

# Analysis of Influencing Factors of Urban Road Traffic Accidents Based on Association Rules

Jiaqi Yang<sup>1, a</sup>

<sup>1</sup>Civil Aviation University of China, Air Traffic Management College, Tianjin, China

<sup>a</sup>1320137085@qq.com

**Abstract.** Urban road traffic conditions are complex and changeable. Affected by multiple factors, the mechanism and main influencing factors of traffic accidents are difficult to define. This paper aims to study the distribution characteristics and main factors influencing urban road traffic accidents. According to the collected data of urban traffic accidents in Britain, the basic characteristics including trunk road flap, weather conditions, time, maximum speed limit, first road classification and other factors are all analyzed, with association rules used to explore its influence mechanism. From aspects of vehicle, environment and road, this paper not only focuses on the influence mechanism of the interaction of two factors and three factors in urban road traffic accidents, but also puts forward targeted accident prevention measures accordingly.

**Keywords:** Urban Road Traffic Accidents; Influencing Factors; Association Rules; Preventive Measures; Apriori Algorithm.

## 1. Introduction

The traffic accident refers to an incident where a vehicle causes personal injury or property loss due to a fault or accident on the road. In addition to being caused by unspecified personnel violating road traffic safety laws and regulations, traffic accidents can also be triggered by irresistible natural disasters such as earthquakes, typhoons, flash floods, lightning strikes, etc. In recent years, China has formulated a national unified road traffic accident information registration standard and established a comprehensive accident management information system [1]. Therefore, studying the characteristics and influencing factors of urban road traffic accidents can provide a scientific basis for formulating accident prevention measures.

Scholars at home and abroad have conducted numerous research on the characteristics and influencing factors of urban road traffic accidents. Besides, basic statistical methods are widely used in earlier related studies. In 2008, the results of an opinion poll finished by Vanlaar [2] using Ontario drivers showed that 58.6% of drivers had fatigued driving experience. In 2015, after analyzing the characteristics of expressway traffic accidents in ice and snow weather, Wan Yujia [3] found that the number of traffic accidents in ice and snow weather was 19% more than that in normal weather conditions. In 2024, Song Ligang [4] made a comprehensive and in-depth analysis of the influencing factors of highway traffic accidents based on the 5M1E model. He divided the influencing factors into unsafe behavior of drivers, poor vehicle conditions, road and environmental factors, abnormal weather conditions and inadequate construction management and safety measures. In 2023, a Logistic regression model was established by Ke Xingan [5] who used the traffic accident data of a city from 2016 to 2020. The results indicated that the right-turn state of large and medium-sized vehicles was more likely to cause serious traffic accidents.

In recent years, the Poisson model, negative binomial model, Bayesian model, etc. have been used to analyze the influencing factors of accidents. In 2014, Li Rui et al. [6] generated SPSS data by using the "Road Traffic Accident Injury Information Collection Form" of traffic accident cases in a city from April to September 2013. They fitted the data with the Poisson regression model and negative binomial regression model respectively, which showed that accidents at night had more tendency to cause death than during the day. In 2018, based on the data from 31 provinces and cities across China

from 2004 to 2015, Zhou Hongmei [7] constructed a panel data model of the number of traffic accidents, the number of traffic accident injuries and the number of traffic accident deaths respectively. Firstly, the ordered Logistic model and the ordered Probit model are used to analyze traffic accident data, with the second-order interaction terms between variables included in the model to mine the relationship between influencing factors. Meanwhile, adding interaction terms also improves the fitting degree and accuracy of the model, but the model structure is complicated. Thus, the Bayesian network model is further used to study the relationship between the influencing factors of traffic accidents. The results indicate that the increase in highway mileage per 10,000 people, the number of public transport vehicles owned by the urban population, the number of urban road lighting per kilometer and the number of medical and health institutions per 10,000 people can reduce the number and severity of traffic accidents, while the increasing urban road area per capita and illiteracy rate will lift the number and severity of traffic accidents.

Compared with traditional statistical analysis methods, it's unnecessary for the dependent variables and independent variables to be determined in advance in association rules analysis. Thus, some valuable associations that are hard to discover in traditional methods can be fully explored in association rules analysis. In 2018, Das [8] used the association rule Eclat algorithm to analyze the traffic accident data in Louisiana. The results showed that the interaction between men, off-peak hours and frontal collisions was highly correlated with traffic accidents. In 2024, Wu Tingchao et al. [9] used SPSS Statistics and SPSS Modeler to analyze the drug dosage form, frequency, medicinal properties, medicinal flavor, meridian return, dosage, association rules and systematic clustering of compounds for the in-depth analysis of prescription medication rules. According to the results, the medication rules of compound prescription for the treatment of T2DM with deficiency of both qi and yin provided inspiration for the related clinical and scientific research. In 2024, Wang Jiadan et al. [10] adopted ecological auxiliary adjustment to realize governance planning based on landscape layout and river bank construction under weighted association rules. Ma Xiaoping [11] et al. also applied the optimized Apriori correlation analysis to find out the correlation between printing defect detection data in 2024, so as to solve printing defect detection in printing enterprises. This method is proven feasible in the analysis of printing defect detection data and plays a guiding role in reducing the printing defect rate.

By analyzing the characteristics of road traffic accidents in Britain in a recent decade, this paper further explains the influence mechanism of multi-factor interaction of urban road traffic accidents by using association rules, and puts forward targeted accident prevention measures.

## **2. Data Collection and Preprocessing**

### **2.1. Data Collection**

This paper selects the traffic data of all urban traffic roads in the UK in 2020 collected by the British Transport Bureau as a sample. After preprocessing and deleting blank values, the sample includes the data of 49,317 accidents, with the preprocessed sample data shown as follows:

#### **2.1.1. Dependent Variables**

The dependent variable in this paper is the number of casualties.

#### **2.1.2. Independent Variables**

The dependent variables of this paper are the characteristics affecting traffic accidents, including the number of vehicles, time, first road classification, maximum speed limit, warning signs, light conditions, weather conditions, road conditions, special situations and main road flags. The data is preprocessed such as integration, deletion of null values, deduplication, discretization, desensitization, etc. to obtain data samples, which are shown in Table 2.

**Table 1.** Example Data Table

Vehicles Number	Casualty Number	Time	First Road Classification	Max Speed Limit	Warning Signs	Lighting	Weathers	Road	Special Situations	Main Road Flags
1	1	Early Morning	C	Low	Many	Good	Bad	Wet	No	Need More
3	2	Early Morning	C	Low	Normal	Good	Sunny	Dry	No	Need More
2	1	Early Morning	A	Low	Normal	Good	Sunny	Dry	No	Need More

**Table 2.** Description of Data Samples

Independent Variable	Data	Tools
Number of Vehicles	Less, Medium, and Many	Classified by the Number of Vehicles in Traffic Accidents
Time	Early Morning, Morning, Afternoon, Evening	Time Classification Table
First Road Classification	Classification Name	Classification Code Table
Maximum Speed Limit	Low, Medium, High	Speed Limit Classification
Warning Signs	Few, Normal, and Many	Divided by the Number of Warning Signs
Lighting Conditions	Intact, Good, Medium, Poor, Completely Damaged	Table of Lighting Conditions Classification
Weather Conditions	Sunny, Cloudy, Light Rain, Light Snow, Heavy Rain, heavy Snow, Dense Fog, Severe	Classified Based on Cloud Height, Visibility, Rain and Snow
Road Conditions	Dry, Wet, Snowy, Rainy, Icy, Muddy	Table of Road Conditions Classification
Main Road Flags	Needs More, Does not Need More	Whether There is a Flag
Special Situations	Automatic Traffic Signals Need to be Updated; Pavement Conditions Need to be Predicted; Automatic Traffic Signals Need to be Added for Prediction and Normal Work	

## 2.2. Data Preprocessing

In a data set, variables without missing values are complete variables, while variables with missing values are incomplete variables. Deleting and filling are two methods to deal with missing values. Adopting the deletion of missing values, this paper replaces the missing value with a null value and deletes the null row.

## 3. Analysis of Basic Characteristics of Urban Road Traffic Accidents

This paper focuses on the analysis of the basic characteristics of urban road traffic accidents according to the uncontrollable environment and controllable environment.

The uncontrollable environment is road time. According to data analysis, the probability of traffic accidents in the early morning is only 5.8%, while that in the daytime is as high as 69%.

The controllable environment includes road classification, lighting conditions, road conditions, weather conditions, etc. Road classification is a vital factor causing traffic accidents. When the road classification is A, the probability of traffic accidents is 50%. It shows that changes affecting traffic have occurred with road classification A, leading to greatly increased traffic accidents. Lighting condition is another crucial factor in traffic accidents. When the lighting condition is intact, there are almost no traffic accidents. Otherwise, the possibility of traffic accidents exceeds 60% when the lighting condition is damaged. Road and weather conditions are also significant influencing factors that can't be ignored. The impacted road condition will increase the possibility of traffic accidents, while the possibility of traffic accidents is only 5.4% when the weather is sunny.

#### 4. Correlation Analysis of Influencing Factors of Urban Road Traffic Accidents

##### 4.1. Association Rule Theory

Compared with traditional methods, association rules analysis does not need to specify independent variables and dependent variables. Traditional methods should calculate the confidence and support of each possible rule, which costs a lot. Thus, the advantages of association rules can make the influencing factors have a more valuable relationship. In this paper, the Apriori algorithm used to analyze the influencing factors of urban road traffic accidents is divided into the following steps. (1) Calculate the support degree of a single itemset, with the frequent itemsets screened out according to the Apriori algorithm. (2) Connect single items to form a binomial set, calculate its support degree, and use the Apriori algorithm to screen it again. (3) Connect to form three itemsets and calculate the support degree to obtain the maximum frequent itemset.

In the Apriori algorithm, support, confidence and promotion are three important indexes in association rules.

Support is defined as  $SupX = \frac{\text{Number of the Item (X) in the Itemset}}{\text{Total Number Recoded in the Itemset}}$ ,

Confidence is defined as  $Con(X \Rightarrow Y) = \frac{Sup(XY)}{Sup(X)}$ ,

Conditional probability is defined as  $P(YX) = \frac{P(XY)}{P(X)}$ ,

Lift is defined as  $Lift(A \rightarrow B) = \frac{Con(A \rightarrow B)}{Sup(B)}$ , that is, the ratio of the probability of B recurring based on the occurrence of A to the probability of B occurring alone.

The promotion reflects the correlation between item A and item B in the association rules. The promotion greater than 1 and above indicates a higher positive correlation, while that less than 1 and below indicates a higher negative correlation. Meanwhile, a promotion equal to 1 indicates no correlation. The promotion is an index used to measure the relationship between antecedents and contexts. The higher the promotion, the more convincing the rule is.

When association rule analysis is used to study the combination of influencing factors that occur simultaneously in urban road traffic accidents, the thresholds of support (S) and confidence (C) should be set in advance. To be specific, the thresholds are set to support (S)  $\geq 30\%$  and confidence (C)  $\geq 90\%$  respectively.

##### 4.2. Results Analysis Based on Association Rules

According to Tables 3-6, the results of binomial and trinomial association rules as well as the corresponding confidence and support degrees are shown.

**Table 3. Binomial Correlation Results**

ID	Preceding Item	Latter Item	Support	Confidence
1	Main Road Flags: Need More	Number of Casualties: Less Serious Accidents	1.00	1.00
2	Weather Conditions: Sunny	Number of Casualties: Less Serious Accidents	0.845	1.00
3	Time: Afternoon	Number of Casualties: Less Serious Accidents	0.387	1.00
4	Time: Evening	Number of Casualties: Less Serious Accidents	0.311	1.00
5	Maximum Speed Limit: Low	Number of Casualties: Less Serious Accidents	0.413	1.00
6	Maximum Speed Limit: High	Number of Casualties: Less Serious Accidents	0.362	1.00
7	First Road Classification: A	Number of Casualties: Less Serious Accidents	0.518	1.00
8	Warning Signs: Normal	Number of Casualties: Less Serious Accidents	0.343	1.00
9	Warning Signs: Few	Number of Casualties: Less Serious Accidents	0.530	1.00
10	Road Conditions: Dry	Number of Casualties: Less Serious accidents	0.746	1.00
11	Weather Conditions: Sunny	Number of Casualties: Serious Accidents	0.800	1.00
12	Time: Morning	Number of Casualties: Serious Accidents	0.400	1.00
13	Time: Evening	Number of Casualties: Serious Accidents	0.400	1.00
14	Maximum Speed Limit: Low	Number of Casualties: Serious Accidents	0.400	1.00
15	Maximum Speed Limit: High	Number of Casualties: Serious Accidents	0.400	1.00
16	Light Conditions: Intact	Number of Casualties: Serious Accidents	0.600	1.00
17	First Road Classification: A	Number of Casualties: Serious Accidents	0.400	1.00
18	First Road Classification: M	Number of Casualties: Serious Accidents	0.400	1.00
19	Warning Signs: Normal	Number of Casualties: Serious Accidents	0.400	1.00
20	Warning Signs: Few	Number of Casualties: Serious Accidents	0.400	1.00
21	Weather Conditions: Light Rain	Number of Casualties: Extremely Serious Accidents	0.500	1.00
22	Weather Conditions: Sunny	Number of Casualties: Extremely Serious Accidents	0.500	1.00
23	Time: Morning	Number of Casualties: Extremely Serious Accidents	0.500	1.00
24	Time: Evening	Number of Casualties: Extremely Serious Accidents	0.500	1.00
25	Road Conditions: Wet	Number of Casualties: Extremely Serious Accidents	0.500	1.00
26	Light Conditions: Intact	Number of Casualties: Extremely Serious Accidents	0.500	1.00
27	Lighting Conditions: Poor	Number of Casualties: Extremely Serious Accidents	0.500	1.00
28	First Road Classification: A	Number of Casualties: Extremely Serious Accidents	0.500	1.00
29	Warning Signs: Normal	Number of Casualties: Extremely Serious Accidents	0.500	1.00

#### 4.2.1. Results Analysis of Binomial Association Rules

According to Table 3, among the binomial association rules, the important association rules are related to the main road flag, weather condition, time, maximum speed limit, first road classification, warning sign, road condition and lighting condition. Rule 1 with the greatest support in Table 3 shows that the less serious accident when the main road flag is incomplete is related to the absence of road signs. In urban roads, if there are no road signs, vehicles will not drive on the correct road, which will greatly increase the possibility of accidents. Therefore, the flags of main roads should be rectified and supplemented in time to prevent traffic accidents.

#### 4.2.2. Results Analysis of Three-item Association Rules

Among the three-item association rules, the highest support in Table 4 is 3. It indicates that when the main road flag is incomplete and the weather is sunny, it's prone to occur less serious traffic accidents, which is consistent with the result of binomial association rules. In addition, Rules 4 and 5 are related to weather conditions. These two rules show that under sunny weather conditions, time and maximum speed limits can cause traffic accidents.

Table 5 demonstrates three-item association rules that cause serious traffic accidents. Besides, 1 and 3 have the highest support, which shows that drivers are easy to go on roads that should not be driving without road signs, thus triggering traffic accidents.

Three-item association rules that cause extremely serious traffic accidents are demonstrated in Table 6. Rules 1 and 2 are related to main road flags, where 1 implies that traffic accidents and tire slippage are easy to occur when there is light rain. Rules 3, 4 and 5 are related to weather conditions. Rule 3 shows that driving vehicles at night is more likely to cause traffic accidents. Rule 4 indicates that intact lighting conditions can greatly reduce traffic accidents, and Rule 5 proves that intact warning signs can help reduce traffic accidents. Therefore, we should pay attention to the quality of vehicles and road facilities, increase punishment, and formulate more comprehensive and detailed traffic laws and regulations to avoid serious traffic accidents.

**Table 4.** Results of Three-item Association Rules for Less Serious Accidents

1	Main Road Flags: Need More; Time: Afternoon	Number of Casualties: Less Serious Accidents	0.387	1.00
2	Main Road Flags: Need More; Time: Evening	Number of Casualties: Less Serious Accidents	0.311	1.00
3	Main Road Flags: Need More; Weather Conditions: Sunny	Number of Casualties: Less Serious Accidents	0.845	1.00
4	Weather Conditions: Sunny; Time: Afternoon	Number of Casualties: Less Serious Accidents	0.327	1.00
5	Weather Conditions: Sunny; Maximum Speed Limit: Low	Number of Casualties: Less Serious Accidents	0.341	1.00

**Table 5.** Results of Three-item Association Rules for Serious Accidents

1	Main Road Flags: Need More; Weather Conditions: Sunny	Number of Casualties: Serious Accidents	0.800	1.00
2	Weather Conditions: Sunny; Time: Morning	Number of Casualties: Serious Accidents	0.400	1.00
3	Weather Conditions: Sunny; Road Conditions: Dry	Number of Casualties: Serious Accidents	0.800	1.00
4	Maximum Speed Limit: High; First Road Classification: M	Number of Casualties: Serious Accidents	0.400	1.00
5	Light Condition: Intact; First Road Classification: A	Number of Casualties: Serious Accidents	0.400	1.00

**Table 6.** Results of Three-item Association Rules for Extremely Serious Accidents

1	Main Road Flags: Need More; Weather Conditions: Light Rain	Number of Casualties: Extremely Serious Accidents	0.500	1.00
2	Main Road Flags: Need More; Weather Conditions: Sunny	Number of Casualties: Extremely Serious Accidents	0.500	1.00
3	Weather Conditions: Light Rain; Time: Evening	Number of Casualties: Extremely Serious Accidents	0.500	1.00
4	Weather Conditions: Light Rain; Light Conditions: Intact	Number of Casualties: Extremely Serious Accidents	0.500	1.00
5	Weather Conditions: Light Rain; Warning Signs: Normal	Number of Casualties: Extremely Serious Accidents	0.500	1.00

## 5. Conclusions

According to the data information on British urban roads, this paper first uses descriptive statistical analysis to analyze the characteristics of traffic accidents from the aspects of main road flag, weather condition, time, maximum speed limit, first road classification, warning sign, road condition and lighting condition. Then, the Apriori algorithm of association rules is used to analyze the influencing factors of traffic accidents. By screening and analyzing the results of association rules, it shows the interactive influence of various factors on urban road traffic accidents, with conclusions listed as follows:

- (1) In the absence of main road flags, driving in the wrong position is a crucial cause of traffic accidents.
- (2) If it is raining, the road surface is wet, causing the driver to experience emergencies such as tire slippage, and traffic accidents will occur when it cannot be avoided.
- (3) Speed limit is also one of the important factors. On roads with high speed limit, the probability of traffic accidents is usually higher.
- (4) On roads with poor lighting conditions, the influence of lights will result in drivers' mistakes in judgment of road conditions, leading to traffic accidents.

Based on the above conclusions, this paper puts forward the following suggestions:

- (1) We should not only regularly update the flags and lights of main roads, but also publish the latest flags of main roads on recognized websites or navigation maps, so as to facilitate reasonable and correct traffic.
- (2) In bad weather conditions, we should issue weather forecasts to remind drivers of vigilant driving to reduce traffic accidents.
- (3) Given that the danger of each section determines its speed limit, we should allocate each part of urban roads reasonably. By limiting the speed of vehicles, traffic accidents can be reduced.

These conclusions show that urban road traffic accidents involve complex interactions among many factors. Through the exploration of association rules, we can better grasp the internal mechanism of urban road traffic accidents and provide a scientific basis for formulating corresponding preventive measures. In addition, this paper has some limitations. The adopted data and association rules methods can be optimized in future research.

## References

- [1] Chen, L., Li, C. Y., Zhan, L. et al. (2022). Influencing factor analysis and prediction of urban road traffic accident patterns. *Journal of Chang'an University (Natural Science Edition)*, 42(04), 98-107. DOI:10.19721/j.cnki.1671-8879.2022.04.010.

- [2] Vanlaar, W., Simpson, H., Mayhew, D. & Robertson, R. (2008). Fatigued and drowsy driving: A survey of attitudes, opinions and behaviors. *Journal of Safety Research*, 39(3): 303-309.
- [3] Wan, J. Y. (2015). Analysis and preventive countermeasures of expressway traffic accidents in ice and snow weather. Nanjing: Master's Dissertation of Southeast University.
- [4] Song, L. G. (2024). Research on the influencing factors and countermeasures of traffic accidents in the operation area of highway renovation and expansion based on 5M1E. *Transport Energy Conservation & Environmental Protection*, 1-5.  
Retrieved from <http://kns.cnki.net/kcms/detail/10.1261.U.20240521.0845.002.html>.
- [5] Ke, X. A., Ding, L. M. & Zhao, D. (2023). Analysis on the influencing factors of traffic accidents severity for electric bicycles based on logistic-TAN. *Journal of People's Public Security University of China (Science and Technology)*, 29(02), 47-54.
- [6] Li, R. & Zhao, L. H. (2014). Analysis on the influencing factors of traffic accidents. *Journal of Mathematical Medicine*, 27(06), 634-636.
- [7] Cui, X. X. (2018). Analysis of road traffic accidents and the influential factors of severity. Dalian: Master's Dissertation of Dalian University of Technology.
- [8] Das, S., Dutta, A., Jalayer, M., Bibeka, A. & Wu, L. T. (2018). Factors influencing the patterns of wrong-way driving crashes on freeway exit ramps and median crossovers: Exploration using 'Eclat' association rules to promote safety. *International Journal of Transportation Science and Technology*, 7(2): 114-123.
- [9] Wu, T. C., He, M. M., Feng, H. Y. et al. (2024). Medication rule analysis of traditional Chinese medicine compound in treatment of type 2 diabetes mellitus with qi and yin deficiencies based on data mining. *Chinese Traditional and Herbal Drugs*, 55(14), 4811-4823.
- [10] Wang, J. D., Huang, Y., Zhou, K. Y. et al. (2024). River regulation planning method of water conservancy project based on weighted association rules. *Pearl River Water Transport*, (13): 117-119.
- [11] Ma, X. P., Cao, S. Z. & Li, Y. (2024). Correlation analysis of print inspection data based on optimized Apriori algorithm. *Journal of Beijing Institute of Graphic Communication*, (06): 22-26. DOI:10.19461/j.cnki.1004-8626.2024.06.003.