the analytic hierarchy process to study the safety of driving behavior. Wang Keyin and other [14] extracted vehicle driving trajectories, selected representative evaluation indicators, and used the K-means clustering method to classify drivers' driving styles into cautious, normal, and aggressive. Zhang Zhihong [15] used the K-means clustering algorithm to select indicators such as average speed, maximum speed, and overspeed to classify driving tendencies.

However, the above studies did not consider the influence of drivers' driving age on their driving behaviors when selecting characteristic indicators to evaluate drivers' driving behaviors. Usually, driving age is closely related to driving behavior, so when evaluating a driver's driving behavior, differences in driver's driving age will also affect their performance in driving behavior. Therefore, this study takes the driver's driving age into consideration when selecting the characteristic indicators.

4. Summary

Based on the bus GPS data, this paper selects the standard variance of vehicle speed, the number of idling speeds, the number of constant speeds, the number of rapid accelerations, the number of rapid decelerations, the number of speeding times and the driving age as characteristic indicators from the perspective of economy, safety and drivers themselves, so as to ensure the driving behavior evaluation. The scientificity and feasibility of the system. K-means clustering was used to preliminarily classify the types of driving behavior, and the driving behavior was divided into three types: aggressive type, general type and cautious type. Then, through the combination of entropy weight method and TOPSIS model, the objective and accurate evaluation of the driver's driving behavior of No. 1 bus in Rizhao City is realized. The results show that the weight of each driving behavior characteristic index is in line with the actual situation, and drivers with younger driving experience, unstable speed and intense driving can be accurately identified. At the same time, driving age as an important factor affecting the driving behavior of bus drivers should arouse our attention. In order to ensure the safety of passengers and improve the favorability of public transportation, urban traffic managers can reasonably set industry access thresholds, formulate a driver assessment system, and strictly control the driving level of drivers, so as to reduce bad driving behaviors and traffic accidents from the root. occur.

References

[1] The Traffic Management Bureau of the Ministry of Public Security. Statistical Report of Road Traffic Accidents (2014) [R]. Wuxi: Traffic Management Science Research Institute, Ministry of Public Security, 2015.

[2] Zhang Qian, Yang Peizhao, Yan Cilei, et al. Analysis of road traffic accidents in China [J]. Practical technology for automobile, 2016 (6): 7-9.

[3] Shen Xiaoyan, Li Xiaonan, Xie Pei, Xie Chenjiang. Statistical analysis and study of 886 dangerous goods tank vehicle road transport accidents [J]. China Safety Production Science and Technology, 2012,8 (11): 43-48.

[4] Iwona Grabarek and Sylwia Beczkowska. Analysis of Factors Determining Ergonomic Conditions of Driver's Workplace and Safety in Transport of Dangerous Goods[J]. Archives of Transport, 2012, 24(3) : 297-306.

[5] Wang Haixing, Wang Xiangyu, Wang Zhaoxian, Li Xuedong. Cluclustering analysis of dangerous cargo risk based on data mining [J]. Transportation System Engineering and Information, 2020,20 (01): 183-189.

[6] Wang Hongan, Chen Guoming. Defect marking method of magnetic flux leakage in oil and gas pipelines based on K-means clustering [J]. Science Technology and Engineering, 2020, 20 (21): 8643-8646.

[7] Zheng Xin, Cheng Yanmei, Ren Caifeng, Zhou Lizhi. Evaluation of the resource and environmental carrying capacity in Ordos city based on the entropy right TOPSIS model [J]. Ecological Science, 2020,39 (02): 95-103.

[8] Yin Chuanzhong, Zhang Xiangdong, Qu Siyuan, Zhao Ming, Liu Yongdong. Discussion on the evaluation of planned railway in Yangtze River Delta region based on entropy right TOPSIS model [J]. Railway Transportation and Economy, 2021,43 (01): 77-82,101.

[9] Yang Man, Wu Chaozhong, Zhang Hui, Li Siyao. Study on the factors affecting driving risk in driving safety incidents [J]. Traffic Information and Safety, 2018,36 (05): 34-39.

[10] Lei Cailin, Zhong Tianyi, Cai Xiaoyu, Tang Xiaoyong. Road safety evaluation method based on the OBD data of the Internet of Vehicles [J]. Highway and Transportation, 2019 (01): 30-36.

[11] Shi Kun, Zhang Yongjin, Sun Chaoyong. Reliability assessment of safe driving based on multi-state driving environment [J]. Operations Research and Fuzzy, 2018, 8(3) : 83-89.

[12] Zhang Xiaobo, Chen Xi. Analysis of driving behavior of commercial vehicles based on internet of vehicles [J]. Sci-Tech & Development of Enterprise, 2020, 36(5) : 61-63.

[13] Xu Zhiqi, Yao Yuan, Zhou Xingshe, et al. Research on driving safety evaluation technology based on fuzzy comprehensive evaluation [J]. Computer Measurement & Control, 2019, 27 (7): 261-265.

[14] Wang Keyin, Yang Yahui, Wang Sishan, et al. Research of clustering and recognition for driving style [J]. Journal of Hubei University of Automotive Technology, 2021, 35(3) : 1-6, 10.

[15] Zhang Zhihong. Research on driving behavior based on OBD data analysis [D]. Xi'an: Chang'an University, 2017.