ADVANCES IN TRANSPORTATION STUDIES
An International Journal

Special Editors: Zhixiang Hou

2020 Special Issue, Vol. 1

Contents

Advances in Transportation Studies an international Journal 2020 Special Issue, Vol. 1

ISBN 978-88-255-3186-2 – ISSN 1824-5463 – DOI 10.4399/97888255318621 – pagg. 3-12

Prediction method of driving risk in complex environment based on fuzzy comprehensive evaluation model

X.L. Chen[1] G.J. Jim[2]

[1]*Department of Mechanical Engineering, Zhejiang Industry Polytechnic College, Shaoxing 312000, China*
[2]*Department of Civil and Environmental Engineering, Seoul National University, Seoul 08826, Korea*

Abstract

In order to solve the problems of low prediction accuracy and poor real-time prediction in traditional driving risk prediction methods of vehicle, a fuzzy comprehensive evaluation model based driving risk prediction method in complex environment is proposed. In this method, the driving sample data are filtered first, and the incomplete and unstable driving data are removed. The vehicle's driving model is constructed with the processed data, and the force and dynamics of the vehicle's driving model are analyzed with the particle dynamics. The risk factor analysis set is constructed, the entropy weight method is used to determine the risk factor weight, and the five level classification method is used to construct the risk assessment set according to the analysis results of the weight method. The fuzzy evaluation matrix of vehicle's driving risk in complex environment is designed, and the risk prediction model is constructed to realize the prediction of vehicle's driving risk in complex environment. The experimental results show that the prediction accuracy of the proposed method is as high as 0.98, and the prediction is real-time and reliable.

Keywords- fuzzy comprehensive evaluation, model, complex environment, vehicle, driving, risk prediction

1. Introduction

In recent years, the number of motor vehicles in China has increased in a straight line, and the number of existing vehicles has reached 278 million. The increase of vehicles has led to an explosive increase in the number of traffic accidents. According to the official statement of the WHO, the number of road traffic accident deaths and injuries in the world has reached more than 1.1 million and more than 40 million, which has a great negative impact on the social security and economic development of each country. In addition, the road infrastructure in low-income countries is not advanced, and the probability ratio of road traffic accidents increases gradually. China is the largest developing country in the world, and the problem of road traffic safety is increasingly prominent [5]. According to the road traffic accident data of the Ministry of public security of China, in 2014, there were nearly 200000 traffic accidents in China, nearly 60000 people died in traffic accidents, nearly 220000 people were injured, and the economic loss reached 1075.43 million yuan [16]. Among them, there are more than 800 serious traffic accidents, with the death toll as high as 2900; there are two very serious traffic accidents, with 99 deaths. It can be seen that with the optimization of people's living standards, there are more and more vehicles, and the number of road traffic accidents shows a trend of rising but not falling. Therefore, improving vehicle safety is a very important thing [17].

At present, there are few researches on the prediction of vehicle's driving risk in the actual traffic scene. In Ali et al. [1], a driving risk prediction method of vehicle based on decision tree algorithm is proposed. In order to reduce the driver's driving risk, based on the analysis of the causes of driving risk, this paper deeply excavates and analyzes the driver's experience, vehicle condition, road condition and other influencing factors, designs the key influencing variables related to the vehicle's driving risk identification, and establishes the vehicle's driving risk identification model by using the decision tree algorithm, so as to realize the vehicle's driving risk prediction according to the model. This method has the problem of low prediction accuracy, so it is difficult to achieve the ideal application effect. In Guo et al. [4], a method of vehicle's driving risk prediction based on rough set and support vector machine is proposed. Firstly, the driving process is analyzed, and 15 driving risk assessment indicators are preliminarily selected. Then the historical driving data of 100 high-risk driving sections of vehicles are selected, and according to the rough set theory, combined with genetic algorithm and Johnson algorithm the evaluation items are reduced, to obtain 8 core indicators; finally, support vector machine (SVM) algorithm is used to establish the risk prediction model, and Matlab is used to carry out simulation test, but the method has high complexity and poor real-time prediction, which is difficult to meet the actual needs. Yu et al. [15] proposes a vehicle's driving risk prediction technology based on the improved Apriori algorithm. First of all, by analyzing the characteristics of vehicle's driving risk, the risk database is established, the relevant information of risk database is extracted. The Apriori algorithm is applied to deal with this problem, and the correlation and occurrence law of risk factors are found out. The spatial complexity of Apriori algorithm is reduced by using the improved candidate extraction method to improve the efficiency. The calculation results of Apriori algorithm are analyzed, the association rules between risk events and risk results are discovered, and the results are analyzed, to complete the prediction of vehicle driving risk. However, this method has the problems of low prediction accuracy and low real-time prediction.

However, the above methods are used to predict the driving risk of the vehicle in the ordinary environment, which is not suitable for the complex environment. The existing methods mainly analyze the driving control at the low level and the driving mode at the strategic level. There are many analysis contents and more requirements for the test conditions, which are difficult to some extent. Therefore, this paper focuses on these problems, and puts forward the driving risk prediction method based on the fuzzy comprehensive evaluation model. This method has the characteristics of high prediction accuracy and real-time performance, which can effectively solve the problems of traditional methods and improve the driving safety of vehicles in complex environment.

2. Analysis of vehicle's driving risk prediction method in complex environment

2.1. Selection and processing of vehicle's driving sample data

In order to improve the reliability of driving sample data, statistical principles are used to implement screening [6], and data irrelevant to the driving sample analysis of experimental vehicles are deleted.

(1) Remove incomplete data

Because the test items are not complete, the incomplete data is removed.

(2) Remove unstable data

If the data indicator value obtained by the detected target does not change abnormally based on the specified interval, it means that the data dispersion of this indicator is small and there is an example. According to this method of removing unstable data, sample data is removed.

$$\left\{\left|\frac{Y-Q}{\delta}\right|\right\} < \phi \tag{1}$$

where, Y describes the variance of a certain type of repeated detection data of the driver; Q describes the mean value of the variance of the Y-th repeated monitoring data; δ describes the standard deviation of the variance of the Y-th repeated monitoring data; ϕ describes the probability of the Y-th set.

2.2. Establishment of fuzzy comprehensive evaluation model of vehicle's driving safety state

The process of building the mathematical model of the fuzzy comprehensive evaluation method by using the fuzzy comprehensive evaluation theory is as follows: set the target layer A of the evaluation model, the factor set of the criteria layer A (the first level indicator) is set as $V = \{v_1, v_2, \dots, v_n\}$. The evaluation set (secondary indicator) of the scheme level is set as $U = \{u_1, u_2, \dots, u_n\}$; the weight set $\varpi = \{\varpi_1, \varpi_2, \dots, \varpi_n\}$ of related factor set V is built. A fuzzy evaluation matrix is established, assuming that the fuzzy evaluation vector of the j-th factor in the factor set related to the evaluation set U is $P_j = \{p_{j1}, p_{j2}, \dots, p_{jm}\}$. Then the fuzzy evaluation matrix is $P = (p_{ji})_{n \times m}$, which describes the relative relationship between the factor set V and the evaluation set U; according to the fuzzy operator $C = \varpi \cdot p$, the target layer A of the evaluation model is analyzed and evaluated [10].

2.2.1. Establishment of the analysis set of vehicle's driving risk factors

In this paper, CA1046L2 light truck is set as the research goal, and the particle dynamic model of the vehicle model is constructed to obtain the vehicle speed data based on the complex environment. When the vehicle is running in the complex environment, there are high requirements for the driver's overall control of driving speed, road conditions, wind direction interference and vehicle load weight M when the vehicle is completed. Particle dynamics is used to carry out force and dynamics analysis on the moving vehicle model [18]. The stress model is shown in Figure 1 and Figure 2:

In Figure 1, the horizontal direction of the vehicle particle model is successively affected by the engine power H_d and the horizontal wind H_u. The motion analysis based on dynamics includes the horizontal velocity d_x, vertical velocity d_y, tangential acceleration d_r and normal acceleration d_n of the particle. Where θ describes the angle between the line between the particle O and the curvature center R and the X axis in the horizontal direction, with radian units. φ describes the radius of curvature of the particle around the center [19, 8].

In Figure 2, considering the slope level of the road surface, the road surface where the vehicle operates is set to have an inward slope of μ angle.

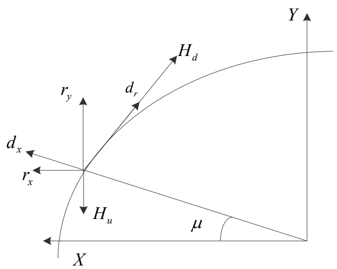

Fig. 1 - Horizontal dynamic motion model

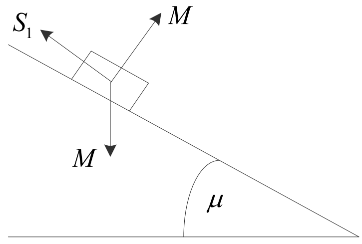

Fig. 2 - Vertical interface force model

This is that the force of the particles based on the vertical interface belongs to the plane force system, which is successively affected by the vehicle load in the vertical direction, the forward and reverse friction force S_1 of the particles and the supporting force of the road surface on the particle model of the vehicle. According to the dynamic equilibrium theory in dynamics, it can build the equilibrium equation with automobile as the research goal [20].

Based on the force analysis of horizontal and cross-section, according to the friction coefficient of the ground, vehicle gravity, cross wind and other elements, the following object force equilibrium equation of three-dimensional space is constructed:

$$\begin{cases} S_1\cos\mu - S_2\cos\mu - H_u\cos d + S_2\sin d = Md \\ S_2\cos\mu + H_u\sin\mu = Md \\ \sin\mu + H_u = Md \end{cases} \tag{2}$$

where, the driving force of the device engine is S_2. Based on the above calculation, nine factors, i.e. load taking mass, vehicle speed, longitudinal safety distance, transverse safety distance, wheel pressure, steering power [7], brake pipe leakage, brake shoe wear and abnormal light, are set as the evaluation indicators of driving risk state of vehicles in complex environment. Then, the factor set of criteria layer in the evaluation model is $V = \{v_1, v_2, v_3, \dots, v_9\}$.

2.2.2. Weight calculation of risk factors based on entropy weight method

Information entropy describes the abnormal level and disorder of the system, which is regarded as the probability weighted statistical mean of information quantity. Then:

$$\varpi = -\sum_{j=1}^{m} q_j \lg q_j \tag{3}$$

where q_j describes the probability of time.

Entropy method is a weight operation method to describe the local difference, and it describes the key level according to the deviation level between the observation values of the same indicator. Compared with the previous subjective weighting method and analytic hierarchy process, entropy weighting method uses information entropy to evaluate the order and utility of obtaining, and prevents subjectivity of each factor weight as much as possible, so the evaluation result is similar to the actual situation [14, 2].

Based on the prediction of vehicle’s driving risk, the weight is set through the operation of "entropy", that is, the weight of each indicator is set according to the deviation level of each observation standard value. If the indicator value of each evaluation target has a large difference and a small entropy value, it means that the indicator carries a large amount of effective information and a large weight; if the indicator gap is small and the entropy value is large, it means that the indicator carries a small amount of effective information and a small weight.

Tab. 1 - 1~9 evaluation scale information

Scale	Content
1	Describe b_j as important as b_i
3	Describe the importance of b_j and b_i
5	Describe the importance of b_j and b_i
7	Describe the importance of b_j and b_i
9	Describe the extreme degree of importance between b_j and b_i
2/4/6/8	There is the same relationship as above

Firstly, the evaluation matrix is constructed according to the evaluation results. For example, the second level indicator set of road traffic V_2 in the first level indicator is $\{u_1, u_2, u_3, u_4\}$, which has nine second level indicators.

According to the evaluation scale (Table 1), the evaluation of these nine second level indicators is established as a matrix B:

$$B = \begin{bmatrix} 1 & 5/4 & 6/4 & 2 & 3 & 4 & 9/2 & 5 & 6 \\ 4/5 & 1 & 8/5 & 2 & 9/4 & 3 & 4 & 4 & 5 \\ 4/6 & 5/8 & 1 & 6/5 & 2 & 5/2 & 3 & 4 & 5 \\ 1/2 & 1/2 & 5/6 & 1 & 7/6 & 2 & 5/2 & 3 & 4 \\ 1/3 & 4/9 & 1/26/7 & 1 & 5/4 & 2 & 5/2 & 3 \\ 1/4 & 1/3 & 2/51/2 & 1/2 & 1 & 2 & 2 & 3 \\ 2/9 & 1/4 & 1/32/5 & 1/2 & 1/2 & 1 & 3/2 & 2 \\ 1/5 & 1/4 & 1/41/3 & 2/5 & 1/22/3 & 1 & 4/3 \\ 1/6 & 1/5 & 1/51/4 & 1/3 & 1/31/2 & 3/4 & 1 \end{bmatrix} \tag{4}$$

The entropy of the driving risk evaluation indicator under the j-th complex environment is $G(I)$:

$$G(I) = -\sum_{i=1}^{m} g_{ji} \left[\lg(b_{ji})\right] \tag{5}$$

where, b_{ji} describes the parameters in the matrix of vehicle's driving risk assessment in complex environment; g_{ji} describes the entropy factor, $g_{ji} = b_{ji}/\sum_{i=1}^{m} b_{ji}$ and g_{ji} conforms to $\sum_{i=1}^{m} b_{ji} = 1$. If b_{ji} is 0, $G(I)$ is equal to 0.

After setting the entropy value of the vehicle's driving risk evaluation indicator under the j-th complex environment, the entropy weight of the j-th indicator is:

$$\varpi_j = \frac{1-G(I)}{\sum_{j=1}^{n}(G(I))} \qquad 0 \le \varpi_j, \sum_{j=1}^{n} \varpi_j = 1 \tag{6}$$

where, n describes the number of indicators.

The entropy weight of each element in each row of judgment matrix B is calculated and the normalization is implemented to obtain the weight vector ϖ : $\varpi = 0.226,0.189,0.160,0.124,0.095,0.082,0.054,0.039,0.030$. MATLAB is used to get the maximum eigenvalue $\beta_{max} = 9.083$ of judgment matrix B. The consistency of judgment matrix is checked as [11, 12]:

$$X_1 = \frac{\beta_{max} - m}{m-1} = \frac{9.083-9}{9-1} = 0.010 \tag{7}$$

X_2 represents the average random consistency indicator of judgment matrix. For judgment matrix of order 1-9, the value of X_2 is shown in Table 2.

The random consistency ratio X_3 of judgment matrix B is:

$$X_3 = \frac{X_1}{X_2} = \frac{0.010}{1.46} = 0.007 < 0.1 \tag{8}$$

According to the criteria, we can know that the difference level of decision matrix B is acceptable. So the criteria level factor centralization vector ϖ is:

$$\varpi = [0.227,0.190,0.161,0.125,0.096,0.083,0.055,0.040,0.031] \tag{9}$$

Tab. 2 - 1-9 X_2 values of order judgment matrix

m	X_2
1	0
2	0
3	0.59
4	1.8
5	1.13
6	1.25
7	1.33
8	1.42
9	1.46

Tab. 3 - Indicators of factor evaluation

Factors	u_1	u_2	u_3	u_4	u_5
v_1	0-701	701-1751	1751-2801	2801-3501	More than 3501
v_2	0-31	31-51	51-81	81-91	More than 91
v_3	Greater than 1.51	1.51-1.41	1.41-1.21	1.21-1.01	Less than 1.01
v_4	0.81-1.1	0.61-0.81	0.41-0.61	0.21-0.41	0-0.21
v_5	531	581	501	More than 641	Less than 451
v_6	More than 601	601-581	581-551	551-451	Less than 451
v_7	-	0	-	-	1.1
v_8	-	0	-	-	1.1
v_9	-	0	-	-	1.1

2.2.3. Establishment of a risk prediction model

The evaluation set of the scheme level belongs to the language description of vehicle's driving risk state evaluation [3, 13]. The five level classification method is used to build the evaluation set of vehicle's driving risk state, then $U = \{u_1, u_2, u_3, u_4, u_5\}$, where u_1, u_2, u_3, u_4 and u_5 successively describe very safe, safe, general safe, little dangerous and dangerous. According to the technical conditions for motor vehicle operation safety (GB7258-2018) and road test, the evaluation indicators of the set of factors are shown in Table 3.

According to the scheme level evaluation set and factor evaluation indicator (Table 3), the factor set in the criteria level is analyzed and evaluated, to obtain the fuzzy evaluation matrix P of vehicle's driving risk in complex environment.

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{15} \\ P_{21} & P_{22} & \cdots & P_{25} \\ \vdots & \vdots & & \vdots \\ P_{91} & P_{92} & \cdots & P_{95} \end{bmatrix} \tag{10}$$

The fuzzy evaluation set $C = \varpi \cdot P = [c_1, c_2, c_3, c_4, c_5]$ is obtained by using the weight vector ϖ and the fuzzy evaluation matrix P operation of the factor set of the criteria level. Assuming that the vector established by the rank of each factor in the fuzzy evaluation set is $\{1,2,3,4,5\}$, then $U = [u_1, u_2, u_3, u_4, u_5] = \{1,2,3,4,5\}$, the weighted average principle operation is used to obtain the maximum membership degree c of the fuzzy evaluation set C of vehicle's driving risk status. According to the results, the prediction model of vehicle's driving risk is constructed as follows:

$$Q = \sum_{i=1}^{5} c_i c - P \sum_{j=1}^{n} \big(G(I)\big) \tag{11}$$

According to the maximum "close degree" criteria, according to the maximum membership degree c of vehicle's driving dangerous state, the dangerous state of vehicle's driving is evaluated, so as to complete the prediction of dangerous state of vehicle's driving in complex environment.

3. Driving risk prediction experiment and results in complex environment

In order to verify the practical application effect of vehicle's driving risk prediction technology in complex environment based on fuzzy comprehensive evaluation model, a simulation experiment is carried out. The overall scheme of the experiment is designed as follows:

(1) Setting up the experimental environment, the operating system is Windows 10, the simulation software is Matlab, the CPU is Intel Xeon, the memory capacity is 8GB, and the external hard disk capacity is 4T.

(2) Experimental data: the location of complex environment is a one-way street in a park on December 12, 2019. Most people cross the one-way street. In order to obtain the running state parameters of the vehicle in the complex environment, based on the wince environment of UNO2171 industrial computer, this paper sets EVC as a development tool, uses the embedded development method, integrates the multi-sensor acquisition method and GPS, obtains the experimental data, and proposes incomplete data and noise data, so as to ensure that the data can run on the simulation platform.

(3) Experimental methods: the methods in Ali et al. [1] and Guo et al. [4] and the proposed method are selected for comparative experiments.

(4) Experimental indicators:

Prediction accuracy: prediction accuracy refers to the fitting degree between the prediction results of different research methods and the actual situation. The higher the prediction accuracy is, the closer the prediction results are to the actual situation, and the stronger the authenticity of the prediction results is.

Real-time prediction: real-time refers to the response time of different methods to achieve the expected goal, the better the real-time prediction is, the shorter the output time of risk prediction results is, and the higher the vehicle safety is.

The specific experimental process is as follows:

(1) Verification of prediction accuracy of the method in this paper

In the interception experiment, the driving state parameters of the evaluation factor set in a certain time period are shown in Table 4, and the driving state of the vehicle in this time period is predicted by the method in this paper.

Tab. 4 - Driving state parameters of the corresponding evaluation factor set in a certain period of time in the experiment

Time	Loading quality /kg	The speed of the car / $km \cdot h^{-1}$	Horizontal distance /m	The longitudinal distance /m	Tire pressure /kPa	Steering /kPa	Pipeline leak	Shoe wear	The light failure
2019-12-12-1:00	1686	15.501	1.507	120.898	514	576	-	-	-
2019-12-12-2:00	1691	27.001	1.802	95.708	515	571	-	-	-
2019-12-12-3:00	1696	27.301	1.715	58.002	574	574	-	-	1.1
2019-12-12-4:00	1696	18.701	1.894	33.255	515	577	-	-	1.1
2019-12-12-5:00	1701	10.801	1.109	14.063	517	576	-	-	1.1
2019-12-12-6:00	1711	2.601	0.906	8.453	519	576	-	-	1.1
2019-12-12-7:00	1716	14.601	2.151	92.606	519	575	-	-	-
2019-12-12-8:00	1706	16.301	2.054	71.772	518	575	-	-	-
2019-12-12-9:00	1696	18.501	1.986	48.381	516	573	-	-	-
2019-12-12-10:00	1696	18.001	2.015	65.483	515	571	-	-	-

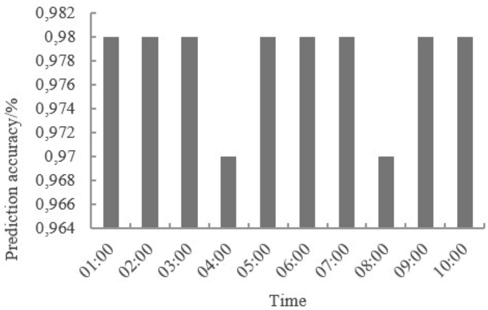

Fig. 3 - The method in this paper predicts the test results of structural reliability

The prediction results of the proposed method is compared with the actual situation, so as to judge the reliability of the prediction results of the proposed method, as shown in Figure 3.

As shown in Figure 3, the prediction accuracy of the method in this paper is as high as 0.98 in the complex environment of 1:00 ~ 10:00 on December 12, 2019, so the prediction results of the method in this paper are reliable.

(2) Comparison of prediction accuracy of different methods

In order to further test the prediction performance of the proposed method, two kinds of complex environments are set up: one is a one-way street in a park on December 12, 2019, where most people cross the road; the other is a vehicle angle of a city's main traffic line on December 12, 2019.

The method in this paper, the method in Ali et al. [1] and the method in Guo et al. [4] are used to predict the same prediction target, and the prediction accuracy of the three methods is tested. The results are shown in Figure 4 and Figure 5.

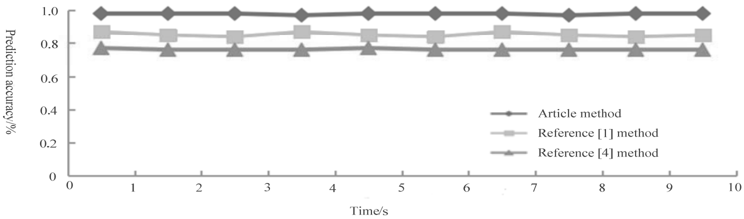

Fig. 4 - Prediction accuracy of three methods in complex environment 1

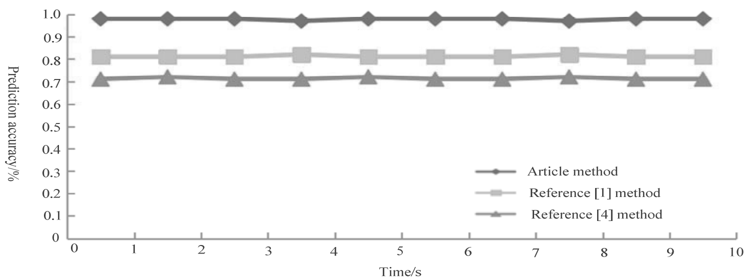

Fig. 5 - Prediction accuracy of three methods in complex environment 2

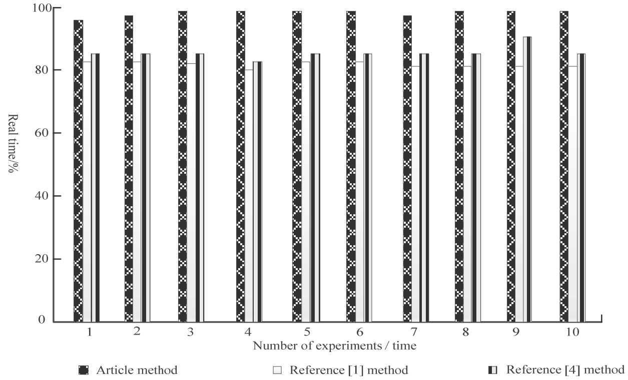

Fig. 6 - Three methods predict real-time performance

In Figure 4 and Figure 5, under the two complex environment settings, the maximum prediction accuracy of the proposed method for vehicle's driving risk is 0.98, and the maximum prediction accuracy of the methods in Ali et al. [1] and Guo et al. [4] for vehicle's driving risk is always lower than that of the proposed method, which is not more than 0.90. In different complex environments, the prediction accuracy of the proposed method for driving risk is the best.

(3) Comparison of real-time prediction by different methods

In the above experiments, three methods are tested to predict the real-time performance, and the results are shown in Figure 6.

In Figure 6, the real-time indicator of the proposed method for vehicle's driving risk prediction in complex environment is as high as 98.88%, and the highest predictive real-time indicator of methods in Ali et al. [1] and Guo et al. [4] is 82.55% and 88.89%, respectively. In contrast, the proposed method has the highest real-time performance.

In conclusion, the prediction accuracy of the proposed method is as high as 0.98, and the prediction is real-time and reliable. The main reason why this method can realize the prediction of driving risk in the complex environment is that it mainly analyzes the multi-level driving control and strategic driving mode, which has many analysis contents and needs more test conditions. Besides, the incomplete and unstable driving data are removed, which improves the data quality and lays the foundation for the next step of risk prediction. The vehicle's driving model is constructed with the processed data, and the force and dynamics of the vehicle's driving model are analyzed with the particle dynamics. The risk factor analysis set is constructed, the entropy weight method is used to determine the risk factor weight, and the five level classification method is used to construct the risk assessment set according to the analysis results of the weight method. The fuzzy evaluation matrix of vehicle's driving risk in complex environment is designed and the risk prediction model is constructed. Through these steps, the driving risk of vehicles in complex environment can be accurately predicted.

4. Conclusions

In this paper, a method based on the fuzzy comprehensive evaluation model is proposed to predict the driving risk of complex vehicles, and it is applied to the experiment to verify its effectiveness. It is verified that the accuracy of the method is as high as 0.98 in the complex environment of 1:00 ~ 10:00 on December 12, 2019, so the prediction result of the proposed method is credible. Under the two kinds of complex environment settings, the maximum prediction accuracy of the proposed method is 0.98, and the prediction accuracy is high. In the complex environment, the real-time indicator of the proposed method is 98.88%, and the real-time prediction is high. The proposed method can be used to predict the driving risk of vehicles in complex environment with high accuracy and real-time.

Acknowledgments

This work was supported by Shaoxing Public Welfare Technology Application Research Plan Project no.2015B70011

References

1. Ali, A.N., Sang, H.L., Javaan, C. (2017). Vehicle's driving risk prediction method based on decision tree algorithm. Iet Image Processing 11(5), PP. 292-300.
2. Altche, F., Qian, X. (2017). An Algorithm for Supervised Driving of Cooperative Semi-Autonomous Vehicles (Extended). IEEE Transactions on Intelligent Transportation Systems PP, 99, pp. 1-13.
3. Ba, Y.T., Zhang, W., Peng, Q.J. (2016). Risk-taking on the road and in the mind: behavioural and neural patterns of decision making between risky and safe drivers. Ergonomics, 59(1), PP. 27-38.
4. Guo, Y., Yu, S. Z., Li, H. (2019). Vehicle's driving risk prediction method based on rough set and support vector machine. Journal of Power Supply 17(3), PP. 172-181.
5. Hu, X.C., Witold, P., Wang, X.M. (2017). Granular Fuzzy Rule-based Models: A Study in a Comprehensive Evaluation of Fuzzy Models. IEEE Transactions on Fuzzy Systems, 99, pp. 1-11.
6. Ji, G.J., Wang, Y.L. (2016). Study on control strategy of energy-saving elevator with super capacitor based on fuzzy control. Chinese Journal of Power Sources 40(08), PP. 1673-1674+1730.
7. Noah, J. (2017). From Trolleys to Risk: Models for Ethical Autonomous Driving. American Journal of Public Health 107(4), pp. 496-496.
8. Peter, W., Anna, W., Hans, P. H. (2017). SUV driving "masculinizes" risk behavior in females: a public health challenge. Wiener Klinische Wochenschrift 129(17), pp. 625-629.
9. Qin, Z.C., Wan, T., Zhao, H.Q. (2017). Hybrid clustering of data and vague concepts based on labels semantics. Annals of Operations Research, 256(3), PP. 1-24.
10. Richard, S. J., Robert, J. (2016). Comment on 'The Global Impacts of Extreme Sea-Level Rise: A Comprehensive Economic Assessment'. Environmental & Resource Economics, 64(2), pp. 225-253.
11. Sun, J.S., Chae, H.S., Yi, K.S. (2018). Stochastic Model Predictive Control for Lane Change Decision of Automated Driving Vehicles. IEEE Transactions on Vehicular Technology, 99, PP. 1-12.
12. Sukru, K., Diane, E.G., Neil, J.M. (2016). Driving a better driving experience: a questionnaire survey of older compared with younger drivers. Ergonomics, 60(4), PP. 1-8.
13. Song, G., Krzysztof, J., Daniel, R. M. (2017). A data-synthesis-driven method for detecting and extracting vague cognitive regions. Intern. Journal of Geographical Information Science, 31(6), 1-27.
14. Wu, H. L., Ren, X. Y. (2016). Specific Data Mining Algorithm Based on Fuzzy Constraint Database.Computer Simulation 33(10), pp. 240-243+375.
15. Yu, W.S., Ling, J.S., Guo, L.W. (2016). Vehicle's driving risk prediction technology based on improved Apriori algorithm. Iet Generation Transmission & Distribution, 10(15), PP. 3855-3864.
16. Zhou, L.T., Li, H.X., Sun, K.B. (2016). Teaching performance evaluation by means of a hierarchical multifactorial evaluation model based on type-2 fuzzy sets. Applied Intelligence 46(1), pp. 1-11.
17. Eustace, D., Alanazi, F, Hovey, P.W. (2019). Investigation of the effect of vehicle color on safety, Advances in Transportation Studies, an International Journal, 47, pp. 77-88.
18. Zhang, L.N., Ma, Y., Zhang, J. J. (2016). Comparison and Simulation of PID control and fuzzy PID control in motor speed control system. Automation & Instrumentation 46(7), pp. 52-54.
19. Zheng, C., Liang, L., Bing, j.Y. (2016). Multimode Energy Management for Plug-In Hybrid Electric Buses Based on Driving Cycles Prediction. IEEE Transactions on Intelligent Transportation Systems 17(10), pp. 2811-2821.
20. Zhou, X. (2018). Colimit in Category of Fuzzy Modules. Journal of Jilin University (Science Edition) 56(04), pp. 799-804.

Advances in Transportation Studies an international Journal 2020 Special Issue, Vol. 1

ISBN 978-88-255-3186-2 – ISSN 1824-5463 – DOI 10.4399/97888255318622 – pagg. 13-22

Modeling and analysis of influencing factors for traffic safety accident considering information entropy model

J. Liu[1] Q.G. Ma[2] N. Zhou[3] B. Chen[4]

[1]*Aliyun School of Big Data, Changzhou University, Changzhou 213164, China*
[2]*Changzhou University HuaiDe College, Taizhou 214500, China*
[3]*School of Petroleum Engineering, Changzhou University, Changzhou, Jiangsu 213164, China*
[4]*China Academy of Safety Production Research, Beijing 100012, China*

Abstract

In order to solve the problems of poor correlation and large error of analysis results in the existing analysis methods of influencing factors of traffic safety accidents, a new analysis method of influencing factors of traffic safety accidents is constructed by considering the information entropy model. Collect the information of traffic safety accident influencing factors, and get the characteristics of traffic safety accident influencing factors. The information entropy model is used to calculate the maximum eigenvalue of the judgment matrix, and the eigenvector is used as the corresponding weight of each factor to realize the construction of the influencing factor model of traffic safety accidents. The influencing factors of traffic safety accidents are analyzed by the established model. The experimental results show that the model has high factor correlation performance and low analysis error, which can be widely used in urban traffic safety management.

Keywords - information entropy model, traffic accidents, influencing factors, modeling analysis

1. Introduction

At present, traffic accidents have become a global recognized problem, and people are more and more concerned about road safety [12]. Reducing the number of traffic accidents and improving the safety level of roads have become the focus and urgent needs of the whole society. Traffic safety accidents have already become the killer of destroying people's property and life safety. The scope of influence, the severity of consequences and the high frequency of the damage are incomparable to other safety accidents [15]. In order to achieve traffic harmony, we must try our best to reduce the occurrence of traffic accidents and ensure the road safety of the people to the greatest extent. China also regards traffic harmony as a part of building a harmonious society, and actively creates a harmonious traffic situation that is fast, convenient, smooth, safe, orderly and civilized, and abides by the rules [19]. In order to realize the harmony of traffic, we must take unimpeded traffic as the purpose, safety as the guarantee, order as the safety and rule as the basis, strive to realize the harmony among road vehicles, everyone, vehicle and people, and finally realize the harmony among the natural environment, road, vehicle and people, so as to embody the unity of value and life, the unity of safety and rapidity, the unity of rules and vitality, and the unity of rule of law and civilization, to build a harmonious road culture and promote the promotion of social quality and civilization. In order to achieve this goal, it is necessary to analyze the influencing factors of traffic safety accidents, so as to accurately obtain the influencing factors of traffic safety accidents, and reduce the probability of occurrence [1].

In Deng and Zeng [5], a modeling and analysis method for influencing factors of traffic safety accidents based on AHP and hybrid Apriori-Genetic algorithm is proposed. In this method, AHP is used to rank the influencing factors of traffic accidents according to the importance, and the influencing factors of traffic safety accidents are analyzed objectively. Hybrid Apriori-Genetic algorithm is used to calculate the association rules between the influencing factors of traffic safety accidents. The analysis model is built to analyze the influencing factors of traffic accidents, but the accuracy of the method needs to be further improved. In Ge et al. [7], a method of modeling and analyzing the influencing factors of traffic safety accidents based on PreScan is proposed. In this method, the acceleration characteristics of vehicles are calculated by integral graph method, and the detection of rear and side vehicles is realized by combining with cascade classifier to determine the cause of traffic safety accidents. Based on the theory of MeanShift, the texture features of the influencing factors of traffic safety accidents are collected. At last, the influencing factor analysis model is constructed by using PreScan to complete the modeling analysis of the influencing factors of traffic safety accidents. However, the analysis correlation of this method is poor and the analysis results obtained are not accurate. Zhu et al. [20] propose a modeling and analysis method for influencing factors of traffic safety accident based on scene control features, which analyzes the complexity and coupling of traffic safety accidents, and constructs the control structure of traffic safety demand scene. Based on the scene control features, this paper analyzes the variable relationship between the influencing factors of traffic safety accidents, constructs the influencing factors model of traffic safety accident, and completes the modeling and analysis of the influencing factors of traffic safety accidents, but this method has the problem of large error in the analysis results.

In view of the problems existing in the above methods, a new modeling and analysis method for influencing factors of traffic safety accident is studied by considering the information entropy model.

2. Design of the modeling and analysis method for influencing factors of traffic safety accident considering information entropy model

2.1. Information collection and feature acquisition of influencing factors of traffic safety accidents

2.1.1. Information collection of influencing factors of traffic safety accidents

In the database of traffic safety accidents, the information of influencing factors of traffic safety accidents is collected, including the investigation information of accidents after traffic safety accidents and the field information recorded during traffic safety accidents, and the influencing factors of traffic safety accidents are described preliminarily [7]. Among them, the investigation information of the accident after the occurrence of the traffic safety accident refers to the information obtained from the analysis of the scene of the traffic safety accident, including witness testimony, inquiry of the parties, analysis of the scattered objects on the scene, road trace collection, survey environment, etc., as well as the information screened from the related database, including the vehicle insurance records, the attributes of the accident related parties, the records of the previous legal violations, and maintenance vehicle information, etc.; the on-site information recorded in case of traffic safety accident includes calling detector, video and other monitoring information to obtain the real-time data in case of traffic safety accident (the specific form of accident, accident environment, speed, etc.) [20]. The collected influencing factor information of traffic safety accident is sorted into influencing factor information table of traffic safety accident, as shown in Table 1.

Tab. 1 - Information of influencing factors of traffic safety accidents

Serial number	data item	Serial number	data item
1	Accident time	21	cause of death
2	Accident location	22	Number plates
3	Casualties	23	Vehicle number
4	Accident pattern	24	Real carrying number
5	Scene form	25	Legal status of vehicle
6	Whether dangerous goods are loaded	26	Vehicle safety condition
7	Consequence of dangerous goods accident	27	Vehicle running status
8	Cause of accident preliminary investigation	28	Nature of vehicle use
9	Direct property loss	29	Mileage of highway passenger transport section
10	Weather	30	Operation mode of highway passenger transport
11	Visibility	31	Types of dangerous goods carried
12	Whether to detect escape accidents	32	Road type
13	Pavement condition	33	Administrative grade of highway
14	Road table condition	34	terrain
15	Traffic signal mode	35	Road alignment
16	Lighting conditions	36	Road section type
17	Cause of accident identification	37	Physical isolation of road
18	ID number	38	pavement structure
19	Code of household registration administrative division	39	Type of roadside protection facilities
20	Party attribute	40	Name unit name

According to the information table of influencing factors of traffic safety accidents, the contents of influencing factors of traffic safety accidents are obtained, and the influencing factors of traffic safety accidents are preliminarily described [9].

2.1.2. Characteristics acquisition of influencing factors of traffic safety accidents

According to the obtained influencing factors content of traffic safety accident, the characteristics of influencing factors of traffic safety accident are obtained and summarized, including the specific distribution characteristics of time, type, vehicle and personnel, environment and road [6].

The characteristic function of the influencing factors of traffic safety accidents is as follows:

$$Z_i = \sum_j w_j a_{ij} + k_j \tag{1}$$

In the formula, w_j is the correlation weight of influencing factors, a_{ij} is the probability of influencing factors of traffic safety accidents, and k_j is the bias parameter of influencing factors.

Then the characteristic output calculation formula of the influencing factors is:

$$f(t) = p(\sum_j w_j a_{ij} + k_j) \tag{2}$$

In the formula, $f(t)$ represents the characteristic output of the influencing factors of traffic accidents, and p represents the characteristic output factor.

If the maximum number of influencing factors of traffic safety accidents is K, the output probability of influencing factors is a^K, and the input function of influencing factors of traffic safety accidents is $y(x)$, then the characteristic loss function of influencing factors of traffic safety is:

$$L(x) = \frac{1}{2}\sum_x \|y(x) - a^K(x)\|^2 \tag{3}$$

After completing the loss function calculation, the error of influencing factors of traffic safety accident is determined as:

$$\delta^K = \frac{\partial L(x)}{\partial a^K} \cdot \frac{\partial a^K}{\partial Z^K} = \nabla_a L(x) \cdot f(t) \tag{4}$$

The influencing factors of traffic safety accidents are as follows:

$$C = Q(\sum_K w_K a_K - \delta^K) \tag{5}$$

The specific distribution characteristics of types are shown in Table 2, the specific distribution characteristics of vehicles and personnel are shown in Table 3, and the specific distribution characteristics of environment and road are shown in Table 4 [14].

Tab. 2 - Specific distribution characteristics of types

Features	Classification	Concrete content
Specific distribution characteristics of types	Cause of accident	The frequency of accidents, the number of injuries and deaths on Friday and Saturday are significantly lower than those on Friday and Saturday
	Accident pattern	The rest of Sunday and weekday are basically the same from Monday to Thursday. The road traffic safety level on rest days is generally better than that on working days.

Tab. 3 - Specific distribution characteristics of vehicles and personnel

Features	Classification	Concrete Content
Specific distribution characteristics of vehicles and personnel	Mode of transportation	China's road traffic safety management department determines the cause of the accident according to the motor vehicle illegal fault, motor vehicle non illegal fault, non motor vehicle illegal, pedestrian passenger illegal and road reasons. In the case of multiple reasons, it is generally classified according to the main responsibility, that is to say, the causes of the accident are determined according to the main responsible party and the order of motor vehicles, non motor vehicles, pedestrians and others under the same responsibility.
	Driving age distribution	Road traffic accidents are mainly divided into collision, rolling, scraping, overturning, falling, fire and other forms. Front collision, side collision and tail collision are the main accident forms, and three types of forms account for the total accident frequency and injured persons
	Age distribution	Number and 80% of deaths. The contradiction between motor vehicles and pedestrians in the limited road space is further prominent.

Tab. 4 -Specific distribution characteristics of environment and road

Features	Classification	Concrete Content
Specific distribution characteristics of environment and road	Distribution characteristics of traffic accidents in road sections	Driving motor vehicles (buses, vans) occupy an absolute majority. Pedestrian accidents cause higher death rate and more serious accidents.
	Influence of physical isolation form on road traffic safety	In the current accident information collection data system, driving age is an objective indicator to reflect driving experience and level. The data distribution confirms that low-age drivers are the frequent traffic accident population, and the overall traffic safety potential risk of low-age drivers is high.
	Traffic control	The age of the person responsible for the accident is basically consistent in the distribution of accident frequency, number of injuries and number of deaths, and the distribution characteristics are basically normal. As the age group composition of drivers is consistent with that of social activities, the age group of 26-40 is not only the most active group, but also the group with frequent accidents.
	The relationship between the day night distribution of traffic accidents and lighting conditions	More than 90% of the accident frequency, the number of injured and the number of dead at the intersection of three or four branches are consistent with the actual road network structure.

2.2. Construction of influencing factor model of traffic safety accident

2.2.1. Construction of decision table

According to the characteristics of influencing factors of traffic safety accidents, the model of influencing factors of traffic safety accidents is constructed based on information entropy model [2]. First of all, it needs to build a decision table, which is represented by the following formula:

$$S = C, U \cup D \tag{6}$$

where, S represents decision table; U represents decision information set; C represents condition attribute set; D represents decision attribute set. In the construction of influencing factor model of traffic safety accident, the condition attribute set is traffic safety accident, and the decision attribute set is accident influencing factor [3, 16].

2.2.2. Construction of judgment matrix

Through the reduction algorithm to reduce the condition attributes contained in the decision table, the core attribute set is obtained. Then, the judgment matrix is constructed, and the core attribute set is set as $C(c_1, c_2, \ldots, c_m)$, where c_{ij} represents the relative importance ratio of element c_i and element c_j, and its value is shown in Table 6 [10, 8, 17]. According to the information entropy model, the importance corresponding calculation is carried out for each core attribute, and the judgment matrix is obtained by combining the above table:

$$T = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1m} \\ c_{21} & c_{22} & \cdots & c_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ c_{m1} & c_{m2} & \cdots & c_{mm} \end{bmatrix} \tag{7}$$

where T represents the judgment matrix.

Tab. 5 - Decision table

Serial number	Driving age	Age	Fatigue driving	Nature of vehicle use	Road alignment	Hour	Weather	Accident type
1	8	36	yes	Operation vehicle	General bending	1	Yin	Summary procedure accident
2	11	35	no	Private car	Flatness	11	Fine	General accident
3	8	42	no	Operation vehicle	General slope	5	rain	Summary procedure accident
4	7	36	yes	Operation vehicle	Flatness	3	Snow	General accident
5	2	28	no	Private car	Flatness	8	Fine	Summary procedure accident
6	11	42	yes	Operation vehicle	Flatness	2	Fog	General accident
7	1	33	yes	Private car	Flatness	15	Fine	Summary procedure accident
8	7	36	no	Private car	General curved slope	18	Snow	General accident
9	1	23	yes	Private car	Flatness	1	Fine	Summary procedure accident
10	8	36	yes	Private car	Flatness	23	Yin	Summary procedure accident

Tab. 6 - Value of c_{ij}

Serial number	Scale value	Meaning
1	1	Represents the comparison of c_i and c_j with equal importance
2	3	Represents the comparison between c_i and c_j, c_i is slightly more important than c_j
3	5	Represents the comparison between c_i and c_j, c_i is obviously more important than c_j
4	7	Represents the comparison between c_i and c_j, and c_i is more important than c_j
5	9	Represents the comparison between c_i and c_j, c_i is extremely important than c_j
6	8,6,4,2	Mean values of 1-3, 3-5, 5-7, 7-9 of adjacent judgments
7	Reciprocal	Judgment of comparison between c_i and c_j :$c_{ji} = \frac{1}{c_{ij}}$

2.2.3. Importance calculation based on information entropy model

The maximum eigenvalue of the judgment matrix is calculated, and its eigenvector is taken as the corresponding weight of each factor.

$$U_T = g(Rs_T) \tag{8}$$

In the formula, U_T represents the output vector of the matrix, and R represents the weight matrix of the factors affecting traffic safety.

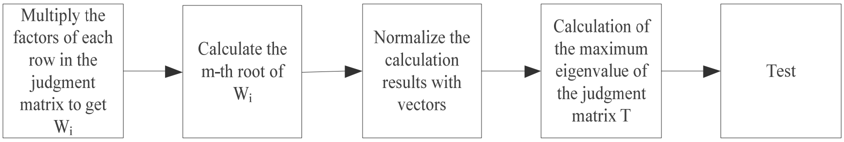

Fig. 1- Calculation steps of maximum characteristic root

The calculation formula of influencing factor dimension s_T is as follows:

$$s_T = f(Wx_t + Hs_{t-1}) \tag{9}$$

The corresponding weight calculation formula of each factor is:

$$\sigma_n = \frac{1}{m}\sum_{i=1}^{m}(U_T - \lambda_n) \tag{10}$$

The specific steps of importance calculation are shown in Fig. 1 [4,11,13]. Through the above steps, the construction of influencing factor model of traffic safety accident is completed [18].

3. Experimental verification

3.1. Experiment process

In this paper, the modeling and analysis method for influencing factors of traffic safety accident by considering information entropy model is used to carry out the experiment on modeling and analysis of influencing factors of traffic safety accident. The experimental data are shown in Table 7.

The above experimental data are used to model and analyze the influencing factors of traffic safety accidents. In order to ensure the contrast of the experimental results, the proposed method is compared with the methods in Deng and Zeng [5], Ge et al. [7] and Zhu et al. [20]. The indexes of experimental comparison are factor correlation performance and analysis efficiency. The higher the factor correlation degree and the analysis efficiency are, the higher the effectiveness of the analysis method is.

Tab. 7- Experimental data

	classification	Number of injuries or deaths			
Road table condition	Dry	536	717	238	463
	Damp	44	81	15	49
	Accumulated water	19	12	8	11
	Ice and snow	3	95	0	41
	Other	5	17	3	9
Visibility	1-50 meters	29	122	8	103
	50-100 meters	83	125	51	89
	100-200 meters	111	179	59	113
	200 meters or more	214	361	160	250
Traffic signal mode	No signal	139	257	60	137
	Marking line	111	109	0	1
	Signal lamp	43	38	27	29
	Sign	314	510	174	399
	Other	11	9	3	6

3.2. Comparison of factor's correlation performance

The correlation performance comparison results of the four analysis methods are shown in Fig. 2. According to the comparison results of factor's correlation performance in Figure 2, the overall factor correlation value of the proposed method is higher than that of the three methods in references, and the highest factor's correlation value of the proposed method can reach 0.87, while the highest correlation values of the methods in Deng and Zeng [5], Ge et al. [7] and Zhu et al. [20] are 0.62, 0.51, and 0.49, respectively, which are lower than that of the proposed method. Therefore, it fully shows that the proposed method has high factor's correlation performance.

3.3. Comparison of analysis efficiency

In order to further verify the effectiveness of the proposed method, the analysis efficiency is used as the comparison index for the experiment, and the comparison results of analysis efficiency of the four methods are shown in Figure 3. According to Fig. 3, the analysis efficiency of the proposed method is higher than that of the methods in Deng and Zeng [5], Ge et al. [7] and Zhu et al. [20] with the increasing number of experiments. When the number of experiments is 60, the analysis efficiency of the proposed method is 94%, while that of Deng and Zeng [5], Ge et al. [7] and Zhu et al. [20] is 78%, 63% and 82%, so the analysis efficiency of the proposed method is the highest.

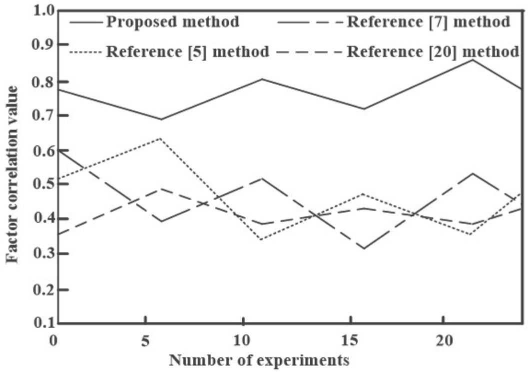

Fig. 2 - Comparison results of factor's correlation performance

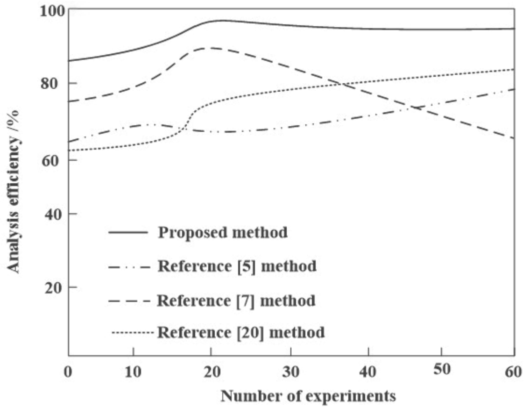

Fig. 3 - Comparison results of analysis efficiency