Modeling of single-vehicle crash severity at intersections based on full Bayesian model

Zhenggan Cai[1]*, Xin Li[1], Long Chen[1], Xiaoyan Wu[1], Xu Yao[2]

[1]School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo, Shandong, 255000, China

[2]School of Management, Changchun University, Changchun, Jilin, 130000, China

*corresponding author: Dr. Zhenggan Cai; E-mail: zhenggancai@163.com

ABSTRACT

In order to explore the factors that have a significant impact on the severity of single-vehicle crashes at intersections, we take 2940 crashes at intersections in Zibo City from 2017 to 2019 as samples. And then, we use the sample data to establish a Bayesian logit model. The modeling results show that the probability of fatal crash of male drivers is 0.53 times of that of female drivers. The probability of fatal crash of unused seat belts is 2.41 times of that of using seat belts. The probability of fatal crash of drunk driving is 3.75 times of that of non-drunk driving. The probability of fatal crash of elderly drivers is 1.47 times of that of middle-aged drivers. The probability of fatal crash in the night is 1.46 times of that in the morning peak.

Keywords: traffic safety; single-vehicle crash; intersections; full Bayesian model

INTRODUCTION

According to the data released by the World Health Organization in 2018, at least 1.5 million people die in traffic crash every year. And road traffic crashes have become the eighth largest cause of death in the world. Among them, Single-vehicle (SV) crashes account for 59% of the total deaths. SV crashes have become a major threat to road traffic safety. In all SV crashes, the crashes at intersections account for more than 52% of the total crashes. In addition, the probability of driver death caused by SV crashes at intersection is higher than that at road section. Therefore, it is necessary to conduct a systematic study on the SV crashes at intersections.

Domestic and foreign researches on single vehicle accidents mainly focus on regional factors or macro level accidents [1-4]. For example, Cai, et al. [1] in the research process of SV crashes, firstly, the mixed logit model was established by using the SV crashes data of Zibo City. It was found that the rainfall weather would significantly increase the severity of driver injury, and the snowfall weather would significantly reduce the severity of the accident. Moreover, the probability of serious injury caused by single vehicle accident in rural areas was significantly higher than that in cities. Then, Cai, et al. [2] modeled the SV crasehs in rural areas, and found that the probability of serious injury to drivers in winter was low. Based on the existing research, Hu, et al. [3] established a multi-layer ordered logit model for SV crashes in Zibo City, and studied the factors affecting the severity of driver injury, and reached the same conclusion. Huang, et al. [5] Based on the Bayesian logit model, the influencing factors of the severity of SV crashes were analyzed, and it was found that the probability of fatal crash in un-sunny weather was 1.127 times of that in sunny weather.

In the research process of traffic safety, scholars have established a large number of discrete choice models, among which logit model has been widely used because of its simple structure and mature theory [5]. For example, Sivasankaran and Balasubramanian [5] used binary logit model to study SV crashes, and identified the impact of different factors on the severity of crashes. Ahmed et al. [6] took truck collisions in rural areas as samples to establish Bayesian binary logit model. The modeling results show that the Bayesian logit model can reasonably predict the severity of the crash. Therefore, this paper also uses logit model to analyze the influence factors of SV crash severity at intersections.

To sum up, scholars have done little research on SV crashes at intersections. On the basis of previous studies, this paper establishes a binary logit model with the severity of driver injury as the dependent variable, and discusses the factors that have a significant impact on SV crashes at intersections.

DATA

Based on the traffic crash data of Zibo City, this paper extracts 3149 SV crashes at intersections from 2017 to 2019. After eliminating the samples with incomplete records, a total of 2940 SV crashes were used for subsequent modeling. When dividing the severity of the accident, the injury status of the driver is taken as the severity level of the crash. If the driver died within seven days after the crash, the crash will be regarded as a fatal crash. If the driver is not dead, the crash will be regarded as a non-fatal crash. Therefore, the dependent variables of the model are death and non-death. The independent variables of the model include 10 factors such as driver's gender, age, accident time and season. The specific variable statistics and coding are shown in TABLE 1.

TABLE 1: Classification and coding statistics of variables

Variables	Code
Severity (y)	0=Non-death; 1=death
Driver gender(x_1)	0=Female; 1=male
Seat belt/helmet (x_2)	0=Used; 1=Not used
Fatigue driving (x_3)	0=No; 1=yes
Drunk driving (x_4)	0=No; 1=yes
Speeding (x_5)	0=No; 1=yes
Driver age (x_6)	0=25-59; 1=<25
	2=>59
Crash type (x_{11})	0=Collison with non-fixed object;
	1=Collison with fixed object
Weather (x_{13})	0=Clear weather
	1=Non clear weather
Time (x_{15})	0=Morning peak (7:00-10:00)
	1=Noon (10:00-17:00)
	2=Evening peak (17:00-20:00)
	3=Night (20:00-7:00)

METHODOLOGY

The standard Logit model is essentially a generalized linear model (GLM), which treats model parameters as fixed unknown constants and uses data to estimate the best values of the parameters. In the Bayesian method, parameters are treated as random variables, and data are fitted to the parameters to evaluate their distribution characteristics. In this study, a uniform prior, which obeys a normal distribution, is used to estimate the initial parameters of the model using the maximum likelihood method. In addition, Bayesian method is different from classical statistics. Bayesian derives the posterior probability distribution of parameters, constructs credible intervals, and interprets the parameters with probability. It also gives an accurate measurement of the posterior distribution of the parameters, which overcomes the problem of overestimating the accuracy of the maximum likelihood method due to ignoring the uncertainty of parameter estimation. It effectively avoids the overfitting problem of the model when the number of samples is limited and the number of variables is large. Bayesian model is given as:

$$P(Y|X) \propto P(X|Y)P(X) \qquad (1)$$

Where Y denote the binary dependent variable of the model, $P(X)$ is the prior probability; $P(Y|X)$ is the posterior probability; $P(X|Y)$ is the likelihood function.

Bayesian Logit Model constructs a fitting model between the binary dependent variable Y and the independent variable X such as weather, time, season and visibility. Formula (2) for Logit regression model:

$$Y = \mathrm{logit}(p) = \beta X + \beta_0 \qquad (2)$$

Where β_0 is the model intercept; β represents the regression coefficient of the model; X is the set of model independent variables; P represent the probability of Y=1. The Bayesian Logit regression model is to analyze and describe the probability of death and its influencing factors in traffic accidents with mathematical expression methods. As mentioned above, the judgment of accident casualties shall be based on the most serious injury. If the driver is killed in the collision, the accident is considered to be a fatal crash and coded as 1, otherwise it is classified as a non-fatal crash and coded as 0.

RESULT ANALYSIS

The Bayesian logit model established in this paper is implemented by stata16.0 software, and the significance level is 0.1 to test the significant influence of independent variables on dependent variables. The results show that driver's gender, unused seat belt, drunk driving, elderly driver, collision fixture and accident time have significant effects on the severity of driver injury. The specific model estimation results are shown in Table 2. As can be seen from table 2, the ROC value of the model is 0.793. Ahmed et al. [7] showed that when the ROC value was greater than 0.785, the fitting effect of the model was very good. Therefore, the goodness of fit of the model is reasonable

TABLE 2: Parameter estimation results of Bayesian model

Variables	Attribute	Reference	Coefficient	S.E.	OR
Driver gender	Male	Female	-1.057	0.051	0.53
Seat belt/helmet	Not used	Used	1.761	0.070	2.41
Drunk driving	Yes	No	2.075	0.183	3.75
Driver age	>59	25-59	0.871	0.081	1.47
Crash type	Fixed object	Non-fixed object	1.840	0.021	2.17
Time	Night	Morning peak	0.974	0.107	1.46
Intercept	-	-	-2.720	0.079	0.19
ROC			0.793		
DIC			1435.79		
MCMC			20000		
Burn-in			5000		

It can be seen from TABLE 2 that compared with female drivers, male drivers are 0.53 times more likely to have a fatal collision than female drivers. This shows that female drivers are more vulnerable to serious injury in collision. This finding can be proved in the existing literature [4].

The probability of death and injury caused by not using seat belt is 2.41 times of using seat belt. This conclusion is reasonable. As an important part of passive safety protection measures, seat belts can effectively protect the lives of drivers in case of collision. This conclusion has been widely accepted [3].

The probability of death collision of drunk driving is 3.75 times that of non-drunk driving, which is consistent with Cai et al. [3]. It is obvious that alcohol will affect the driver's reaction sensitivity and driving stability, and is more vulnerable to serious injury in case of emergency.

Older drivers (25-59) are 1.47 times more likely to have a fatal crash than middle-aged drivers (25-59), which is consistent with previous studies [1].

The probability of driver's death caused by collision with fixed object is 1.46 times of that caused by collision with non-fixed object. This conclusion is reasonable. Because it is easy to cause vehicle rollover when colliding with fixed objects, and rollover is more likely to cause serious injury to the driver [6].

The probability of death collision at night is 1.46 times that of the morning peak, which is consistent with the existing literature [7].

CONCLUSION

In order to explore the factors that have a significant impact on the severity of single vehicle accidents at intersections, we use the collision data of Zibo City in 2017-2019 to build Bayesian logit model. The sample number is 2940 traffic collisions. The results show that non use of seat belts, drunk driving, elderly drivers, collision fixtures and night time are significantly positively correlated with the severity of the collision. However, there was a significant negative correlation between male drivers and crash severity.

In the future, clustering technology will be used to analyze the single vehicle collisions at intersections to mine the heterogeneity of data.

REFERENCE

[1] Z. G. Cai, S. S. Hu, R. R. Wang, et al., "Severity analysis for single-vehicle crashes based on mixed logit model," *Journal of Scientific and Engineering Research*, vol. 7. no. 6, pp. 71-78, 2020.

[2] Z. G. Cai, S. S. Hu, R. R. Wang, et al., "Analysis of driver-injury severity in rural single-vehicle crashes based on multinomial logit model," *Journal of Scientific and Engineering Research*, vol. 7, no. 5, pp. 206-211, 2020.

[3] S. S. Hu, Z. G. Cai, Y. Q. Zhou, et al., "Analysis of factors affecting the severity of single-vehicle accidents in Chinese cities based on multi-layer order logit model," *Journal of Scientific and Engineering Research*, vol. 7 no. 6, pp. 71-78, 2020.

[4] Z. G. Cai, S. S. Hu, Z. Zhen, et al., "Traffic mode selection prediction model based on standard multinomial logit model." *Journal of Scientific and Engineering Research*, vol. 7 no. 6 pp. 155-159, 2020.

[5] X. H. Huang, X. Yao, Y. Mei, et al., "Investigation of driver injury severities in single-vehicle crashes using Bayesian binary logit model," *RAS Engineering & Technology*, vol. 1, no. 1, pp. 1-4, 2020.

[6] Sivasankaran, S. K., Balasubramanian, V. (2020). Exploring the severity of bicycle–vehicle crashes using latent class clustering approach in India. *Journal of Safety Research*, 72, 127-138.

[7] M. M. Ahmed, R. Franke, K. Ksaibati, et al., "Effects of truck traffic on crash injury severity on rural highways in Wyoming using Bayesian binary logit models," *Accident Analysis & Prevention*, vol. 117, pp. 106-113, 2018.