Acta Technica *62* (2017), No. 6A, 119–126

Cluster analysis of freeway traffic accidents based on external environmental factors

Bing Zhang[2, 3, 4, 6], Jianbiao Ning[5], Xiaoyong Zhou[4]

Abstract. Analyzing the traffic accidents during the period from 2007 to 2012 at the DaGuang freeway, the influence of human, vehicle, road and environment on the traffic accidents are considered in the light of the seriousness of the accidents and the different weight coefficient θ of death accidents and injured accidents, which are converted into the number of different accidents with the same severity. Combining the correlation of the external factors such as vehicle condition, traffic condition and weather regime_to analyze the relationship of different illegal driving behavior of motor vehicle drivers, and using the fuzzy clustering and systematical Clustering methods for analysis with regarding the external factors as statistical indicators and the different illegal driving behaviors as statistical samples. The results show that the clustering results of two methods are the same when $\lambda \in [0.9552,0.9736]$ and $k = 5$, the 12 kinds of common illegal driving behaviors are divided into five categories, which are proportional to the difference of direct economic losses, in line with the actual situation. And standardize the severity of the accidents which results from the various illegal driving behaviors under the combined action of external influencing factors and provide a strong theoretical basis for further traffic security management.

Key words. Vibration, trapezoidal plate, taper constants, thermal gradient, aspect ratio, non-homogeneity.

[1]Acknowledgment - It is very grateful to the anonymous referees for their insightful and constructive comments and suggestions that have led to an improved version of this paper. The work also get supports from Jiangxi Provincial Focus on research and Development Projects (20161BBG70080) and Jiangxi Provincial Department of Transportation Science and Technology Project (2015C0012 and 2014R0011).

[2]High Speed Rail and Regional Development Research Center, East China Jiao Tong University, Nanchang, Jiangxi, China

[3]School of Highway, Chang'an University, Xi'an, Shanxi, China

[4]Jiangxi Transportation Institute, Nanchang, Jiangxi, China

[5]School of transportation and logistics, East China Jiao Tong University, Nanchang, Jiangxi, China

[6]Corresponding Author, e-mail: Zhangbing191@sohu.com

1. Introduction

According to the World Health Organization (WHO), more than 1.17 million people die in road accidents around the world every year, and every year, over 10 million people are disabled or injured from road accidents. More than 6 million will die and as much as 60 million will be injured during the next 10 years in developing countries unless urgent action is taken [1]. The occurrence of freeway traffic accidents is related to the road environment, but it is mainly caused by human factors. 90.3% of traffic accidents are related to human factors (illegal driving behavior, bad driving condition, physical, psychological and other aspects) [2]. Traffic accidents as an important reason for the abnormal death and the accidental loss of property in peacetime are highly concerned. Many scholars have done this in-depth study. There exists the closely relevance between vehicle accidents and geometric design of road sections. For example, Miaou et al. [3] investigated statistical relationships between traffic crashes and traffic flows in Toronto, and Poch and Mannering [4] have investigated negative binomial models of crash counts in Washington. The factors on the severity of traffic accidents in freeway tunnels were analyzed by Ma Jinglin, Shao Chunfu, et al. [5] and nine variables were selected from the three aspects of time, environment and traffic dynamics to study their impact on the severity of traffic accidents. Li Jinlong, Sun Wanhua [6] consider the unique driving environment and the influence of comprehensive human, vehicle, road and environment, the paper analyzes the main causes of freeway traffic accidents by using analytic hierarchy process. Wen Bin, Cao Dongwei et al. [7] consider the impact of external weather factors and road conditions on freeway traffic accidents, obtains the relationship between rainy weather accident rate, rainy weather accident rate and pavement resistance (SFC).

On the study of freeway traffic accidents, most scholars focus on the traffic accident rate as a statistic, combined with driving behavior, road environment, vehicle status and so on to analyze the characteristics and causes of traffic accidents. In this paper, the author takes the number of traffic accidents as the statistic, taking into account the different severity of the accident, combining human factors such as human, car, road and environment and external environmental factors, using fuzzy clustering and system clustering, In recent years, traffic accidents occurred in the case of external factors to consider the case of driver driving behavior of the cluster analysis, to regulate the driver driving behavior and reduce the road traffic accident to provide a theoretical basis.

2. Mathematical analysis

2.1. Data processing

The data are derived from the traffic accident statistics for the period from 2007 to 2012 for a total of 6 years from the DaGuang freeway (2952 km + 350 m $\sim$ 3044 km + 450 m). Taking the motor vehicle driver driving behavior as a sample a_i, $i = 1, 2, \ldots, 12$. Three types of external factors such as vehicle type, road type and weather condition are taken as evaluation indexes b_k, $k = 1, 2, 3$. The number of traffic

accidents x_{ik} under the combined influence a_i and b_k is:

$$x_{ik} = \sum_{l=1,r=1}^{l=3,r=3} \theta_l n_r + N_i \eta_k \quad (i = 1, 2, \ldots, 12;\ k = 1, 2, 3)\,, \tag{1}$$

Where:

N_i — during the survey period, the number of traffic accidents caused by the i driving behavior;

n_r — Under the combined effect a_i and b_k: ($r = 1$: death toll; $r = 2$: number of seriously injured; $r = 3$: minor injuries).

2.2. Statistical analysis of raw data

By the above formula on the original data processing, in a variety of illegal driving behavior under the influence of external factors occurred in the same severity of the number of accidents as shown in Table 1, the illegal driving behavior caused by direct traffic losses caused by traffic accidents as shown in Table 2 As shown.

Table 1. Violation Driving Behavior Influencing Factors Incident

illegal driving behavior a_i	Evaluation index b_k			illegal driving behaviora_i	Evaluation indexb_k		
	Weather condition	Road line type	Vehicle type		Weather condition	Road line type	Vehicle type
Speeding a_1	405	225	138	Illegal overtaking a_7	15	9	14
Fatigue driving a_2	60	194	287	Illegal driving a_8	80	59	6
No-Deceleration through the construction of the road a_3	4	14	20	Overload a_9	54	42	16
Change lane illegally a_4	11	30	88	Driving without license a_{10}	24	4	16
Not keep a safe distance a_5	153	19	238	Drunk driving a_{11}	34	34	11
Driving an illegal vehicle a_6	45	88	201	Other illegal driving behaviors a_{12}	53	48	22

Table 2. Illegal driving behavior caused by direct traffic losses caused by traffic accidents (Unit: 10000 yuan)

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}
738	900	157	362	1010	1224	27	114	90	30	40	113

2.3. Data standardization

The data x'_{ik} is obtained by deviating the data in Table 1 (Equation 2), and data standardization [9] is standardized by mathematical statistics. For x'_{ik} application of the differential method (Equation 3), to obtain standardized data x''_{ik}, $x''_{ik} \in [0, 1]$, as shown in Table 3. The formula is as follows:

$$x'_{ik} = \frac{x_{ik} - \overline{x_k}}{\sqrt{S_{kk}}}, \tag{2}$$

Where: $\overline{x_k} = \frac{1}{n}\sum_{i=1}^{n} x_{ik}, S_{kk} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ik} - \overline{x_k})^2$

$$x''_{ik} = \frac{x'_{ik} - x'_{ij\,\min}}{x'_{ik\,\max} - x'_{ik\,\min}}. \tag{3}$$

Table 3. Standardized data

Illegal driving behavior a_i	Evaluation index b_k		
	Weather	Road line type	Vehicle Type
Speeding a_1	1.0000	1.0000	0.4698
Fatigue driving a_2	0.1396	0.8597	1.0000
No-Deceleration through the construction of the road a_3	0.0000	0.0452	0.0498
Change lane illegally a_4	0.0174	0.1176	0.2918
Not keep a safe distance a_5	0.3716	0.0679	0.8256
Driving an illegal vehicle a_6	0.1022	0.3801	0.6940
Illegal overtaking a_7	0.0274	0.0226	0.0285
Illegal driving a_8	0.1895	0.2488	0.0000
Overload a_9	0.1247	0.1719	0.0356
Driving without license a_{10}	0.0499	0.0000	0.0356
Drunk driving a_{11}	0.0748	0.1357	0.0178
Other illegal driving behaviors a_{12}	0.1222	0.1991	0.0569

2.4. Cluster analyses

In this paper, the system clustering method and fuzzy clustering method are used to cluster the driving behavior of motor vehicle drivers under the common influence of external environment.

(1) System clustering method

System clustering ideas: N samples, starting as a class, which is class N, and specifying the distance between the sample and the sample and the distance between the class and class. Then the nearest merged into a new class, calculate the distance

between the new class and other types; repeat two recent categories of mergers. Each time a class is reduced until all the samples are merged into a large class. The distance between the class and the class (similarity) can be calculated from the distance:

$$d_{ij}^2 = \sum_{k=1}^{3} (x_{ik} - x_{jk})^2 . \tag{4}$$

For the six kinds of system clustering methods, the squares of the square sums of the deviations are related to the number of samples of the two classes. The two categories tend to have a large distance and should not be merged. This is more in line with the clustering Practical requirements, so the deviation of the square sum method is a better system clustering method [10]. In this paper, the clustering results are generated by R-Project [11] using the squared sum method. The clustering results are shown in Figure 1.

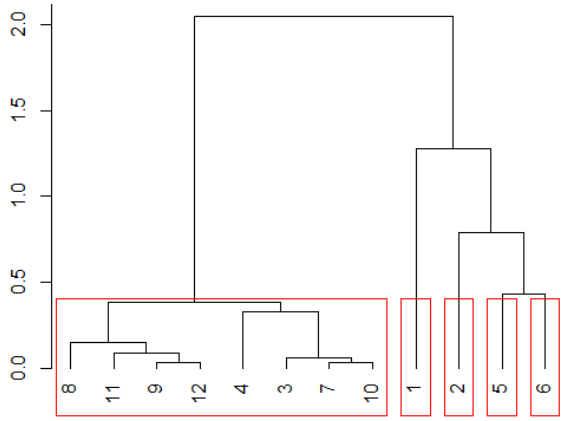

Fig. 1. The results of the squared sum method

(2) Fuzzy clustering method

Fuzzy clustering introduces the concept of fuzzy mathematics into clustering analysis. By constructing fuzzy equivalence relation, we construct the equivalent matrix, classify the objective things with fuzzy equivalence relation, and classify them with fuzziness.

The fuzzy clustering method is as follows:

1) Calculate the similarity statistics r_{ij}

For each object a_i to be classified, the same k dimension variable b_k is used, and the classification object is regarded as a sample. The similarity r_{ij} of the sample can be obtained by the related method such as angle cosine method and similarity coefficient method.

2) Construct the equivalent matrix R

The fuzzy matrix $R = (r_{ij})_{n\times k}$ only has reflexive and symmetry, and does not have transitivity. Equalization of the closure by means of the method, namely the

equivalent matrix on the domain.

$$R^2 = R \circ R = \bigvee_{k=1}^{m} (r_{ij} \bigwedge_{i,j=1}^{m} r_{ij}),$$

And so on, when $R^{2k} = R^k \circ R^k = R^k$, R^k [12] is the equivalent matrix.

3) Solve the Boolean matrix R_λ

In line with R^k select a different λ ($\lambda \in [0,1]$). And obtaining the dynamic cluster result by the different λ. This paper uses the matlab software [13] compiler to process the data. The fuzzy similarity matrix $R = (r_{ij})_{n\times k}$ is calculated by using the similarity coefficient method, and the equivalent matrix R^k and λ truncated matrix $R_\lambda = (\lambda r_{ij})_{n\times k}$ are obtained.

Among them:

$$\lambda r_{ij} = \begin{cases} 1 & (r_{ij} \geq \lambda), \\ 0 & (r_{ij} < \lambda). \end{cases} \tag{5}$$

Obtaining the equivalent matrix R^4 by sequential operation: $R^8 = R^4 \circ R^4 = R^4$.

In line with the equivalent matrix and selecting a different λ by formula (5) obtained the dynamic clustering results as shown Table 4. In the above clustering results, there are two cases about $\lambda \in [0, 0.8016]$ and $\lambda \in (0.9736, 1]$. The former ignores the difference of external environmental factors on traffic accidents caused by illegal driving behaviors, and the classification is too rough. The later emphasizes the difference of the influence of external environmental factors on traffic accidents caused by different illegal driving behaviors, and thus ignores the correlation among them, which results in no classification.

Table 4. Fuzzy clustering results

Ranges of λ	Clustering results	Number of clusters
[0, 0.8016]	{a1 a2 a3 a4 a5 a6 a7 a8 a9 a10 a11 a12}	One categorie
(0.8016, 0.9552]	{a1}, {a2 a3 a4 a5 a6 a7 a8 a9 a10 a11 a12}	Two categories
(0.9552, 0.9736]	{a1}, {a2}, {a5}, {a6}, {a3 a4 a7 a8 a9 a10 a11 a12}	Five categories
(0.9736, 1]	{a1}, {a2}, {a3}, {a4}, {a5}, {a6}, {a7}, {a8}, {a9}, {a10}, {a11}, {a12}	Twelve categories

3. Results analysis

When the above two kinds of clustering methods in clustering number $k = 5$, the clustering results are the same.

(1) When the DaGuang freeway traffic accidents resulted in the illegal driving behavior as a statistical sample and the external factors as a target variable, the accident can be divided into five categories. At the same time, according to the clustering results shown in Figure 1 and Table 2 can be seen: class spacing and

various types of illegal driving behavior caused by the direct economic loss of traffic accidents is proportional to the difference, in line with the actual situation.

(2) According to the clustering results and the economic losses caused by traffic accidents, it can be seen that the traffic accidents caused by the illegal driving behavior of the motor vehicle drivers under the influence of external environmental factors such as weather conditions, road type and vehicle condition can be seen The magnitude of the loss is $\{a_6\} > \{a_5\} > \{a_2\} > \{a_1\} > \{a_3, a_4, a_7, a_8, a_9, a_{10}, a_{11}, a_{12}\}$.

4. Conclusion

Through the cluster analysis method, it is well seen that the driving behavior of the traffic accident caused by the merger is related to the external influencing factors, and the driving behavior of the accident is clustered. The clustering result is the improvement of traffic safety the situation provides a clear direction: When the external environment cannot change, we can start from the people, educate the driver civilized driving, improve the driver's driving quality, for further traffic safety management work to provide a strong theoretical and empirical basis. For reduce the DaGuang freeway traffic accidents, we should focus on remedying driving illegal vehicles and not maintain safe driving distance and other illegal driving behaviors, followed by focus on remedying fatigue driving and speeding driving and other illegal driving behaviors, and finally should focus on prevention of illegal change lanes, illegal overtaking and other driving behaviors.

References

[1] M. Hosseinpour, A. S. Yahaya, S. M. Gradiri, etc: *Application of adaptive neuro-fuzzy inference system for road accident prediction.* KSCE Journal of Civil Engineering *17* (2013), 1761–1772.

[2] S. G. Klauter, T. A. Dingus, V. L. Neale, et al: *Comparing real-world behaviors of drivers with high versus low rates of crashes and near crashes.* Washington, DC: National Highway Traffic Safety Administration, 2009, 25–28.

[3] S. P. Miaou, D. Lord: *Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes methods.* Transportation Research Record *184* (2003), 31–40.

[4] M. Poch, F. L. Mannering: *Negative binomial analysis of intersection-accident frequency.* Journal of Transportation Engineering *122* (1996), 105–113.

[5] Z. L. Ma, C. F. Shao, X. Li: *Analysis of influence factors on severity for traffic accidents of expressway tunnel.* Journal of Beijing JiaoTong University *33* (2009), 52–55.

[6] J. L. Li, W. H. Sun: *Cause analysis of traffic accidents on express highway and study on their countermeasures.* China Safety Science Journal *15* (2005), 59–62.

[7] B. Wen, D. W. Cao: *Statistical analysis of traffic accident and skid-resistance of expressway pavement.* Journal of Highway and Transportation Research and Development *23* (2006), 72–75.

[8] C. S. Zhang, R. G. Ma: *Identifying method of traffic accident black spots on mountain freeway.* Journal of Chang'an University (Natural Science Edition) *30* (2010), 76–80.

[9] Z. Sheng, S. Q. Xie, C. Y. Pan: *Probability and Mathematical Statistics (Fourth Edition).* Beijing: Higher Education Press, 2008, 10–16.

[10] B. Tian: *Research on system clustering and its application.* YinShan Academic Journal *28* (2014), 11–16.

[11] Q. X. Li: *R language combat: programming basis, statistical analysis and data mining Collection.* Beijing: Electronic Industry Press, 2016, 210–216.

[12] H. C. Sung, B. P. Jin, Y. H. Joo, K C Lin: *Robust digital implementation of fuzzy control for uncertain systems and its application to active magnetic bearing system.* International Journal of Control Automation & Systems *10* (2012), 603–612.

[13] M. Chen, C. Y. Zheng, Z. Zhang: *Matlab function and examples of quick manual.* Beijing: People's Posts and Telecommunications Press, 2014, 125–156.

Received October 19, 2017