Research Article

The Model of Severity Prediction of Traffic Crash on the Curve

Jian-feng Xi,^{1,2} Hai-zhu Liu,² Wei Cheng,³ Zhong-hao Zhao,² and Tong-qiang Ding²

¹ State Key Laboratory of Automotive Simulation and Control, Jilin University, Changchun 130022, China

² College of Transportation, Jilin University, Changchun 130022, China

³ Faculty of Transportation Engineering, Kunming University of Science and Technology, Kunming 650224, China

Correspondence should be addressed to Wei Cheng; chengwei.ding@163.com

Received 5 September 2013; Revised 20 November 2013; Accepted 29 November 2013; Published 9 January 2014

Academic Editor: Wuhong Wang

Copyright © 2014 Jian-feng Xi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the study of traffic crashes on curved road segments as the focus of research, a logistic regression based curve road crash severity prediction model was established based on a sample crash database of 20000 entries collected from 4 regions of China and 15 evaluation indicators involving driver, driving environment, and traffic environment factors. Maximum Likelihood Estimation and step-back technique were deployed for data analysis, the conclusion of which is that the three main contributory factors on curve road crash severity are weather, roadside protection facility, and pavement structure. Hosmer and Lemeshow tests were used to verify the reliability of the model, and the model variables were discussed to a certain degree as well.

1. Introduction

As a major component of road geometric design, curved road segments, due to their alignment characteristics, are most prone to traffic crashes among all road geometric elements. In 2010, crashes on curved segments accounted for 10.5% of the total number of traffic crashes in China. Correspondingly, the number of deaths accounted for 12.89% of the total number of deaths, as shown in Figure 1.

Many researches have been conducted with regard to the characteristics and causes of crash on curved segments, ranging from those based on macroeconomic statistics, where crash rates of straight and curved segments are compared [1, 2], and those based on vehicle dynamics, where the focuses are on improving vehicle's safety and through performance on curves [3, 4], to those that are based on the study of human factors, where the studies of traffic crash have been heavily focused on drivers and driver behavior [5]. Vehicle speed and geometric design are also topics of study when curve road crashes are concerned [6, 7]. Regarding crash severity, certain mathematical models have been established, such as the Bayesian Method [8], the Ordered Probit Model [9], and the Neuronal Network Approach [10]. However, among the aforementioned studies and researches, few had aimed

specifically at investigating crash severity on curved road segments.

Analysis and prediction models are based on crash data. In the past, data used by traffic crash studies on curve roads was mostly gathered from specific segments. Crash samples and crash-related factors had not been duly considered either. For these reasons, crash analysis would often fail to explain the real causes of crash, and the models were often questionable in terms of reliability and adaptability.

In this study, the crash data consists of 500 randomly chosen samples, according to the data gathering and management standards of China's "National Road Traffic Crash Information System," from a crash database of 20000 valid crash data entries that covers 4 regions of China from 2004 to 2008. The database comprises 60 items of crash information, which is comprehensive enough to recreate the process of a crash and provide an important basis for the analysis of causes. In terms of analysis method, the crash data is the foundation of analysis of the cause of crash and the structure and form of the crash determines the model for crash causation analysis. Since the attributed causes of each crash cannot be isolated from a single point of interpretation [11] and various internal and external factors are affecting drivers' reaction to road traffic safety, this paper made a

FIGURE 1: National crash loss on the curve in 2010.

comprehensive analysis of the severity of crash on curves from three aspects that include driver, vehicle type, and the road environment.

In recent years, due to its innate suitability, the logistic regression model has been widely used in practical problems, such as Banking [12], Genomics [13], and Psychopathology [14]. Many researchers used the logistic regression model to analyze the safety of roads, such as crashes of single-vehicle motorcycles [15], and crash prevention [16, 17]. Descriptive factors are defined as natural numbers that start from 0 and can be transformed into discrete variables. By this transformation, traffic crash data can meet the requirements of logistic regression model; therefore, the severity prediction model of traffic crash on curved segments may be built based on these data. The model [18] can overcome the deficiencies of analysis method and linear regression analysis of the traditional Mantel-Haenszel model, for it can contain multiple influencing factors that include an analysis of both discrete and continuous variables. The model can effectively analyze mixed influences and interactions from external variables and therefore also provides a methodological basis for the quantitative description relationship between multiple influencing factors and the prediction of severity on curves.

2. Establishing the Logistic Regression Model

2.1. Binary Logistic Regression Probability Formula. The severity of traffic crash on curved segments is regarded as dependent variable y , when the i th crash has any number of deaths, $y_i = 1$; otherwise, $y_i = 0$.

Assuming there are p influence factors that are related to dependent variable y and are denoted $X = (x_1, x_2, \dots, x_p)$, the probability of fatal crash under the impact of influencing factors is

$$P = \left\{ y_i = \frac{1}{X} \right\} = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}, \quad (1)$$

where x_i ($i = 1, 2, \dots, p$) are the influencing factors of crash severity; it can be a continuous variable, categorical variables, or dummy variables; β_i ($i = 1, 2, \dots, p$) are the regression coefficients.

2.2. Binary Logistic Regression Model Parameter Estimation. Considering the severity of the crash to be a dichotomous dependent variable, $x_1, \dots, x_p$ are its corresponding independent variables. Let $x_i = (1, x_{i1}, \dots, x_{ip})'$, and $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$ be its corresponding variable vector, $i = 1, \dots, n$. Thus, a logistic model may be established as [15]

$$\text{logit}(y) \triangleq \ln \left(\frac{p_i}{1 - p_i} \right) = x_i' \beta, \quad (2)$$

in which $p_i = P(y_i = 1/x_i) = e^{x_i' \beta} / (1 + e^{x_i' \beta})$.

The model used Maximum Likelihood Estimation method for variable estimation and the likelihood function of variable $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$ can easily be derived from the binary logistic regression model, which is

$$l(\beta) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i}, \quad (3)$$

where $y_i = 0$ or 1 . Therefore, the logarithmic likelihood function is

$$L(\beta) = \ln l(\beta) = \sum_{i=1}^n [y_i x_i' \beta - \ln(1 + x_i' \beta)]. \quad (4)$$

For $(\partial L(\beta))/\partial \beta = 0$, it resulted in

$$\sum_{i=1}^n [x_i (y_i - p_i)] = 0. \quad (5)$$

In the logistic regression model, the Newton-Raphson method is used. Define $s_i = y_i - p_i$, $v_i = p_i(1 - p_i)$, and let $S = (s_1, \dots, s_n)'$, $V = \text{diag}(v_i)$; then

$$\begin{aligned} \frac{\partial L(\beta)}{\partial \beta} &= \sum_{i=1}^n x_i (y_i - p_i) = X' S, \\ -\frac{\partial^2 L(\beta)}{\partial \beta \partial \beta'} &= \frac{\partial}{\partial \beta'} \left(\sum_{i=1}^n x_i (y_i - p_i) \right) \\ &= -\frac{\partial}{\partial \beta'} \left(\sum_{i=1}^n x_i p_i \right) \\ &= \sum_{i=1}^n p_i (1 - p_i) x_i x_i' \\ &= X' V X, \end{aligned} \quad (6)$$

in which

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}. \quad (7)$$

If the number of iterations is k , then the maximum likelihood estimate of variable β becomes

$$\begin{aligned}\hat{\beta}^k &= \hat{\beta}^{k-1} - \left[\left(\frac{\partial^2 L(\beta)}{\partial \beta \partial \beta'} - i \frac{\partial L(\beta)}{\partial \beta} \right) \right]_{\beta=\hat{\beta}^{k-1}} \\ &= \hat{\beta}^{k-1} + \left[(X'VX)^{-1} (X'S) \right]_{\beta=\hat{\beta}^{k-1}}.\end{aligned}\quad (8)$$

3. Selection of Crash Impact Factor

Through a detailed analysis on the structure of the crash database, it is found that the database is a superdimensional structure with radial, multidimensional, and multilevel characteristics. Each crash record contains multiple data attributes and each value reflects a traffic crash's characteristics in one aspect. In addition, the setting of data attributes is based on five factors, which are the information of the crash, personnel information, vehicle information, road information, and environmental information. Therefore, logically the data attributes and five factors of people, vehicles, roads, climate environment and crash basic information formed a two-level formation, in which the five factors of people, vehicles, roads, environment, and crash basic information are in the upper layer, and the data attributes are in the lower layer. Also, different attributes and the attribute value formed a set complying with their specific logical relations and within each attribute there existed a specific hierarchy.

500 crash samples, randomly selected from a crash information database of 20000 entries, have been analyzed. The causes of crash are mostly related to bad subjective judgment and human errors, vehicle performance issues, change of external environment, and change of road conditions. However, vehicle performance is generally not considered in crash analysis and is neglected in this study.

Therefore, from three-level analysis system of drivers, driving environment, and road environment, the model selects 15 evaluation items as independent variables. The evaluation of independent variables is shown in Table 1.

4. Prediction Model of Traffic Crash Severity on Curved Segments

The model used stepwise regression method to analyze the independent variables and step-back technique to obtain the results. In step 1, 15 independent variables were all put into the model and the variables based on the probability of the likelihood ratio for the test were assumed. In step 13, the weather X_6 , roadside protective facility type X_{12} , the road pavement X_{14} , and constant were selected. Results are shown in Table 2.

Taking a significance level of 0.05 and using the reverse stepwise method and 13 times of screening for selection, the model obtained the correlation of crash severity on curves with weather, road-side protective facility type, and road pavement structure. From Hosmer and Lemeshow tests, the results in Table 3 showed that the significance level of 0.674 is greater than 0.05 and thus proved that the original assumption is valid. Also, the chi-square value from 7.856

of the first screening reduced to 0.79 in the 13th screening, which proved that the model is correct.

The probability of crash with death from the logistic regression is expressed by

$$\begin{aligned}P &= \left(y_i = \frac{1}{X} \right) \\ &= \frac{e^{-1.634+2.258X_6-1.321X_{12}-1.46X_{14}}}{1 + e^{-1.634+2.258X_6-1.321X_{12}-1.46X_{14}}}.\end{aligned}\quad (9)$$

5. Conclusion

(1) Not all of the 15 impact factors are selected to put into the curve road crash severity prediction model. Some were excluded partly due to weak correlations, but it does not mean that they have little impact on the severity of a crash. For example, visibility and road surface conditions are closely related to weather conditions and their effects may be indirectly reflected in the final model, if such factors are to be further considered.

(2) During bad weather conditions, road surface friction coefficient decreases and visibility of road condition reduced, and vehicles driving on curved segments will be prone to cross over the median, causing side scraping, rollover, or rear-end crashes. During rainy or foggy weathers, water film will form on the surface of the road reducing tire friction. Therefore, weather is one of the main factors that affect crash severity on curved segments, and the severity will be even greater especially under ice and snow conditions.

(3) Highway safety design often focuses on road alignment and profile, but lacks due considerations toward road-side facilities. However, in the real world, road-side environment in general is closely related to traffic crashes. On curve roads, off-road crashes frequently occur and it is mainly because of improper speed control and road-side environment, especially on mountainous curves where one side of the road is either deep grooves or cliffs. If curved segments have roadside protection facilities, it can greatly reduce the severity of the crash. Hence, better design and installment of road-side protection facilities should be one of the priorities for preventing severe crashes from happening.

(4) Compared to gravel pavement, asphalt roads generally have higher crash rates. Also, compared to the cement concrete pavement, asphalt roads have smoother surface and greater involuntary horizontal movement from vehicles. Therefore, drivers usually feel more comfortable driving on an asphalt-paved road surface, which often leads to high speed driving and speed related crashes. Asphalt pavement is more sensitive to the temperature; its structural strength decreases at high temperature and cracks at low temperature. However, drivers would often neglect the effect of temperature changes on the condition of road surface and their effects on safe driving, which is another very important factor in causing severe crashes.

(5) The existing road traffic crash database of China was originally designed to determine the responsibility of crash for legal purposes. With its 60 data items or attributes,

TABLE 1: Evaluation of independent variables.

Evaluation of the category	Variable symbol	Variable content	Variable assignment
Driver	X_1	Gender	Female = 0, male = 1
	X_2	Age classification	Age from 16 to 25 = 0, 26 to 35 = 1, 36 to 45 = 2, above 46 = 3
	X_3	Household register	Agriculture household = 0, nonagricultural household = 1
	X_4	Driving experience classification	1–5 years = 0, 6–10 years = 1, 11–15 years = 2, 16–20 years = 3, above 20 years = 4
	X_5	The accident responsibility	Minor responsibility = 0, equal responsibility = 1, main responsibility = 2, all responsibility = 3
Driving environment	X_6	Weather	Sunny = 0, not sunny (rain, snow, fog, cloudy, wind) = 1
	X_7	Terrain	Plain = 0, others = 1
	X_8	Visibility	Above 200 m = 0, 100–200 m = 1, 50–100 m = 2, below 50 = 3
	X_9	Road surface condition	Dry = 0, not dry (rain, ice) = 1
	X_{10}	Lighting condition	Day = 0, night with light = 1, night without light = 2
Road environment	X_{11}	Traffic signal type	With signal control = 0, without signal control = 1
	X_{12}	Roadside facilities protection type	With roadside protection = 0, without roadside protection = 1
	X_{13}	Road physical isolation	With isolated = 0, without isolated = 1
	X_{14}	Pavement structure	Asphalt pavement = 0, not asphalt pavement = 1
	X_{15}	Road conditions	Road surface in good condition = 0, pavement damage = 1

TABLE 2: Variable estimation.

Variable	B	S.E.	Wals	df	Sig.	Exp (B)
Weather	2.258	1.443	2.448	1	0.118	9.56
Roadside protective equipment	−1.321	0.853	2.398	1	0.122	0.267
Road pavement	1.46	0.531	7.546	1	0.006	4.304
Constant	−1.634	0.403	16.409	1	0.000	0.195

TABLE 3: Hosmer and Lemeshow test.

Step	Chi-square	Freedom	Significance level
1	7.856	8	0.448
13	0.79	2	0.674

although very comprehensive, it still needs future improvement in order to better describe the whole process of an actual traffic crash and to be used as the basis for crash prediction and prevention studies.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This project was supported by the NSFC (Grant no. 50808093).

References

- [1] T. Miller, D. Lestina, M. Galbraith, T. Schlax, P. Mabery, and R. Deering, "United States passenger-vehicle crashes by crash geometry: direct costs and other losses," *Accident Analysis and Prevention*, vol. 29, no. 3, pp. 343–352, 1997.
- [2] C. Zhang and J. N. Ivan, "Effects of geometric characteristics on head-on crash incidence on two-lane roads in connecticut," *Journal of the Transportation Research Board*, vol. 1908, pp. 159–164, 2006.
- [3] S. K. Rao, "Unit curve for design of highway transitions," *Journal of Transportation Engineering*, vol. 121, no. 2, pp. 169–175, 1995.
- [4] J. B. Beniley and W. E. Gallagher, "Review of highway curve design," *Journal Highway*, vol. 18, no. 11, pp. 7–16, 2005.
- [5] S. Schick, "Accident related factors," Trace Coordinator V3, 2009.
- [6] R. Yuan Yuan, *Research on Driving Dangerous Area and Driving Behavior Model in Road Curved Section*, Traffic Environment and Safety Technology, Jilin University, Changchun, China, 2011.

- [7] W. Chaoshen, *Study on Road Traffic Safety Analysis and Countermeasures in Curve*, Transportation Planning and Management, Chang'an University, Xi'an, China, 2010.
- [8] J. Ma, K. M. Kockelman, and P. Damien, "A multivariate poisson-lognormal regression model for prediction of crash counts by severity, using bayesian methods," *Accident Analysis and Prevention*, vol. 40, no. 3, pp. 964–975, 2008.
- [9] B. Mediouny, Z. Anis, and H. S. Habib, "Motorway traffic crash prediction based on the network approach: application to the ringway of Paris," *Evaluation at Optimization System in the Production Services*, vol. 1, pp. 1–8, 2011.
- [10] S. M. Rifaat and H. C. Chin, "Accident severity analysis using ordered probit model," *Journal of Advanced Transportation*, vol. 41, no. 1, pp. 91–114, 2007.
- [11] R. Howard, *Statistical Modeling and Analysis of Injury Severity Sustained by Occupants of Passenger Vehicle Involved in Crashes with Large Trucks*, University of Nevada, Reno, Nev, USA, 2010.
- [12] S. Y. Sohn and H. S. Kim, "Random effects logistic regression model for default prediction of technology credit guarantee fund," *European Journal of Operational Research*, vol. 183, no. 1, pp. 472–478, 2007.
- [13] C. M. Zhang, H. D. Fu, Y. Jiang, and T. Yu, "High-dimensional pseudo-logistic regression and classification with applications to gene expression data," *Computational Statistics and Data Analysis*, vol. 52, no. 1, pp. 452–470, 2007.
- [14] C. R. García-Alonso, J. Guardiola, and C. Hervás-Martínez, "Logistic evolutionary product-unit neural networks: innovation capacity of poor guatemalan households," *European Journal of Operational Research*, vol. 195, no. 2, pp. 543–551, 2009.
- [15] V. Shankar and F. Mannering, "An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity," *Journal of Safety Research*, vol. 27, no. 3, pp. 183–194, 1996.
- [16] M. A. Abdel-Aty and H. T. Abdelwahab, "Predicting injury severity levels in traffic crashes: a modeling comparison," *Journal of Transportation Engineering*, vol. 130, no. 2, pp. 204–210, 2004.
- [17] R. Harb, E. Radwan, X. Yan, A. Pande, and M. Abdel-Aty, "Freeway work-zone crash analysis and risk identification using multiple and conditional logistic regression," *Journal of Transportation Engineering*, vol. 134, no. 5, pp. 203–214, 2008.
- [18] A. S. Al-Ghamdi, "Using logistic regression to estimate the influence of accident factors on accident severity," *Accident Analysis and Prevention*, vol. 34, no. 6, pp. 729–741, 2002.