Front. Eng. Manag. 2025, 12(2): 414–424
https://doi.org/10.1007/s42524-024-4059-x

RESEARCH ARTICLE

**Shuo LIU, Liujiang KANG, Huijun SUN, Jianjun WU, Samuel AMIHERE**

# Exploring the factors of major road traffic accidents: A case study of China



**Abstract** Since the implementation of the transportation power strategy, China's transportation industry has developed rapidly, yet the number of road traffic accidents has remained high in recent years. Many scholars have investigated the factors influencing traffic accidents to find the underlying mechanisms, thereby enhancing road traffic safety. Compared to general accidents, the factors influencing major road traffic accidents are more complex. This study focuses on examining the relationships between factors affecting major road traffic accidents. Data on 968 major road traffic accidents from 2012 to 2018 in China were collected and organized. The accident information fields were analyzed to identify seven attributes: accident province, accident region, accident quarter, accident time, accident form, accident vehicle, and weather condition. The Apriori association rule algorithm was employed to mine and solve the strong association rules between accident attribute values. The associations between different influencing factors and the form of accident results were analyzed, with a deeper exploration of three-factor and four-factor rules. The results indicate that certain causal factors jointly contribute to major accidents, particularly in the western region, represented by Guangxi. These accidents mainly involved trucks and occurred in rainy and snowy weather during the first quarter. The conclusions of this research can provide the transportation management department with measures to improve urban road traffic safety and reduce the occurrence of traffic accidents.

**Keywords** road safety, traffic accidents, influencing factor, association rules, apriori algorithm

## 1 Introduction

Since the implementation of China's strategy to build a robust transportation sector, there has been a continued emphasis on the transportation industry, accelerating innovative transportation development and significantly improving the overall transportation standards. According to statistical data released by the Traffic Management Bureau of the Ministry of Public Security of the People's Republic of China, China's automobile ownership has reached 319 million by the end of 2022, with a total of 417 million motor vehicles. Additionally, the total operating mileage of railways has reached 155,000 km, and the total mileage of highways has reached 5.35 million kilometers. As China's economy continues to develop, the demand for the transportation industry continues to increase, accompanied by a high frequency of traffic accidents. Consequently, traffic safety has become a critically important concern for the public (Kang et al., 2021). To address this, the State Council's Work Safety Committee issued the "14th Five-Year Plan for National Work Safety," which provides a comprehensive blueprint for work safety during this period. The plan includes key objectives such as a 15% reduction in the number of production safety accident-related deaths and a 20% reduction in the number of major and extremely serious safety accidents. The plan also identifies the urgent need to reduce the occurrence of traffic accidents and prevent them from happening in the first place. According to

Received Apr. 1, 2024; revised May 12, 2024; accepted May 24, 2024

Shuo LIU
School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China; Key Laboratory of Transport Industry of Comprehensive Transportation Theory (Ministry of Transport), Beijing Jiaotong University, Beijing 100044, China

Liujiang KANG (✉), Samuel AMIHERE
School of Systems Science, Beijing Jiaotong University, Beijing 100044, China; Key Laboratory of Transport Industry of Comprehensive Transportation Theory (Ministry of Transport), Beijing Jiaotong University, Beijing 100044, China
E-mail: kanglj@bjtu.edu.cn

Huijun SUN, Jianjun WU
School of Systems Science, Beijing Jiaotong University, Beijing 100044, China

This research was supported by the National Natural Science Foundation of China (Grant Nos. 72288101, 72331001, 72361137003) and the Talent Fund of Beijing Jiaotong University (Grant No. 2023XKRC036).

official data from the National Bureau of Statistics China witnessed a total of 273,098 traffic accidents in 2021, resulting in 62,218 deaths, 281,447 injuries, and direct property losses of 1.45 billion yuan. Over the past decade, the number of deaths caused by accidents has consistently remained high, surpassing 50,000 annually. Among all types of traffic accidents, road traffic accidents account for the highest proportion, with extremely major accidents causing the greatest number of casualties and inflicting significant damage to property and personal security.

According to the Ministry of Public Security and the Ministry of Communications, traffic accidents can be classified into three categories: minor accidents, general accidents, and major accidents. A minor accident refers to a motor vehicle accident that results in light injuries to 1–2 people or property damage of less than 1,000 yuan for motor vehicles and less than 200 yuan for non-motor vehicles. A general accident refers to an accident that causes one to two serious injuries, more than three minor injuries, or property losses of less than 30,000 yuan. A major accident refers to an accident that results in 1–2 deaths, more than 3 but less than 10 serious injuries, or property damage of over 30,000 yuan but less than 60,000 yuan. An extremely serious accident refers to incidents that lead to the death of more than 3 people, or serious injuries involving more than 11 people, or the death of 1 person and serious injuries involving more than 8 people, or the death of 2 people and serious injuries involving more than 5 people, or incidents resulting in property loss exceeding 60,000 yuan. This paper specifically focuses on traffic accidents with fatalities, which are collectively referred to as major traffic accidents.

China, having the largest population and the highest number of motor vehicles in the world, faces significant challenges in improving road safety in terms of planning and operation (Zhang et al., 2022). Existing research indicates that there has been limited specialized investigation into major accidents in foreign countries. This could be attributed to significant disparities in population distribution and road environments between Western nations and China, resulting in a lower incidence of major accidents. Conversely, major accidents in China have remained alarmingly frequent in recent years, highlighting their significant hazards and consequences. Therefore, it is crucial for us to further explore the occurrence mechanisms, patterns, distinctive characteristics, and influencing factors of major accidents in order to develop targeted strategies for reducing or preventing accidents. This paper aims to analyze major road traffic accidents over a period of 7 years from 2012 to 2018, with the objective of identifying and analyzing the influencing factors of these accidents and exploring their underlying correlations.

Data mining is a cutting-edge technology that has emerged in the era of big data. It focuses on integrating and processing vast amounts of data stored in databases, employing statistical analysis methods (Kang et al., 2019a, Kang et al., 2019b). By utilizing data mining techniques, we can gain a comprehensive understanding of the fundamental characteristics and influencing factors of various types of traffic accidents in different fields (Huang et al., 2023). This understanding has significant implications for accident prevention, traffic safety management, the development of relevant laws and regulations, and ultimately, the promotion of traffic safety.

Various modeling methods, including machine learning neural network models and statistical models such as decision trees (Ceven and Albayrak, 2024, Ma et al., 2017), Bayesian networks (Kuang et al., 2019), and the Analytic Hierarchy Process (AHP) (Moslem et al., 2024), can be employed to study accident influencing factors. Compared to neural network models, the association rule method offers higher interpretability and practical significance in identifying accident influencing factors, addressing the limitations of machine learning and neural network models. Furthermore, compared to other statistical methods, association rules enable the exploration of hidden and hard-to-find correlations among the itemsets investigated in the accident database. Therefore, we opted to employ the association rule method to analyze accident data from multiple dimensions, including human, vehicle, road, and environment. This approach will help identify relevant relationships among causal factors and equip traffic management authorities with effective preventive measures.

Association rules were initially proposed by Agrawal to extract potential correlations between different attributes, mining the correlated relationships between item sets in a database. Since then, the algorithm has undergone continuous improvements (Agrawal and Srikant., 1994). Martín et al. (2014) utilized the Apriori algorithm to analyze information on traffic accidents in Andalusia, Spain. They studied the relationship between improvement factors, collision frequency, and dangerous road sections, and highlighted deficiencies in the Spanish government's traffic management practices. Hong et al. (2020) applied the Apriori algorithm to mine related risk factors in a data set of 19,038 freight truck accidents in Republic of Korea from 2008 to 2017, resulting in the generation of 90,951 rules. Their study revealed that speeding in rainy weather was highly associated with accidents, as was fatigue driving in clear weather. The research findings can inform the development of suggestions and policies aimed at reducing accidents.

The relationship between individual characteristics and accident proneness has been a challenging aspect of accident prevention research (Zhang et al., 2016). Ait-Mlouk and Agouti (2019) proposed a framework based on association rules to establish connections between variables. They employed this framework in a traffic accident data set in Morocco to extract related variables and association rules.

While research in various methods based on different data sources, variables, and samples has yielded fruitful results, the current focus of research mostly lies in general accidents. The key contributions of this paper are as follows: First, given the high risk and substantial characteristics of major accidents, along with their complex influencing factors, our study adopts a multi-faceted approach. We analyze the influencing factors of accidents by collecting major road traffic accident data and employing the association rules mining algorithm. Secondly, we aim to identify the interplay between multi-source and multi-dimensional factors at the time of the accident, establishing correlations and potential interactions among these factors. Thirdly, our analysis enhances the understanding of the root causes of accidents, enabling traffic authorities to devise targeted preventive measures to mitigate accidents. Finally, the findings of this study strive to improve road traffic safety, reduce the frequency of major accidents, foster a safer and more reliable road traffic environment, and protect individuals' lives and property.

## 2 Data source

This paper examines the data on major road traffic accidents in China from 2012 to 2018 for research purposes. After cleaning and processing the data, a data table of major traffic accidents is obtained, and preliminary analysis is conducted to identify the relevant factors influencing these accidents. The attributes of the various factors can be classified into accident occurrence time, accident occurrence region, accident form, direct cause of the accident, weather conditions, and collision vehicles. The occurrence of traffic crashes is the result of the combined effects of indicators from one or more dimensions in the system consisting of people, vehicles, roads, and the environment.

The time attribute includes the year, month, day, and specific time of the accident. If the month of the accident is known, it can be categorized into four quarters of the year, providing insights into the impact of different seasons on traffic accident occurrence. If the specific time of the accident is known, it can be grouped into time periods, allowing for the study of the influence of different time periods on traffic accident occurrence. This information can provide insights into the specific time periods that require a focus on traffic safety measures.

The distribution of accidents during different time periods in selected years is shown in Fig. 1.

The region attribute refers to the province where the accident occurred. Geographically, the provinces of China can be divided into four regions: eastern, western, central, and water area. (Note: The eastern region includes Beijing, Tianjin, Hebei, Liaoning, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, Hainan, and the municipalities directly under the central government. The central region includes Shanxi, Jilin, Heilongjiang, Anhui, Jiangxi, Henan, Hubei, and Hunan. The western region includes Inner Mongolia, Guangxi, Sichuan, Guizhou, Yunnan, Xizang, Shaanxi, Gansu, Ningxia, Qinghai, Xinjiang, and Chongqing. The water area includes the sea and the Yangtze River area.) As this paper focuses on road traffic accidents, Table 1 presents the distribution of accidents by region.

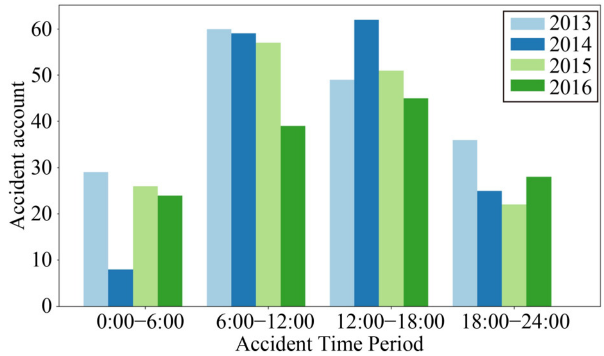


**Fig. 1** Accident distribution by time period chart.

Accidents can occur in various forms, including collisions, rollovers, falls, rear-end collisions, collisions with stationary vehicles, collisions with fixed objects, crushing, explosions, and others. Due to the unpredictable nature of accidents and the need for data analysis, this paper suggests grouping the three subtypes of rear-end collisions, collisions with stationary vehicles, and collisions with fixed objects into the overarching category of collisions.

The direct causes of accidents are diverse and necessitate classification for comprehensive analysis. This paper proposes a refined categorization of the direct causes of accidents by further subcategorizing driver-related errors. These errors include illegal passing, speeding, improper lane changes, and driving against traffic. While these errors can be attributed to driver behavior, their categorization requires further investigation due to the importance of data mining and analysis, as well as individual driving habits.

Weather conditions include clear, foggy, cloudy, rainy, and snowy. Subsequent analysis will examine the varying impacts of these different weather conditions on major road traffic accidents.

Accident vehicles can be primarily classified into three types: coaches, trucks, and hazardous material transport vehicles. Studying the probability of accidents in different vehicle types can inform the development of customized driver training and education programs. The distribution map of accidents in different vehicle types and weather conditions is presented in Fig. 2.

**Table 1** Distribution of accidents by region

| Accident area | Submitting province/autonomous region | Number | Total |
|---|---|---|---|
| Eastern region | Beijing | 5 | 168 |
| | Fujian | 30 | |
| | Guangdong | 64 | |
| | Hainan | 7 | |
| | Hebei | 28 | |
| | Liaoning | 19 | |
| | Jiangsu | 1 | |
| | Shanghai | 7 | |
| | Zhejiang | 18 | |
| | Heilongjiang | 16 | |
| | Shandong | 14 | |
| | Tianjin | 6 | |
| Western region | Gansu | 9 | 552 |
| | Guizhou | 71 | |
| | Inner Mongolia | 14 | |
| | Ningxia | 9 | |
| | Qinghai | 17 | |
| | Shaanxi | 28 | |
| | Guangxi | 127 | |
| | Sichuan | 92 | |
| | Xinjiang | 70 | |
| | Xizang | 10 | |
| | Chongqing | 4 | |
| | Yunan | 77 | |
| Central region | Anhui | 47 | 248 |
| | Henan | 94 | |
| | Hubei | 23 | |
| | Hunan | 30 | |
| | Jilin | 13 | |
| | Jiangxi | 17 | |
| | Shanxi | 12 | |

## 3 Association rule algorithm

Due to the interdependence between various factors affecting road traffic accidents, a certain degree of association exists. The association rule algorithm can reveal the implicit relationship between different factors and identify the combinations of different influencing factors. It also reflects the impact of these rules on accidents, which is helpful in proposing corresponding accident prevention and control measures.

The Apriori algorithm can discover implicit relationship

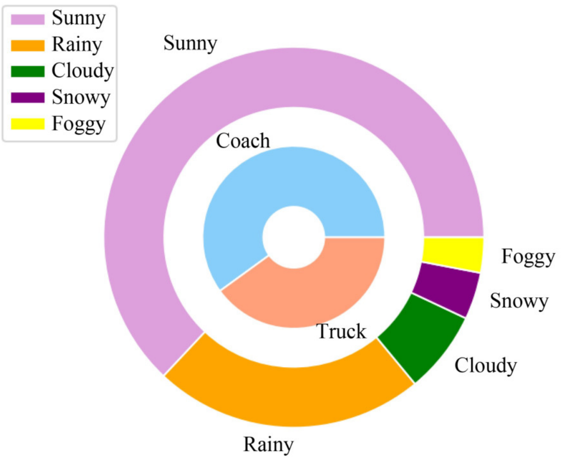


**Fig. 2** Distribution of accidents by weather conditions and vehicle type.

among a wide range of data types by iteratively calculating all frequent itemsets in the database that meet the minimum support threshold (Agrawal and Srikant, 1994). It further filters out association rules that satisfy the minimum confidence threshold.

Association rules are typically represented as $A \Rightarrow B$. Suppose $I = \{I_1, I_2, \ldots, I_m\}$ is the set of all items, where $A \subseteq I$, $B \subseteq I$, and $A \cap B = \phi$. Given the total transportation data warehouse, denoted as $D$, which includes accident types, vehicle types, and time periods, among others.

The support and confidence in the Apriori algorithm represent the proportion of influencing factors and their relatedness, respectively. The calculation methods are shown in Eqs. (1) and (2):

$$Support(A \Rightarrow B) = \frac{Count(A \cap B)}{Count(D)} = P(A \cap B), \quad (1)$$

$$Confidence(A \Rightarrow B) = \frac{P(A \cap B)}{P(A)} = P(B|A). \quad (2)$$

To better evaluate the practical value of a rule, the concept of *Lift* is introduced. *Lift* refers to the ratio between the probability of an outcome occurring when the current term exists and the probability of the outcome occurring when the previous term is absent. Only rules with a *Lift* greater than 1 are considered meaningful. The calculation method is shown in Eq. (3):

$$Lift(A \Rightarrow B) = \frac{P(B|A)}{P(B)} = \frac{Support(A \Rightarrow B)}{Support(A) \times Support(B)}. \quad (3)$$

If an association rule meets both the minimum support threshold (min_sup) and the minimum confidence threshold (min_conf), it can be considered a strong association rule, indicating a strong correlation between factors $A$ and $B$. The values of min_sup and min_conf may vary depending on the specific problem, and the ultimate goal is to calculate the strong association rules that are of research significance, which are obtained through continuous experimentation. Finally, by comparing the elements in itemsets $A$ and $B$ in the obtained strong association rules, corresponding accident prevention measures can be analyzed.

The association rule mining algorithm is illustrated in Fig. 3, which consists of 2 phases. In the first phase, the data set is scanned to obtain various frequent itemsets. In the second phase, the frequent itemsets are self-joined and filtered to obtain strong association rules.

## 4 Analysis of influencing factors results

According to Heinrich (Xiao and Zhu., 2011), the main causes of unsafe human behavior or unsafe object states can be attributed to four problems: incorrect attitude, lack of technical knowledge, physical discomfort, and poor working environment.

In response to these four aspects, Heinrich proposed the 3E principle of accident prevention.

(1) Engineering: Reduce unsafe factors through the use of engineering and technological means.

(2) Education: Acquire the necessary knowledge and skills for safety production through different forms of education and training.

(3) Enforcement: Utilize necessary administrative or legal means such as relevant laws and regulations, rules, and systems to constrain people's behavior.

Heinrich's 3E principle, consisting of Engineering, Education, and Enforcement, provides a comprehensive framework for accident prevention (Xiao and Zhu, 2011). It acknowledges the major causes of unsafe behavior and object states, emphasizing the need for technological measures, knowledge acquisition, and regulatory enforcement. Even with the implementation of engineering and technological measures to reduce and control unsafe factors, it is still necessary to regulate human behavior through education, training, and mandatory measures to avoid unsafe behavior.

Considering the multidimensional attributes of traffic accidents, where multiple factors come into play, analyzing association rules helps identify similarities and differences. To effectively solve the problem from its roots, it is crucial to reduce and prevent traffic accidents. By implementing Heinrich's principles and analyzing these rules, we can develop effective strategies to reduce traffic accidents and create a safer environment for all road users.

Major traffic accidents are frequently associated with multiple influencing factors and possess multidimensional attributes (Yu et al., 2019). To enhance the effectiveness and interpretability of the rules, we excluded the two-item association rules and focused on three- and four-item rules with larger itemsets in order to identify similarities, differences, and conduct analysis.

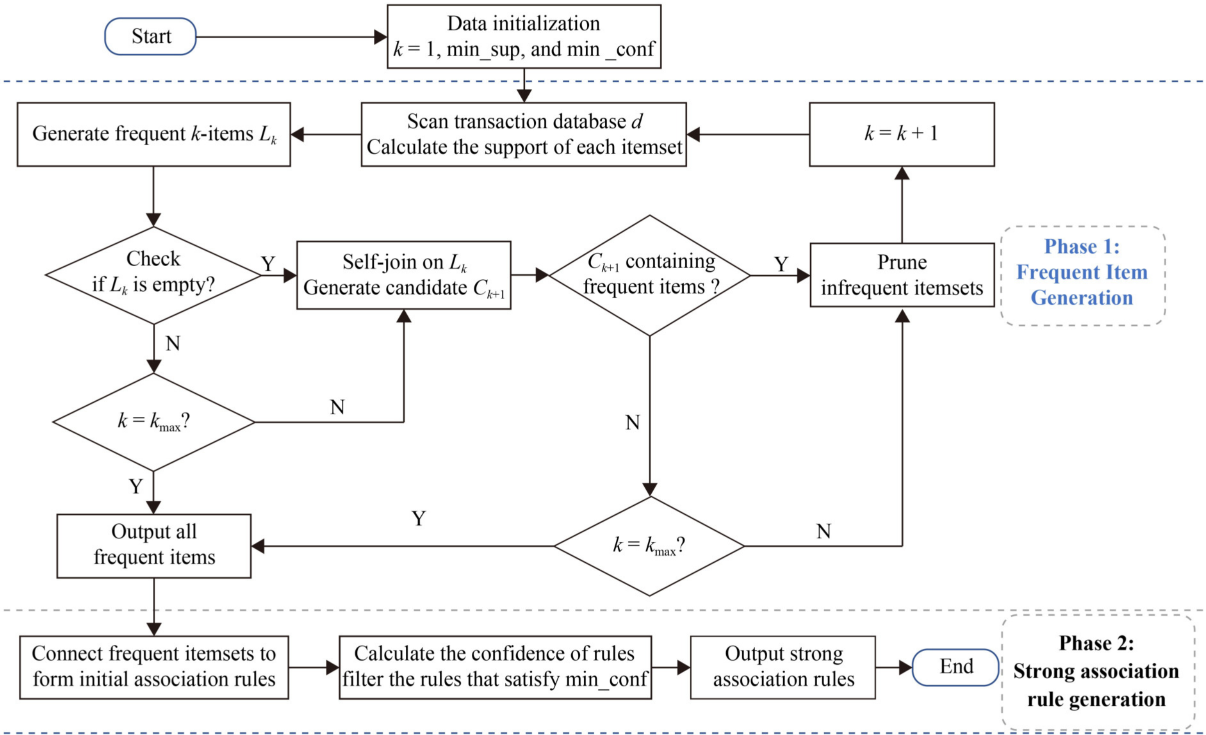


**Fig. 3** Association rule mining process diagram.

4.1 Three-item association rules

Minimum Support and Minimum Confidence are screening criteria subjectively set by the users themselves. They are used to identify relevant factors and rules that exhibit a strong correlation and close relationship (Hou, 2020). Setting the parameter thresholds too high may lead to a small number of rules, potentially overlooking important ones. Conversely, setting the thresholds too low might generate an excessive number of irrelevant rules. Therefore, the parameter thresholds presented in this paper were chosen after a process of continuous adjustment and debugging. Through experimentation and continuous adjustment of parameters, this study ultimately set the minimum support threshold to 0.1, the minimum confidence threshold to 0.5, and the number of itemsets to 3, resulting in 172 three-item association rules.

To ensure the interpretability and rationality of the rules, and to avoid the generation of irrelevant rules, the consequent of the calculated three-item rules was set to the accident types. The resulting association rules were then sorted in descending order based on their confidence levels. This configuration also validates the relationship between the accident type and its influencing factors. Table 2 displays the top ten association rules.

In the table, the antecedent of the rules represents the factors that affect the accidents, while the consequent represents the forms of accidents. The explanations of the obtained association rules are provided below.

Rule [1]: When the accident occurs in Guangxi and the involved vehicle is a truck, the probability of a collision traffic accident reaches 100%.

Rule [2]: When the accident occurs in Guangxi and the time is daytime, the probability of a collision traffic accident reaches 100%.

Rule [3]: During the first quarter of the daytime, the probability of a collision traffic accident reaches 96.1%.

Rule [4]: When the accident occurs in the fourth quarter and the involved vehicle is a truck, the probability of a collision traffic accident reaches 93.7%.

Rule [5]: When the weather is rainy and the involved vehicle is a train, the probability of a collision traffic accident reaches 93.3%.

The explanations for the following five three-item association rules are similar to those mentioned earlier and will not be repeated here.

Finally, we utilized the arulesViz and ggplot2 packages in R to visualize the obtained three-item association rules.

A grouped matrix plot displaying all 172 association rules is depicted in Fig. 4. The antecedent is represented on the horizontal axis, while the consequent is shown on the vertical axis. The association rules are represented by circles, with the size of each circle indicating the support and the color representing the lift of the rule. Similar rules are grouped together, and the entire set of rules is divided into 20 categories.

From the figure, it is evident that there is a higher number of association rules corresponding to the Region attribute being Western, the Form attribute being collision, and the Weather attribute being sunny. The rules in

**Table 2** Partial strongly three-item association rules

| No. | Antecedent | Consequent | Support | Confidence |
|---|---|---|---|---|
| 1 | {Province = Guangxi, Vehicle = truck} | {Form = collision} | 0.140 | 1.000 |
| 2 | {Province = Guangxi, Time = daytime} | {Form = collision} | 0.107 | 1.000 |
| 3 | {Quarter = 1, Time = daytime} | {Form = collision} | 0.206 | 0.961 |
| 4 | {Quarter = 4, Vehicle = truck} | {Form = collision} | 0.123 | 0.937 |
| 5 | {Weather = rainy, Vehicle = truck} | {Form = collision} | 0.115 | 0.933 |
| 6 | {Vehicle = truck, Time = daytime} | {Form = collision} | 0.331 | 0.930 |
| 7 | {Region = Western, Vehicle = truck} | {Form = collision} | 0.330 | 0.909 |
| 8 | {Weather = sunny, Vehicle = truck} | {Form = collision} | 0.330 | 0.869 |
| 9 | {Region = Western, Weather = sunny} | {Form = collision} | 0.280 | 0.829 |
| 10 | {Region = Western, Time = daytime} | {Form = collision} | 0.272 | 0.825 |

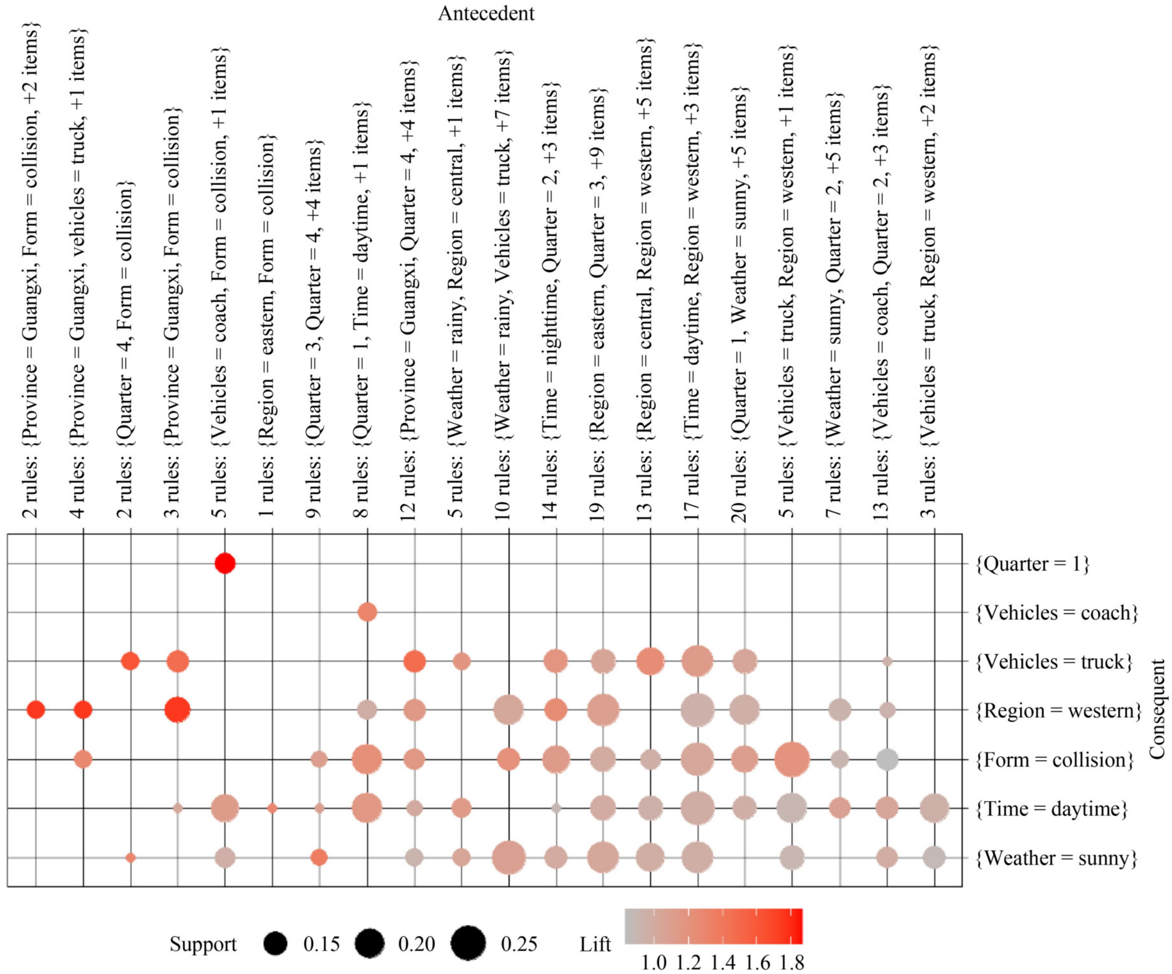


**Fig. 4** Grouped matrix plot of three-item association rules.

the upper left corner of the figure exhibit the highest lift, while the rules in the lower right corner demonstrate the highest support. The accident attributes included in the antecedent of the rules represent frequent itemsets, such as accidents occurring in Guangxi, the Form attribute being collision, and the involved vehicle being a truck. These attributes hold greater importance and should receive more attention when analyzing the factors influencing accidents. However, it is important to note that this grouped matrix plot does not reflect the correlation between different factors.

To address this limitation, a network visualization was created for the top ten three-item association rules, sorted by confidence. This visualization is displayed in Fig. 5.

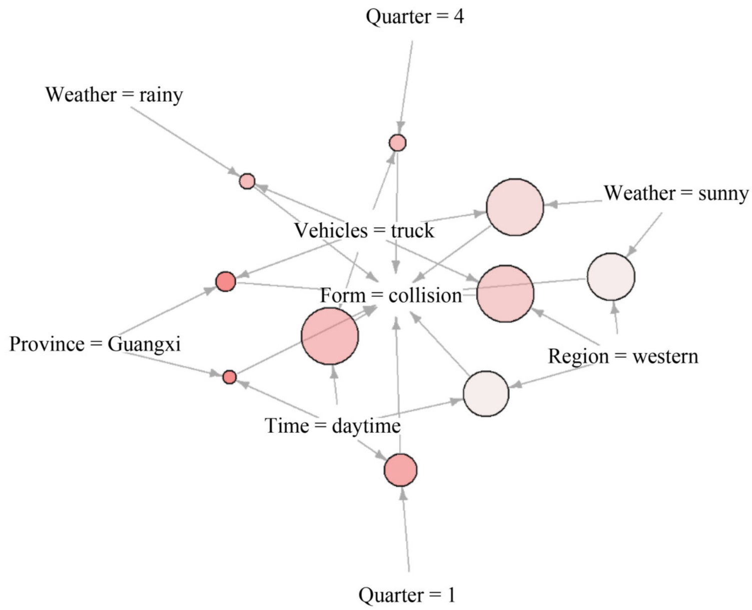


**Fig. 5** Visualized network of three-item association rules.

In this network, each circle represents a different frequent itemset, with the size indicating the support. The larger the circle, the higher the corresponding support. The support range in the figure varies from 0.107 to 0.331. The color depth of the circles represents the lift, which ranges from 1.051 to 1.274. The darker the color, the higher the corresponding lift. Additionally, arrows within the network indicate the internal relationships of the rules.

### 4.2 Four-item association rules

With a minimum support threshold of 0.1, minimum confidence threshold of 0.5, and a set of 4 items, this article calculated 52 four-item association rules. The top 10 association rules sorted by confidence in descending order are shown in Table 3 below:

Moreover, the explanations for the top five rules in the table are as follows:

Rule [1]: This rule indicates that the probability of a collision traffic accident occurring in Western during daytime and involving a truck is 96%.

Rule [2]: This rule indicates that the probability of a collision traffic accident occurring in Western during the second quarter and involving a truck is 94.1%.

Rule [3]: This rule indicates that the probability of a truck being involved in a collision traffic accident during the second quarter in Western is 94.1%.

Rule [4]: This rule indicates that the probability of a collision traffic accident occurring during the first quarter in daytime and involving a coach is 92.9%.

Rule [5]: This rule indicates that the probability of a collision traffic accident occurring in Western during the first quarter and daytime is 92.9%.

The explanations for the following five three-item association rules are similar to those mentioned earlier and will not be repeated here.

Likewise, the grouped matrix plot of the four-item association rules is displayed in Fig. 6. From the plot, it is evident that there are 20 association rules with the consequent being Western. The rules in the upper-left corner of the plot exhibit higher lift, while the rules in the lower-right corner display higher support.

Furthermore, Fig. 7 shows the network visualization of the four-item association rules. The figure demonstrates a support range between 0.107 and 0.215, and a lift range between 1.121 and 1.627. Within the association rule network shown in the figure, {Form = collision}, {Vehicle = truck}, and {Region = Western} are positioned at the center, representing the most commonly shared itemset and their strong correlation to accidents. By following the arrows, the network effectively provides insight into the interplay among different factors influencing accidents, their impact on accident occurrences, and the correlation between accidents.

### 4.3 Comparison of three-item and four-item association rule results

This paper utilizes the output results of an accident causation factor analysis model to propose relevant policy recommendations and measures aimed at reducing the frequency and severity of accidents, as well as the likelihood of major or catastrophic accidents. Moreover, this model can be used to anticipate accident severity in advance, providing data support to management departments for the development of effective accident response strategies, prompt accident handling, improved safety

**Table 3** Partial strongly four-item association rules

| No. | Antecedent | Consequent | Support | Confidence |
|---|---|---|---|---|
| 1 | {Region = Western, Vehicle = truck, Time = daytime} | {Form = collision} | 0.198 | 0.960 |
| 2 | {Region = Western, Quarter = 2, Vehicle = truck} | {Form = collision} | 0.132 | 0.941 |
| 3 | {Region = Western, Quarter = 2, Form = collision} | {Vehicle = truck} | 0.132 | 0.941 |
| 4 | {Quarter = 1, Vehicle = coach, Time = daytime} | {Form = collision} | 0.107 | 0.929 |
| 5 | {Region = Western, Quarter = 1, Time = daytime} | {Form = collision} | 0.107 | 0.929 |
| 6 | {Region = Western, Weather = sunny, Vehicle = truck} | {Form = collision} | 0.182 | 0.880 |
| 7 | {Region = Western, Form = collision, Time = nighttime} | {Vehicle = truck} | 0.132 | 0.696 |
| 8 | {Form = collision, Vehicle = truck, Time = nighttime} | {Region = Western} | 0.132 | 0.696 |
| 9 | {Form = collision, Weather = sunny, Time = nighttime} | {Region = Western} | 0.116 | 0.667 |
| 10 | {Form = collision, Weather = sunny, Time = daytime} | {Vehicle = truck} | 0.215 | 0.667 |

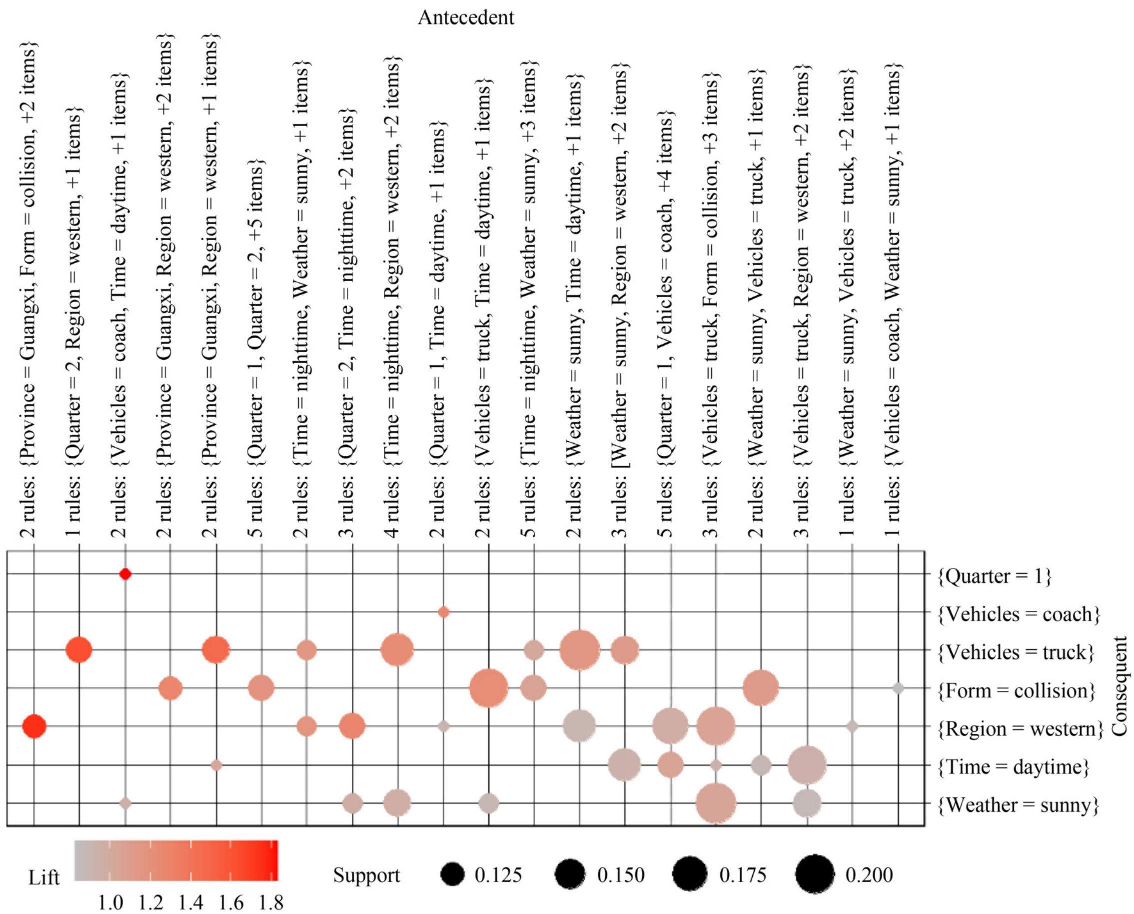


**Fig. 6** Grouped matrix plot of four-item association rules.

management efficiency, and minimization of losses.

Based on the correlation analysis of accident characteristics and different influencing factors, as well as an exploration of the causes and patterns of catastrophic accidents, the following conclusions and preventive measures are drawn:

(1) In the western region of China, exemplified by Guangxi, there is a relatively high probability of collision accidents involving trucks during transportation. This can be attributed to the challenging terrain and numerous mountainous roads present in the western regions. Thus, it is crucial to bolster transportation management in these areas, enhance road conditions, and enforce stricter penalties for hazardous driving behaviors like speeding and illegal overtaking. Additionally, installing warning signs in specific terrains can contribute to improved road safety.

(2) The majority of vehicles involved in traffic accidents

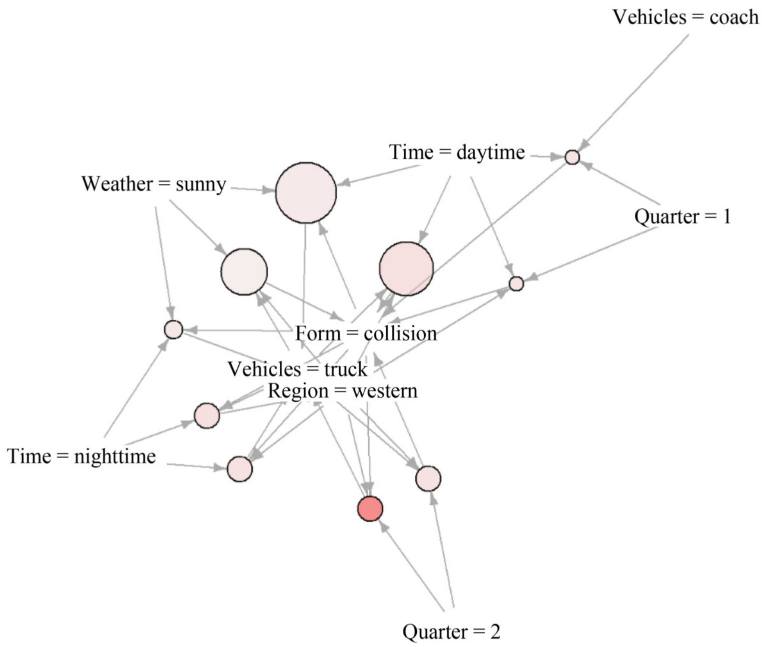


**Fig. 7** Visualized network of three-item association rules.

are trucks, and these accidents are mainly collisions. This could be attributed to the fact that truck drivers often drive continuously to complete tasks in a timely manner, resulting in driver fatigue. Additionally, the long braking distance of large trucks and inadequate following distance maintained by drivers contribute to a higher probability of collision accidents involving trucks. It is crucial to ensure that truck drivers prioritize safe and responsible driving while fulfilling their duties in order to ensure overall safety.

(3) Traffic accidents are more likely to occur during daytime hours in rainy or snowy weather. Wet road surfaces during inclement weather pose challenges for drivers attempting to stop safely in emergency situations. Furthermore, poor weather conditions affect driver visibility and increase the reaction time needed to respond to sudden events. To address this issue, it is necessary to manage road sections prone to water accumulation and uneven surfaces, educate and guide drivers, enforce strict penalties for risky behaviors such as speeding, and rigorously control vehicle speeds. Additionally, it is essential to enhance driver safety warnings by proactively communicating adverse weather conditions via text messages or radio broadcasts and providing safe driving advice.

(4) Traffic accidents are more prevalent in the first quarter of each year. One possible explanation for this trend is the occurrence of behaviors such as returning home for the Lunar New Year, traveling during the Spring Festival holiday, and returning home after the holiday. These activities may lead to drivers becoming relatively relaxed and exhibiting decreased attention and concentration while behind the wheel. Additionally, many regions experience snow-covered roads and adverse weather conditions during winter and early spring. To address this issue, traffic management departments should diligently inspect high-risk trucks and impose stricter penalties for overloading. Moreover, promoting driver safety awareness, upholding the principle of "safety first" at all times, and cautioning drivers to prioritize safe driving practices are crucial for ensuring overall safety.

## 5 Conclusions

Based on the statistical results of major road traffic accidents' impact factors, this paper employs the Apriori association rule algorithm to investigate the internal associations among these factors. The study examines the specific effects of accident forms, weather conditions, and other factors on accident occurrence. Rule analysis is also conducted under different situational conditions. The results indicate that severe traffic accidents are closely associated with factors such as Guangxi, daytime rain and snow weather, and trucks in the western region. To address each causal factor, the paper analyzes strong correlation rules and proposes corresponding preventive measures for transportation management departments at various levels. These measures include rigorous inspection of high-risk trucks, intensified penalties for dangerous driving behaviors like speeding and overloading, and the establishment of warning signs in hazardous areas. These efforts aim at prevent traffic accidents and reduce the likelihood of severe accidents. Although the algorithm used in this study has identified some meaningful rules in traffic accidents, it is important to note that the Apriori algorithm has low computational efficiency. As a result, future research should focus on improving the algorithm's computational efficiency and accuracy.

**Competing Interests** The authors declare that they have no competing interests.

## References


Agrawal R, Srikant R (1994). Fast Algorithms for Mining Association Rules. In: Proceedings of the 20th International Conference on Very Large Data Bases. Santiago: Chile, 487–499

Ait-Mlouk A, Agouti T (2019). DM-MCDA: A web-based platform for data mining and multiple criteria decision analysis: A case study on road accident. SoftwareX, 10: 100323

Çeven S, Albayrak A (2024). Traffic accident severity prediction with ensemble learning methods. Computers & Electrical Engineering, 114: 109101

Hong J, Tamakloe R, Park D (2020). Discovering insightful rules among truck crash characteristics using Apriori algorithm. Journal of Advanced Transportation, 2020(1): 1

Hou Y Q (2020). Analysis of causes of traffic accidents on different sections of expressway based on improved association rules. Dissertation for the Master's Degree. Beijing: Beijing Jiaotong University (in Chinese)

Huang D Z, Liang T T, Hu S P, Loughney S, Wang J (2023). Characteristics analysis of intercontinental sea accidents using weighted association rule mining: Evidence from the Mediterranean Sea and Black Sea. Ocean Engineering, 287: 115839

Kang L J, Li H, Li C, Xiao N, Sun H J, Buhigiro N (2021). Risk warning technologies and emergency response mechanisms in Sichuan–Xizang Railway construction. Frontiers of Engineering Management, 8(4): 582–594

Kang L J, Lu Z Y, Meng Q, Gao S, Wang F W (2019a). Maritime simulator based determination of minimum DCPA and TCPA in head-on ship-to-ship collision avoidance in confined waters. Transportmetrica A: Transport Science, 15(2): 1124–1144

Kang L J, Meng Q, Zhou C B, Gao S (2019b). How do ships pass through L-shaped turnings in the Singapore Strait? Ocean Engineering, 182: 329–342

Kuang L, Yan H, Zhu Y, Tu S, Fan X (2019). Predicting duration of traffic accidents based on cost-sensitive Bayesian network and weighted K-nearest neighbor. Journal of Intelligent Transport Systems, 23(2): 161–174

Ma X, Ding C, Luan S, Wang Y, Wang Y (2017). Prioritizing influential factors for freeway incident clearance time prediction using the gradient boosting decision trees method. IEEE Transactions on Intelligent Transportation Systems, 18(9): 2303–2310

Martín L, Baena L, Garach L, López G, de Oña J (2014). Using data mining techniques to road safety improvement in Spanish roads. Procedia: Social and Behavioral Sciences, 160: 607–614

Moslem S, Farooq D, Esztergar-Kiss D, Yaseen G, Senapati T, Deveci M (2024). A novel spherical decision-making model for measuring the separateness of preferences for drivers' behavior factors associated with road traffic accidents. Expert Systems with Applications, 238: 122318

Xiao G P, Zhu X N (2011). Traffic safety engineering. China: Railway Publishing House (in Chinese)

Yu S, Jia Y H, Sun D Y (2019). Identifying factors that influence the patterns of road crashes using association rules: A case study from Wisconsin, United States. Sustainability, 11(7): 1925

Zhang C J, He J, Bai C G, Yan X T, Wang C W, Guo Y Z (2022). Exploring relationships between months and different crash types on mountainous freeways using a combined modeling approach. Journal of Advanced Transportation, 2022(1): 1

Zhang S L, Jin H Q, Song Y, Yu W S, Sun L D (2016). Genetic study identifies CBLN4 as a novel susceptibility gene for accident proneness. Frontiers of Engineering Management, 3(1): 30–38