IMPACT OF REAL-TIME TRAFFIC CHARACTERISTICS ON FREEWAY CRASH SEVERITY

Meemee Chu, College of Engineering, California State Polytechnic University-Pomona, 3801 W.Temple Ave, Pomona, 91768, 626-731-5080, mtchu@cpp.edu
Wen Cheng, College of Engineering, California State Polytechnic University- Pomona, 3801 W.Temple Ave, Pomona, 91768, 909-869-2957,wcheng@cpp.edu

ABSTRACT

This research study focuses on predicting crash injury types due to real time traffic characteristics on freeway. The vehicle speed and traffic volume on the roads are the main contributing factors considered in the research in predicting the crash occurrence. The crash data collected at different time intervals are applied to validate the Generalized Linear Models (GLM) with the intent to improve the accuracy of the crash severity prediction. The research results are explained in details and shown to identify the relationship between speed, volume and crash severity. The outcomes for this study are expected to help mitigate the crash severity on freeway.

Introduction

Traffic safety has been playing an important role in today's society because people use transportation almost every day of their lives and need to be safe whenever they use vehicles. Therefore, many research studies have been conducted to help enhance the traffic safety. There are many research papers related to traffic issues. Many researchers focus on different aspects of roadway safety. Safety on the road is important as vehicles collisions can lead from light to severe injuries and even fatalities. There are numerous crash related reports published by different sources. According to the year 2013 data of National Highway Traffic Safety Administration (NHTSA) [4], 32,719 people were killed in vehicle crashes in the United States. In addition, Association for Safe International Road Travel (ASIRT) [2] reported that 1.24 million people around the world every year lose their lives and almost 50 million are injured because of roadway crashes. Therefore, creating effective methods to improve roadway safety is required in society. Researchers and engineers have been trying to identify the best approaches or ways to prevent crash occurrence. There are numerous factors that influence on vehicle crashes such as drivers' behaviors, weather conditions, roadway geometry, vehicle characteristics, and traffic characteristics, and many more. Many researchers have high interest in whether there is a relationship between those factors in the case of crashes and how these impacts of these factors can affect the roadway users.

Some researchers have created crash prediction models using the facts of the drivers' behaviors and driving speeds. In addition, the traffic volume, which is the number of vehicles on the roadway per unit of time, is also used as one of the traffic characteristics in analyzing the crashes. Roadway geometries such as horizontal curve, vertical curve, sight distances, and channelization are other influential factors on crash occurrence.

Different research papers use different crash prediction models in predicting the crash frequency or rates. One type of the models is dependent on the real time traffic characteristics, which can date back in 1990s (Preston 1996) [5] Even though the researchers have made progress over the years using the real-time crash models, the different research often leads to inconsistent results. For example, Zheng, Ahn and Monsere (2010) [6] reported that the impact of traffic oscillations are related to crash occurrences on freeway. However, Abdel-Aty et al (2004) [1] concluded that the significant factors affecting the crashes are roadway geometry, time of the day, speed variations and occupancy. Even though most of the previous studies did not report the performances of their prediction models, Abdel-Aty el al [1] indicate that their reported models have high prediction errors such as 38.8% error.). Due to the prediction errors and contradictory conclusions demonstrated in the past and current crash models using real time traffic data, it is imperative that more research of the kind is needed to verify previous research and further understand the relationship between crash and real-time traffic characteristics. To this end, the main objective of this research is to identify the relationship between real-time traffic characteristics and crash injury levels using the data collected from California freeways.

Data Collection

In this research, there are traffic data websites which keep track on the traffic conditions from loop detector and some other sensing device on the roads. The data used for this project were collected from the Caltrans Performance Measurement Systems (PEMS) database website and the Statewide Integrated Traffic Record System (SWITRS) websites. PEMS is maintained by Caltrans and it gives the traffic data such as speed, density and volume of crash time. SWITRS is controlled by California Highway Patrol and it can provide the data for crash time, crash severity, crash location, driver age, gender, and crash type. The collected data locations are taken from the different roadway segments of Interstate freeway I-5 segments in California and these data are data from the year 2010. Since one of the main purposes for this research is to create the crash prediction model in order to be able to identify the influential factors causing the crashes. Table 1 shows the snapshot of the raw data obtained from traffic data websites.

Table 1: Snapshot of extracted raw data

Segment	crash time	date	direction	Freeway	CRASHSEVERVITY	ROADSURF	ROADCOND1	LIGHTING	WERAINLEARTHER1	Volume	Speed
MAIN ST	1805	1/5/2010	S	5	Injury (Complaint of Pain)	Dry	No Unusual Condition	Daylight	CLEAR	672	29.8
LA PAZ	718	1/5/2010	N	5	Injury (Complaint of Pain)	Dry	No Unusual Condition	Dark - Street Lights	CLOUDY	669	62.2
OSO PKWY	735	1/6/2010	N	5	Injury (Complaint of Pain)	Dry	No Unusual Condition	Daylight	CLEAR	641	61
NEWPORT AV	820	1/7/2010	S	5	Injury (Complaint of Pain)	Dry	No Unusual Condition	Daylight	CLEAR	718	43.6
4TH ST	855	1/7/2010	N	5	Injury (Severe)	Dry	No Unusual Condition	Daylight	CLEAR	387	65.9
MAIN ST	1840	1/8/2010	N	5	Injury (Complaint of Pain)	Dry	No Unusual Condition	Dark - Street Lights	CLEAR	744	40
CROWN VALLEY PKWY	1740	1/8/2010	N	5	Injury (Complaint of Pain)	Dry	No Unusual Condition	Dusk - Dawn	CLEAR	337	35.8
EL TORO RD	1700	1/8/2010	N	5	Injury (Complaint of Pain)	Dry	No Unusual Condition	Dark - Street Lights	CLEAR	637	59
REDHILL AV	715	1/9/2010	N	5	Injury (Complaint of Pain)	Dry	No Unusual Condition	Daylight	CLEAR	308	69.6
CAMINO DE ESTRELLA	745	1/11/2010	S	5	Injury (Complaint of Pain)	Dry	No Unusual Condition	Daylight	CLEAR	238	65.1
CHAPMAN AV	615	1/13/2010	N	5	Injury (Complaint of Pain)	WET	No Unusual Condition	Dark - Street Lights	CLOUDY	421	57.9
CULVER DR	1700	1/14/2010	S	5	Injury (Complaint of Pain)	Dry	No Unusual Condition	Dusk - Dawn	CLEAR	738	67.1
MAGNOLIA AV	1710	1/15/2010	S	5	Injury (Severe)	Dry	No Unusual Condition	Daylight	CLEAR	502	60.5

Overall, the data were from 292 segments from the Interstate 5 in California. The detailed information of main contributing factors is shown in Table 2.

Table 2: Detailed information of main contributing factors

Contributing Factors	minimum	maximum	average	Standard Deviation
Volume	97	931	514	589.7
Speed	11.3	71.5	41.4	42.5
Density	0.0178	0.5896	0.3037	0.4043

Research Approach

In this research, the data collected from the roadway segments of the Interstate 5 freeway will be used in creating and validating crash injury prediction models. When crashes occur on freeway, the consequences of the crashes vary depending on the traffic characteristics, roadway factors and environmental conditions. Three injury types are observed during the data collection process as minor injury (complaint of pain-back pain, neck pain, etc.), visible injury types (e.g.: broken leg) and severe injury. These three types will be considered as dependent variables as they depend on other factors. The minor injury is taken as a reference base category. Using the base category, the ratio of the probability of having severe injury over the minor injury during the crash will be determined in the validated models. In addition, the probability ratio of the visible injury over minor injury will be estimated.

Given the categorical feature of the crash injuries, the Multinomial Logit Regression Model, one type of the Generalized Linear Model (GLM), that includes the real time traffic data was developed using R. For this research, R will give the results for. This model is used in data analysis in many areas, such as health, social, behavioral, and educational (El-Habil, 2012) [3]. The logit transformation that will be using for this research is shown as below:

$$log_e\left(\frac{pr(yij)}{pr(yi1)}\right) = f_j(x) for\ j = 2, \dots, m \qquad (1)$$

Where f(x) is the regression model, i is ith observation and j is the crash injury category j.

Research findings/results

As mentioned before, in the R program, the minor injury crash type is used as the reference category. After attaching the excel file and putting the related R codes in the program, the coefficients of the independent variables are able to obtain as outputs. Although the lighting (daytime, night time etc.), roadway factors (roadway and road surface conditions) and environmental situation such as weather are considered as factors affecting the crash occurrence, they were not produced as results as the program recognized them as quite small contributing factors, not giving coefficients for outputs. In the R software data outputs shown in Table 3, only three statistically significant influential factors such as speed, volume and lighting are displayed. Since multinomial logit model is utilized, the observed coefficients are put in the logit transformation equations written as follows:

$$\text{Log}_e\left(\frac{\text{Prob (Severe Injury)}}{\text{Prob (Minor Injury)}}\right) = -3.53 + 0.014*\text{Volume} - 0.0043*\text{Speed} \qquad (2)$$

$$\text{Log}_e \left(\frac{\text{Prob (Visible Injury)}}{\text{Prob (Minor Injury)}} \right) = --1.60 - 0.00167*\text{Volume} + 0.017\text{Speed} \quad (3)$$

The intercepts are produced in the model, however, they cannot be interpreted in terms of variable. The primary interest is the coefficients and the probability ratios are interpreted based on the coefficients related to independent variables (speed, volume, and lighting). Therefore, R software and multinomial logit models indicate the relationship between injury type in terms of traffic volume, speed and lighting. When there is one unit change in the variable speed, the log of the ratio of two probabilities, P(Severe Injury)/P(Minor Injury) is decreased by 0.0043 which implies that the severe injuries during the crash are less likely to occur when there is the variation in speed. In terms of traffic volume, the probability ratio will also be increased by 0.0014, representing that severe injury risk would increase with higher traffic volume.

In another case of comparing visible injury over minor injury, there will be decreased by 0.00167 per unit traffic volume, meaning that the probability ratio of having visible injury over minor injury during crash occurrence will be reduced by that amount. However, these ratios are likely to increase by 0.017when there is one unit change in traffic speed.

Although the coefficients of speed, volume and lightings are given in the results, most of them are not marked as the factor that has high confidence level. In Table 4 shown below, the speed and volume for the other visible injury types are the most influential factors that have 95% and 93% confidence level respectively. In the severe injury type, none of its independent variables have the high confidence level, indicating that the speed, volume and lightings are not the most important factors that causes the severe injury. The probability ratios are also provided in Table 5.

Table 3: Outputs from R using multinomial regression model

Coefficient	Estimate	Std. Error	t-value	Pr(> l t l)
2*:(intercept)	-3.53512	1.32382	-2.6704	0.007576
3**:(intercept)	-1.60641	0.74151	-2.1664	0.030281
2:Volume	0.00141	0.00191	0.7359	0.461805
3:Volume	-0.00167	0.00098	-1.7046	0.088268
2:Speed	-0.00436	0.01673	-0.2607	0.794292
3:Speed	0.01733	0.00959	1.8057	0.070963

Notes: *: 2 represents the linear regression model associated with severe injury level;

**: 3 represents the linear regression model associated with minor injury level;

Table 4: R output using M multinomial logit model for visible injury type on freeway

	Level of Confidence
Visible Injury: Speed	93%
Visible Injury: Volume	91%

Table 5: Probability ratio (odd ratio) from R output

	Speed	Volume
Prob (Severe Injury)/Prob (Minor Injury)	0.995	1.004
Prob (Visible Injury)/Prob (Minor Injury)	0.998	1.017

Conclusion

For this research study, the factors that influence on the crash injury types are observed and run into the programming using multinomial regression model to understand the relationship between the crash injury types and the chosen traffic conditions and roadway conditions and environmental conditions. Real world 2010 data extracted from data collection websites for Interstate I-5 freeway segment are utilized for this research. Using the log of the ratio of two probabilities, the relation between traffic characteristics and lighting conditions are explained.

This paper reports that the probability of crash severe injury over the probability of minor injury will be increased by 1.004, when there is one unit increase in traffic volume. This indicates that the increase in traffic volume increase the chance of having severe injury when crash occurs. In terms of one unit change in speed variable, these two probability ratio will most likely to increase by 0.995. The relative risk of visible injury over minor injury is 0.998 in terms of a unit change in speed. Therefore, the probability of having visible injury during crash is most likely to increase when there is high speed. In the traffic volume variable, its ratio will be increased by 1.017.

Not all of the independent variables such as speed, and volume, have high confidence levels. Only the speed and volume in visible injury type have confidence level of 93% and 91%, indicating that they are the most influential factors in visible injury type on freeway crash occurrence. However, the results regarding severe injury is not conclusive because speed and volume have low confident level showing that these two factors do not impact much on the case of severe crash type. There could be other factors such driver behaviors, traffic density and types of vehicle, roadway geometry (horizontal and vertical curve) may also contribute as influential factors in the case of crashes. In this case study, it is identified that there is the relationship between severe injury type and visible injury and traffic characteristics such as traffic speed and volume.

REFERENCES

[1] Abdel-Aty, M., Uddin, N., Pande, A., Abdalla, F. M., Hsia, L., 2004. Predicting freeway crashes from loop detector data by matched case-control logistic regression. Transportation Research Record: Journal of the Transportation Research Board, 1897(1), 88-95.

[2] Association for Safe International Road Travel: http://www.asirt.org/

[3] El-Habil, Abdalla M. "An Application on Multinomial Logistic Regression Model." *Pakistan Journal of Statistics and Operation Research*, 8.2 (2012): 271-291.

[4] NHTSA (August 2014). Early Estimate of Motor Vehicle Traffic Fatalities for the First Quarter of 2014.

[5] Preston,H, 1996, Potential safety benefits of intelligent transportation system (ITS) technologies. Semisequicentennial Transportation Conference Proceedings, Center for Transportation Research and Education, Iowa State University.

[6] Zheng, Z., Ahn, S., Monsere, C. M., 2010. *Impact of traffic oscillations on freeway crash occurrences. Accident Analysis and Prevention*, 42(2), 626-636.