

Study on Severity and Influencing Factors of Injury at Intersections

Linlin Wang^a, Junyou Zhang^b and Yuan Feng^c

Shandong University of Science and Technology, Qingdao 266590, China

^awll19950705@163.com, ^bjunyouzhang@163.com, ^c2207222659@qq.com

Abstract. Intersections are the places where traffic accidents occur frequently. As the hub of urban transportation, ensuring the efficient and safe operation of vehicles has become a problem that needs to be solved. The research on the mechanism of the severity of traffic accident injury at intersections can provide an important basis for traffic managers or autonomous vehicles to make driving decisions under dangerous conditions. The K-means clustering is used to analyze the traffic accident data, and the vehicle-road-environment is taken as the internal cause variable to characterize the severity of the collision injury, and the clustering results are used to classify the severity of the injury. The severity of injury can be divided into mild injury, moderate injury and severe injury. The clustering results based on the severity of the injury show that the four most significant factors are the type of intersection, the type of collision vehicles, the relative speed and the signal control of the intersection. The results of the study put forward policy suggestions for the planning and management of urban intersections from many angles.

Keywords: intersections, crash injury severity, influencing factors, K-means cluster

1. Introduction

With the continuous development of social economy and the continuous increase of car ownership, traffic accidents occur frequently, resulting in serious loss of life and property. According to the WTO survey, about 1.35 million people die in road traffic accidents every year [1]. This study uses the intersection traffic accident data collected by the United States National Security Administration (NHTSA) to study the mapping relationship between collision severity and vehicles, roads and environment at different intersections, and to determine the risk factors affecting intersection traffic accidents.

Intersections as a frequent point of traffic accidents, in order to deeply understand the internal mechanism leading to frequent traffic accidents at intersections, starting from traffic design and management and vehicle decision-making at intersections, reduce the frequency and severity of traffic accidents at intersections. In recent years, a large number of scholars in various countries are studying the relationship between crash injury severity and potential risk factors. For example, wang [2]use multivariable time series probit model to establish intersection collision accident severity prediction model. It is found that the type of collision, the type of intersection and the weather and lighting conditions are the core factors of the severity of traffic accidents. Alkhlaifi [3]use ordered logistic model to analyze the risk factors of traffic accidents under the intersection, and establish an analysis model for the significant factors that affect the severity of the collision. The results show that road design and traffic signals have a significant impact on the severity of collision damage. Cai [4] analyze the severity of traffic accident injury at intersections and different road sections, excavate the heterogeneity between the severity of collision injury and the influencing factors, and bring the relationship between road characteristics and traffic characteristics on the severity of collision injury into the classification model of the severity of collision injury. The results emphasize the effect of the severity of collision damage at intersections. Zeng [5] use neural network to establish a frequency prediction model of injury severity for traffic accidents occurred in different road sections, and reveal the exact effect of each important explanatory variable. The results show that different collision modes and road characteristics have different effects on the severity of the injury.

The above research explored the development process of crash injury severity classification model from traditional statistical model to modern machine learning model, which promotes the

development of traffic safety and management. As we all know, the crash injury severity prediction model is different in different accident scenarios, but this paper makes a cluster analysis of the traffic accident injury severity at different intersections, according to the accident injury severity in different intersection scenes to determine the internal relationship between the severity of crash injury and influencing factors under different working conditions.

2. Data Preparation

2.1 Crash Data Description

In this study, the vehicle collision accident data comes from the 2015 GES data set in the NASS database of the NHTSA department, which is mainly composed of three sub-data sets, namely, the accident data set, the vehicle data set and the collision participant data set [6-7]. The accident data set includes road conditions, environmental conditions and accident-related characteristics, and the vehicle data set includes a large number of characteristic variables of the vehicles involved in the accident, such as pre-collision speed, vehicle obstacle avoidance decision, etc., the collision participant data set includes a large number of characteristic variables of accident participants (drivers, passengers, pedestrians, cyclists, etc.), such as collision casualty degree, age, gender, etc. Each accident record contained in the three sub datasets is matched by a unified collision accident label (CASENUM). The accident sub-database describes the information of each accident in detail, including speed limit, weather condition, road section type, road gradient, road condition, traffic control equipment, accident type, accident time, accident date, accident location and accident injury severity. The vehicle sub database records in detail each vehicle involved in the accident, including vehicle type, the main cause of the accident, vehicle driving intention and other variables; the personnel sub database records the personnel information involved in the accident in detail, including the degree of injury, age, sex, whether thrown out of the car, seat belt use, airbag deployment, drunk driving, drug driving and other variables.

2.2 Explanatory Variable

GES records the accident data information of vehicle collisions in different regions, including highway accident information and urban road accident information. Combined with the statistical analysis of the data, the number of accidents at intersections accounts for about 60% of the total number of accidents. In the accident data, vehicle-road-environment is the main factor to evaluate the severity of crash injury. The specific classification is shown in Table 1.

Table 1. Traffic accident data discretization table.

| Variable name | A | B | C |
|---------------------------|------------------------|----------------|----------------|
| Collision vehicle type | Standard passenger car | Bus | Heavy truck |
| Relative speed | 0-20mph | 20-40mph | 40-60mph |
| Traffic control equipment | Traffic Signals | No Controls | |
| Intersection type | Four-Way Intersection | T-Intersection | Y-Intersection |
| Date of collision | Working Day | Rest Day | |
| Road gradient | Level | Grade | |
| Roadway condition | Dry | Wet | Snow/Ice |
| Weather conditions | Sunny/Cloudy day | Rainy day | Snowy day |
| Curvature | Straight | Curve | |

In the final data set estimated by the model, after excluding the collisions with incomplete data, a total of 1200 accidents involving intersections and vehicles were considered.

3. Methodology

K-means clustering algorithm is a machine learning algorithm for clustering [8-10], which takes the Euclidean distance as the similarity measure and the sum of square error criterion function as the clustering criterion function. The specific steps are as follows:

Step1: The initial accident data sample is used to determine the clustering number k and the initial clustering center;

Step2: The distances between each book and k initial center points are calculated, and the samples are assigned to the nearest initial center points according to the distance. To calculate the distance, which is expressed as follows:

$$L_{ij} = \sqrt{\sum_{k=1}^m (Z_{ik} - Z_{jk})^2} \quad (1)$$

Step3: The average values of accident data samples of each class are calculated and used as k new clustering centers;

Step4: The squared error cost function are used to verify the convergence, which is expressed by eq.(2). Repeat steps (2), (3) to E swing is very small, clustering ends.

$$E = \sum_{i=1}^k \sum_{Z_q \in C_i} (Z_q - m_i)^2 \quad (2)$$

Where C_i is the sample set; m_i is the sample centers; Z_q is the sample.

4. Results and Discussion

Based on the variable grouping of intersection-vehicle crash accidents, the severity of vehicle crash injury severity is divided into three categories: mild injury, moderate injury or serious injury. In order to select the appropriate number of clusters, different clustering models are estimated, ranging from 1 to 5. Three clusters are determined, and the results are as follows: cluster 1 accounted for 53.8% of the sample, cluster 2 accounted for 27.9% of the sample, and cluster 3 accounted for 18.3% of the sample.

Cluster1. This group includes traffic accidents mainly at four-way intersections. In this group, the relative speed between vehicles is 0-20mph, and the intersections have good signal control. On the other hand, in terms of environmental characteristics, it is a sunny day without the influence of any disaster weather. In addition, the road features are both level and dry. Therefore, this kind of injury severity classification is defined as mild injury.

Cluster2. This group occurs under different types of intersections, the signal control of intersections is similar to the first group, the relative speed between vehicles is different, most of the relative speeds are higher, which occur in different road environments, it is worth mentioning that environmental characteristics and road characteristics affect the clustering results of the severity of injury at this kind of intersections. The clustering result of this category is moderate injury.

Cluster3. The accidents in this group are mainly T-intersections or Y-intersections. The relative speed between vehicles is mostly 40-60mph. In addition, most of the accidents occurred on weekdays, and most of them were affected by the weather, such as rain and snow. in addition, vehicle accidents occurred on some steep or curved roads.

In this paper, the K-means cluster analysis is used to study the severity of damage in intersection collision accidents. Based on the cluster analysis of vehicle-road-environment as the main influencing factors in the accident data, the internal relationship between injury severity and influencing factors is fully observed, with emphasis on the cross-level interaction of characteristics such as collision severity and vehicle environment and road. these factors show different levels of injury severity in the severity of injury.

5. Conclusion

This study attempts to analyze the relationship between influencing factors and the crash injury severity at intersections. The K-means cluster is employed using the crash data collected from NHTSA. The contribution of this study and its importance for improving traffic safety at intersections can be summarized from following two aspects:

(1) Using K-means cluster analyze the relationship between the crash injury severity and influencing factors. This results will be definitely helpful to understand the possibility or severity of accidents occurred at different intersections.

(2) According to the results of the analysis, the realtive speed and collision vehicle type could not be ignored. To a certain extent, the results were basically consistent with the previous research, which further demonstrated the rationality of the K-means cluster models in this paper.

References

- [1]. Information on: www.who.int/mediacentre/factsheets/fs358/en/.
- [2]. Kai Wang, Tanmoy Bhowmik, Shamsunnahar Yasmin, et al. Multivariate copula temporal modeling of intersection crash consequence metrics: a joint estimation of injury severity, crash type, vehicle damage and driver error. *Accident; analysis and prevention*. Vol. 125 (2019) p. 188-297.
- [3]. Alkhlaifi A, Galadari A. Identifying the risk factors affecting crash severity at intersections with considering crash characteristics and signal configuration using an ordered logistic model. *2018 Advances in Science and Engineering Technology International Conferences*. Dubai. 2018-03-21, p. 1-7.
- [4]. Cai Q, Abdel-Aty M, Lee J, et al. Developing a grouped random parameters multivariate spatial model to explore zonal effects for segment and intersection crash modeling. *Analytic methods in accident research*. Vol. 19 (2018) p. 1-15.
- [5]. Zeng Q, Huang H, Pei X, et al. Modeling nonlinear relationship between crash frequency by severity and contributing factors by neural networks. *Analytic methods in accident research*. Vol. 10 (2016) p. 12-25.
- [6]. Information on: www.nhtsa.gov/research-data/national-%20automotive-sampling-system-nass.
- [7]. Liao Y, Zhang J, Wang S, et al. Study on Crash Injury Severity Prediction of Autonomous Vehicles for Different Emergency Decisions Based on Support Vector Machine Model. *Electronics*. Vol. 7 (2018) p.381.
- [8]. Casado-Sanz N, Guirao B, Lara Galera A, et al. Investigating the Risk Factors Associated with the Severity of the Pedestrians Injured on Spanish Crosstown Roads. *Sustainability*. Vol. 11 (2019). p. 1-18.
- [9]. Kouhi Esfahani R, Shahbazi F, Akbarzadeh M. Three-phase classification of an uninterrupted traffic flow: a k-means clustering study. *Transportmetrica B: Transport Dynamics*. Vol. 7 (2019). p. 546-558.
- [10]. Saha A, Chandra S, Ghosh I. Assessment of level of service for urban signalized intersections in India. *CURRENT SCIENCE*.Vol.117 (2019). p. 1516.