Comparative analysis of motor and non-motor vehicle accidents factors at day and night using association rules

Jianghui CHEN*, Jinghua Xu, Taizhou WANG
College of Transportation Engineering, Chang'an University, Xi'an, Shaanxi 710064, China;
*cjh@chd.edu.cn

ABSTRACT

In order to investigate the differences in causes of motor vehicle and non-motor vehicle accidents between day and night, reduce the losses caused by accidents, and establish the Ant Colony Algorithm (ACA) to mine the association rules between different severity levels of motor-nonmotor traffic accidents and various factors. Firstly, the association rule generation will be transformed into the Traveling Salesman Problem, and then the Ant Colony Algorithm will be utilized for solving it, aiming to enhance the quality of the rules. Secondly, directional constraints will be applied to the rules to reduce the generation of redundant rules. Lastly, an analysis was conducted on 2178 motor-nonmotor accident data. The results indicate that the proposed ACA algorithm can significantly reduce the number of invalid rules and accurately identify the correlation between potential accident factors and severity levels. Compared to daytime, there is a greater tendency for higher severity level accidents to occur at night due to environmental and non-motor vehicle driver factors. Furthermore, there are significant differences in the correlation between factors such as rain, fog, snow, driver alcohol consumption, and speeding, with the severity of accidents occurring during the night and daytime.

Keywords: Traffic safety, motor and non-motor vehicle accidents, causes of severity of accidents at day and night, association rules, ant colony algorithm

1. INTRODUCTION

Non-motor vehicles are an important component of urban transportation in China, and they are favored by short-distance travelers for their environmentally friendly and convenient characteristics. However, due to the high flexibility, vulnerability to disturbances, low cost of non-compliance, and frequent sharing of road space with motor vehicles, conflicts between non-motor vehicle riders and motor vehicles have become increasingly severe[1]. Especially with the increase in nighttime activities, there has been a rise in the occurrence of motor-nonmotor accidents and an escalation of both human and economic losses [2]. As a result, nighttime accidents have become a key issue in urban traffic management and regulation [3]. Therefore, investigating the differential impact of various potential factors on the severity of motor-nonmotor accidents between day and night can help implement targeted preventive measures to reduce accident hazards.

The methods used in the study of factors influencing motor-nonmotor traffic accidents primarily involve constructing Logistic regression models to analyze whether various study variables have a significant impact on accident severity [4-5]. A decision tree model was employed to investigate the correlation between bicycle collision severity and factors such as infrastructure, cyclists, and environment. The random forest model was utilized to identify the factors that significantly affect the severity of injuries to cyclists[6]. Some studies quantified the marginal effects of accident factors on non-motor vehicle collision risk by constructing structural equation models and binary probability models [7]. Traffic accident factors are complex and interrelated with each other. However, the aforementioned methods only consider the impact of each factor on accident severity, neglecting the inherent relationships among accident factors.To explore the relationships between accident influencing factors, association rule algorithms are introduced to discover potential patterns in the occurrence of traffic accidents. It can effectively simplify and process accident data, deducing the impact of various factors on accident occurrence and their correlations [8]. The most classic association rule mining algorithm is the Apriori algorithm, which frequently scans the entire dataset during the rule generation process, resulting in the generation of a large number of candidate sets [9].In order to avoid artificially setting thresholds of association rules, particle swarm optimization is used to search for optimal support and confidence thresholds [10]. Most of the studies on the correlation of factors affecting road traffic accidents focus on motor vehicle collisions, but there are few studies on the correlation between motor vehicle and non-motor vehicle accident factors.

International Conference on Smart Transportation and City Engineering (STCE 2023),
edited by Miroslava Mikusova, Proc. of SPIE Vol. 13018, 130183S
 · doi: 10.1117/12.3024030

To sum up, in order to compare and analyze the correlation characteristics between the different severity levels of day and night motor vehicle and non-motor vehicle traffic accidents and various factors, and improve the efficiency of rule mining, this paper constructs ant colony algorithm (Ant Colony Algorithm, ACA) to mine accident-cause association rules . In analyzing the impact of factors such as motor vehicle drivers, non-motor vehicle drivers, roads, and environmental factors on the severity of traffic accidents, a detailed analysis is conducted to explore the coupling relationships among these factors. The research can provide references for addressing the issues of motor-nonmotor traffic safety during day and night, and serve as a theoretical basis and technical support for precise accident prevention and control.

2. ANALYSIS METHODOLOGY

2.1 Principal of association rule

Association rules are used to discover the internal relationship between different attribute items in the data set. It uses three indicators of support, confidence, and lift to measure the degree of association between data item sets. Association rule is an implication like form $A => B$, where A is the causal layer, also known as Left-Hand-Side(LHS), and B is the result layer, also known as Right-Hand-Side(RHS). The $support(A,B) = P(AB)$, refers to the probability that A and B appear at the same time. The $confidence(A,B) = P(B \mid A)$ refers to the probability of the occurrence of the RHS given the premise of the LHS. The higher the confidence, the more reliable the rule. The $lift(A,B) = P(B \mid A) / P(B)$ refers to the degree of improvement in the probability of the occurrence of B given the occurrence of A. When $lift>1$, it indicates that the association rule between A and B is valid, and the higher the value of the lift, the stronger the association.

2.2 Modeling of ant colony association rules mining

The Ant Colony Algorithm (ACA) transforms the association rule mining problem into the Traveling Salesman Problem (TSP). Each factor in the accident data set is used as a node to construct an undirected graph, the support S_{ij} of the 2-item set formed by any node i, j is used as the weight of the edge, and the distance between two nodes is $d_{ij} = 1/S_{ij}$. Select m ants starting from different nodes, the transition probability for ant k is:

$$P_{ij}^{k}(t) = \begin{cases} \dfrac{\tau_{ij}^{\alpha}\eta_{ij}^{\beta}}{\sum\limits_{S \in allowed_k} \tau_{is}^{\alpha}(t)\eta_{is}^{\beta}(t)}, & j \in allowed_k \\ 0 & , otherwise \end{cases} \tag{1}$$

where $\tau_{ij}(t)$ is the concentration of pheromones on the edge between i and j at time t, setting $\tau_{ij}(0) = \tau_{\max}$. η_{ij} indicates the heuristic function that reflects the expected level between two nodes and set as $\eta_{ij}(t) = 1/d_{ij}$. α and β represent the relative importance of pheromones and heuristic factors. $allowed_k$ is the set of nodes that ant k can choose from for its next step. When the ant completes a transfer, it needs to update the local information of the path. The update rule is:

$$\tau_{ij}\left(t+1\right) = \varepsilon \cdot \tau_{ij}\left(t\right) + \left(1-\varepsilon\right) \cdot S_{ij} \tag{2}$$

where $\varepsilon \in (0,1)$ is pheromone volatile factor, which indicates the decay degree of pheromone on the path with time. After the selected node j is added to the path of the ant, it is judged whether the items on the path meet the requirements of S_{min}, and if so, the frequent itemset is extracted, otherwise, j is removed from the path. When the ant has traversed the path, it can obtain the frequent itemset L_{best} with the maximum support in this cycle. At this point, the global pheromone is updated according to the rule specified in equation (3).

$$\tau_{ij}\left(t+1\right) = \rho \cdot \tau_{ij}\left(t\right) + \left(1-\rho\right) \cdot \Delta\tau_{ij}\left(t\right) \tag{3}$$

where $\rho \in (0,1)$ is the pheromone residual factor. $\Delta\tau_{ij}(t)$ represents the pheromone increment, and $\Delta\tau_{ij}(t) = 1/L_{best}$. If the path does not contain ij, then $\Delta\tau_{ij}(t) = 0$.

2.3 Algorithm steps

Considering that the particle swarm optimization algorithm has the advantages of easy implementation and fast convergence speed, reference [10] uses it to obtain the optimal support and confidence.The former item is constrained as the accident factor, and the latter item is constrained as the accident level. The specific flow of the ant colony association rule mining algorithm is shown in Figure 1.

Step 1: Initialize the parameters of the algorithm within a given range, input the maximum number of iterations and the minimum support S_{min} and minimum confidence C_{min} calculated by the particle swarm algorithm to generate the initial population.

Step 2: Select *m* ants to traverse the undirected graph from different items.

Step 3: Calculate the transition probability to determine the next step *j* of ant *k*, and judge whether *j* meets the set minimum support and confidence after joining the path. If it is satisfied, keep the frequent itemset; if it is not satisfied, list *j* into the taboo of ant *k* table, and remove *j* from the path.

Step 4: Update the local pheromone, and judge whether all nodes have been traversed, if not, return to step 3.

Step5: Update global pheromone information, check if the maximum iteration limit is reached. If satisfied, generate frequent itemsets; otherwise, return to step 2.

Step6: Check if the frequent itemsets satisfy the constraint conditions. The antecedent constraint is the accident factor, and the consequent constraint is the accident level. If satisfied, output the association rule; otherwise, eliminate it.

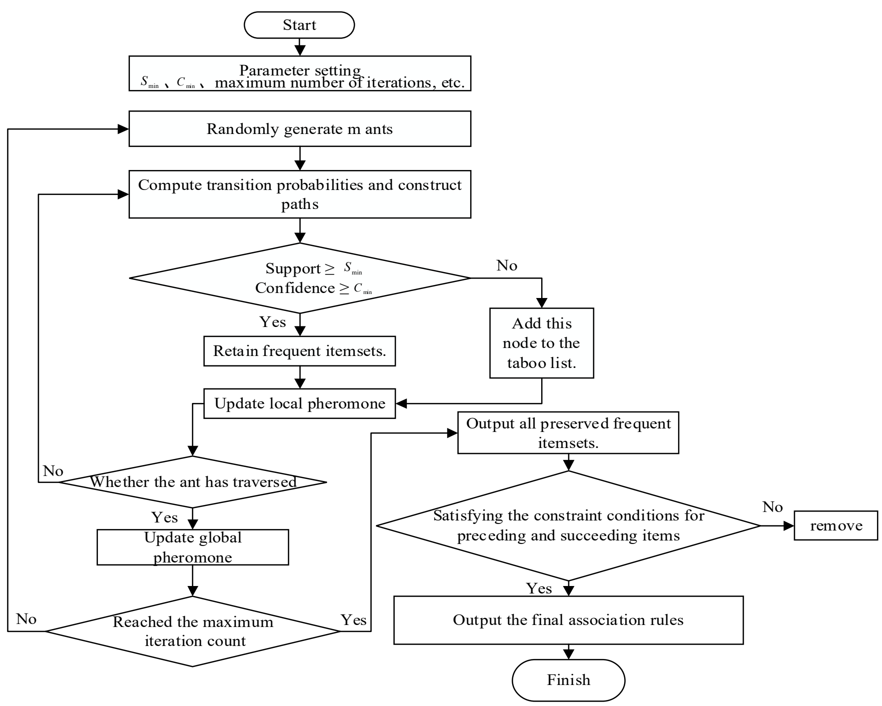


Figure 1. Ant colony association rules mining flow chart

3. DATA DESCRIPTION

The data is sourced from accident statistics compiled by a local traffic management department in China. A total of 2,178 incidents of motor and non-motor accidents were selected. The data statistics are shown in Table 1. According to the peak characteristics of local residents' travel activities, it is set to take 6:00-19:00 during the daytime and 19:00-6:00 the next day at night. According to the classification standard of the Ministry of Public Security[11], the severity of the accident is divided into slight, general, serious, and extraordinarily serious accidents. Since the proportion of extremely serious accidents is too small, it is difficult to form rules, so the extremely serious accidents are merged into serious accidents. Table 2 shows the distribution statistics of accident severity during day and night. It can be observed that the occurrence rate of major accidents during the night is 11.4% higher than during the day. Therefore, it is necessary to analyze the differences in accident severity factors between day and night.

Table 1. Statistics of accident influencing factors

Item	Related factors	Description	Item	Related factors	Description
Non-motor vehicle driver	Age	≤20	Road	Lane	1 lane
		(20,40]			2 lane
		(40,60]			3 lane
		>60			4 lane
	Gender	Female			5 lane
		Male			≥6 lane
Motor vehicle driver	Age	≤20		Accident location	Driveway access
		(20,40]			Road entrance/exit
		(40,60]			Intersection
		>60		Road direction	One-way
	Gender	Female			Two-way, not divided
		Male			Two-way, divided, unprotected median
	Alcohol Involve	Alcohol			Two-way, divided, protected median
		No alcohol		Slope	No slope
	Speed Involve	Speeding			Uphill
		No speeding			Downhill
Environment	Weather conditions	Sunny		Surface	Dry
		Cloudy			Wet
		Rain|fog|snow	Accident Severity		Slight
	Light condition	Light			General
		Dusk			
		No light			Serious

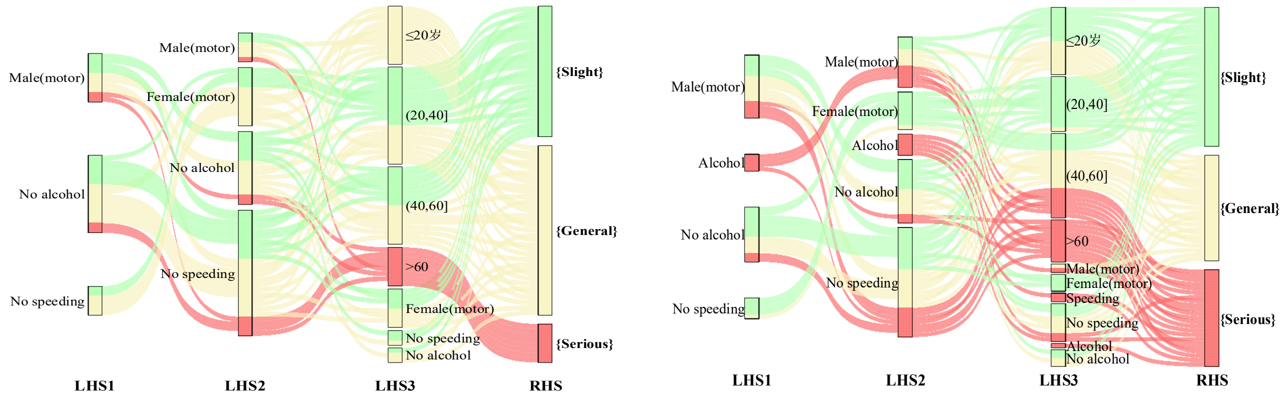
Table 2. Distribution of accident severity at day and night

Time	Slight(%)	General(%)	Serious(%)
Day	36.2	55.7	8.1
Night	31.4	49.1	19.5

4. ANALYSIS RESULTS

4.1 Differences analysis of single-dimensional association rules

1) Motor vehicle driver dimension.

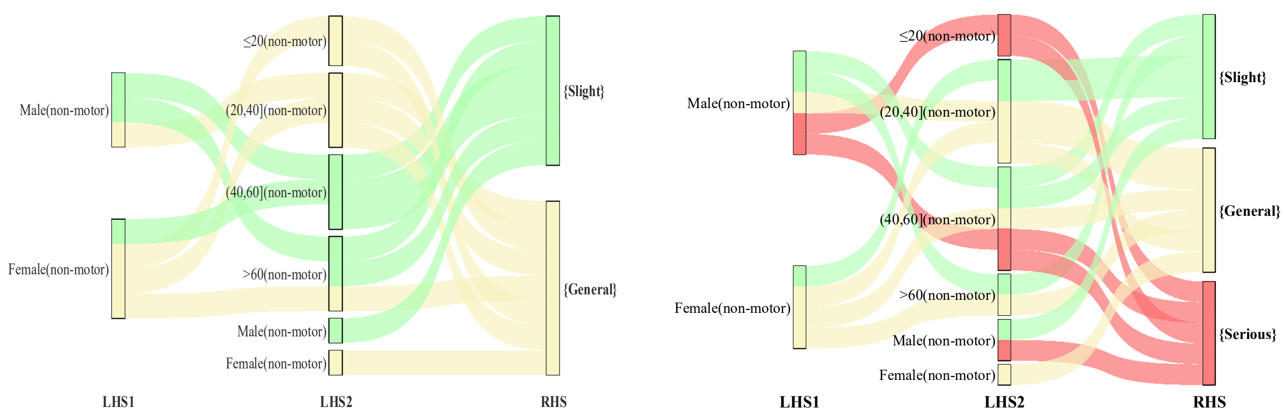


(a) Day (b) Night

Figure 2. Path diagram of association rules for motor vehicle drivers

Filter the rules that only involve motor vehicle drivers and have lift than 1 and draw an association path diagram, as shown in Figure 2. The last column in Figure 2 is the RHS, while the remaining columns are the LHS. The line represents a regular path, with green, yellow, and red colors indicating rules with slight, general, and serious accidents.The severity of accidents caused by motor vehicle drivers drinking alcohol or speeding at night tends to be serious, while the absence of effective rules for drinking alcohol or speeding during the day may be due to the difficulty of drivers avoiding unexpected situations during drinking or speeding at night, resulting in an aggravation of the severity of the accident. In addition, male motor vehicles aged 40, 60 years old tend to have slight and general levels of accidents during the day without drinking alcohol or speeding, while accidents at night tend to be severe.

2) Non-motor vehicle driver dimension



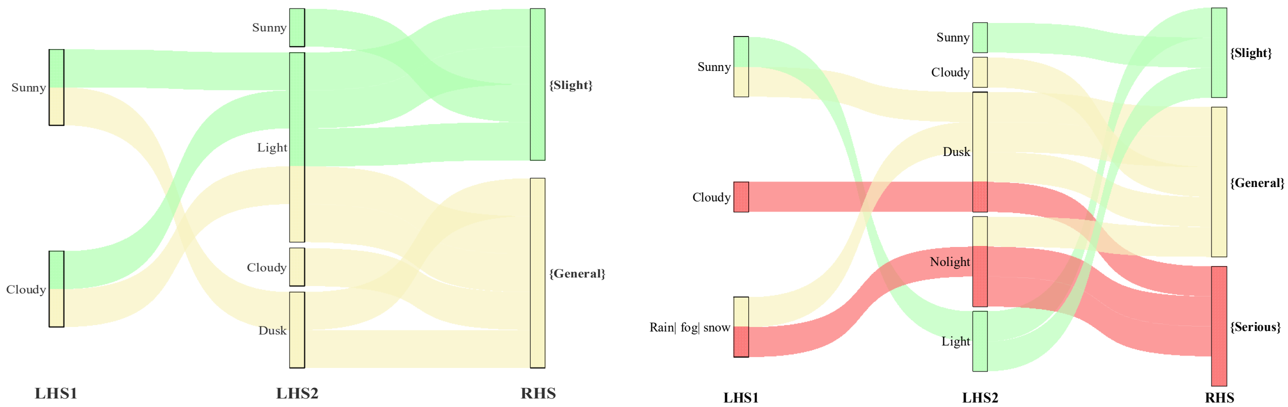
(a) Day (b) Night

Figure 3. Path diagram of association rules for non-motor vehicle drivers

Only select rules that are related to non-motor vehicle drivers and have a lift than 1, and draw an associated path diagram as shown in Figure 3. From Figure 3, it can be observed that male non-motor vehicle drivers aged (40, 60] and >60 years old tend to have slight levels of traffic accidents both during the day and at night. However, female non-motor vehicle drivers over 60 years old show a tendency towards increased severity of accidents.As shown in Figure 3, no effective rules were generated during the daytime where the LHS is non-motor vehicle drivers and the RHS is serious accidents. However, corresponding rules were observed during the nighttime. This indicates that compared to daytime, there is a greater tendency for serious-level accidents to occur during nighttime due to non-motor vehicle driver factors.

Non-motor vehicle drivers aged ≤20 and (40, 60] years old tend to have accidents of slight and general severity during the daytime, while there is a possibility of accidents with a serious severity during the nighttime. Among women aged (40, 60], accidents during the daytime are primarily of slight severity, while there is a tendency for an increase in accident severity during the nighttime.

3) Environment dimension

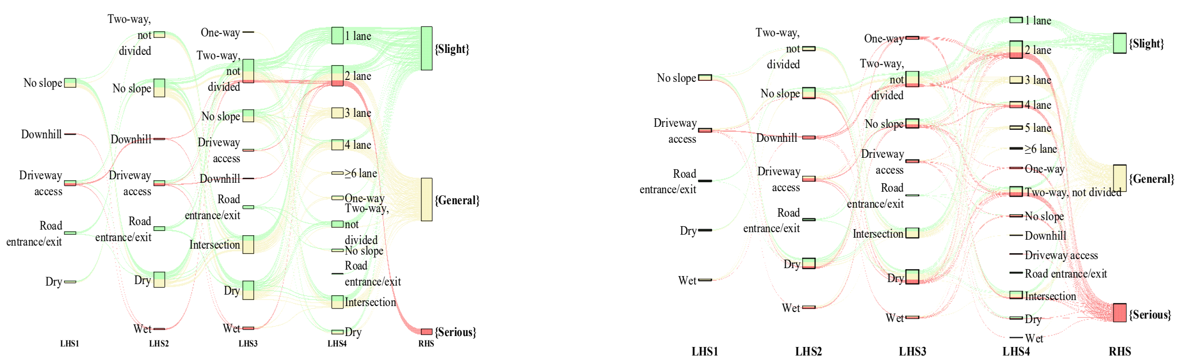


(a) Day (b) Night

Figure 4. Path diagram of association rules for environmental factors

Only select rules that are related to environmental factors and have a lift than 1, and draw an associated path diagram as shown in Figure 4.It can be observed that in sunny and light environments, the majority of traffic accidents occurring during both daytime and nighttime are of slight severity.However, in dusk conditions, there is a tendency for an increased severity of traffic accidents during both daytime and nighttime. As shown in Figure 4, no effective rules were generated during the daytime where the LHS is environmental factors and the RHS is serious accidents. However, corresponding rules were observed during the nighttime. This indicates that in cloudy, rainy, foggy, snowy, and dusk environments, there is a much higher tendency for serious-level accidents to occur during the nighttime compared to the daytime.

4) Road Dimension



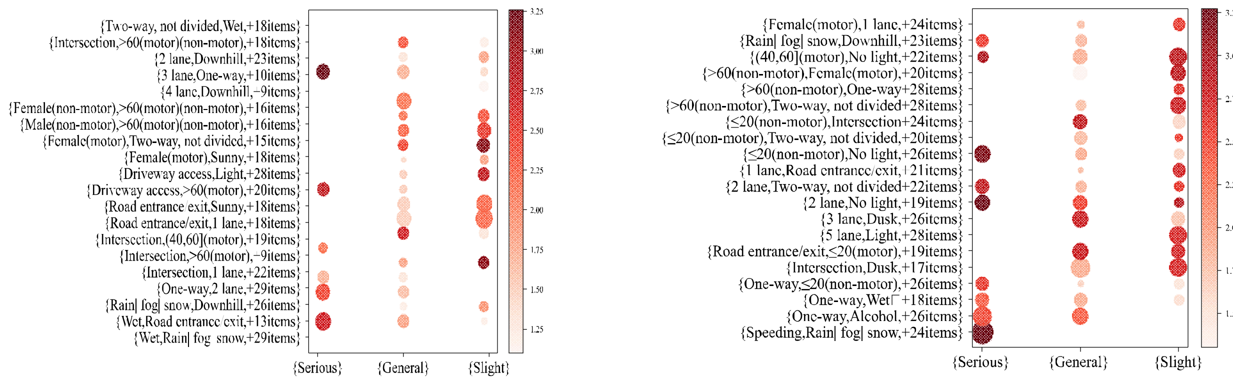
(a) Day (b) Night

Figure 5. Path diagram of association rules for road factors

Only select rules that are related to road factors and have a lift than 1, and draw an associated path diagram as shown in Figure 5.From Figure 5, it can be observed that the severity of traffic accidents during both daytime and nighttime is mostly slight in road segments with entrances and exits and no slopes.Under conditions of undivided or slippery four-lane roads, the severity of accidents that occur during the nighttime is higher than during the daytime. This may be due to the influence of slippery road conditions affecting braking distance and leading to accidents. Additionally, there is a tendency for an increased severity of accidents at intersections and downhill road segments during nighttime compared to daytime.

4.2 Difference analysis of multi-dimensional association rules

Taking into account the impact of factors such as motor vehicle drivers, non-motor vehicle drivers, environment, and road conditions on the severity of traffic accidents, select associated rules with *lift>1*. In order to better explore the relationship between various factors, the rules with *lift>1* are drawn into a bubble diagram.As shown in Figure 6, the horizontal axis represents the regulated post-items, and the vertical axis represents the pre-regulated items. In the "*A,B+n* items" notation, *A* and *B* are the most frequent items in the accident group, and "*+n* items" indicates that the group also includes n additional items.The size of the bubbles in the figure indicates the size of the cumulative support, the color of the bubbles indicates the lift, and the darker the color, the larger the value.



(a) Day (b) Night

Figure 6. Grouped balloon plot of rules at day and night

According to Figure 6, slight and general level traffic accidents occur more frequently at intersections, undivided two-way roads, road entrances and exits, and one-way roads, both during the day and at night. Among non-motorized vehicle drivers over the age of 60, there is a tendency for slight and general level accidents to occur primarily at intersections during the day, while at night, they are more likely to occur on undivided one-way and two-way roads.Serious-level traffic accidents are prone to occur during both daytime and nighttime in conditions such as rain, fog, snow, slippery roads, one-way roads, downhill slopes, and fewer lanes. However, during the nighttime, the likelihood of serious-level accidents is much higher due to insufficient lighting and the increased possibility of drunk driving and speeding by motorized vehicle drivers compared to the daytime.

5. CONCLUSION

In this paper, the Ant colony optimization algorithms is established to mine the association rules of day and night traffic accidents. The main conclusions drawn are:Motor vehicle drivers over the age of 60 are prone to serious accidents during the day and night, while non motor vehicle drivers under the age of 20 are more prone to major accidents at night without light than during the day.There is a strong correlation between serious accidents during the day and sections with downhill and fewer lanes, while there is a strong correlation between nighttime and adverse weather; In addition, the tendency for serious level accidents caused by motor vehicle drivers drinking alcohol or speeding at night is higher than during the day.

REFERENCES

[1] Qian, Y., Sun, Q., Fei, G., et al. Riding behavior and electric bike traffic crashes: A Chinese case-control study[J]. Traffic injury prevention, 2020, 21(1): 24-28. https://doi.org/10.1080/15389588.2019.1696963

[2] HE, Q., MA, J.F., LI, Y., et al. A Study of Characteristics of Road Accidents Involve Electric Bicycles Based on 122 Emergency Number[J]. Journal of Transport Information and Safety, 2018, 36(5): 132-138. DOI:10.3963/j.issn.1674-4861.2018.05.017.

[3] JI, X.F., ZHU, G. Comparative analysis on causes of severity of traffic accidents at day and night in megacities : taking Shenzhen as example[J]. Journal of Safety Science and Technology, 2020, 16(2): 142-148. DOI:10.11731/j.issn.1673-193x.2020.02.023.

[4] Hertach P, Uhr A, Niemann S, et al. Characteristics of single-vehicle crashes with e-bikes in Switzerland[J]. Accident Analysis & Prevention, 2018, 117: 232-238. https://doi.org/10.1016/j.aap.2018.04.021.
[5] LIN, Q.F., DENG, Y.C., HU. J.H. Logistic Regression Analysis of Driver's Fault and Accident Severity in Motor-Bicycle Accidents[J]. Safety and Environmental Engineering, 2019, 26(5): 187-193. DOI:10.13578/j.cnki.issn.1671-1556.2019.05.028.
[6] LI, Y.S., ZHANG, X., WANG, W.J, et al. Factors Affecting Electric Bicycle Rider Injury in Accident Based on Random Forest Model[J]. Journal of Transportation Systems Engineering and Information Technology, 2021, 21(1): 196-200. DOI:10.16097/j.cnki.1009-6744.2021.01.030.
[7] Guo, Y.Y., Zhou, J.B., Wu, Y., et al. Evaluation of Factors Affecting E-Bike Involved Crash and E-Bike License Plate Use in China Using a Bivariate Probit Model[J]. Journal of Advanced Transportation, 2017: 12. https://doi.org/10.1155/2017/2142659
[8] Weng, J., Zhu, J.Z., Yan, X., et al. Investigation of work zone crash casualty patterns using association rules[J]. Accident Analysis & Prevention, 2016, 92: 43-52. https://doi.org/10.1016/j.aap.2016.03.017
[9] Weng, J., Li, G. Exploring shipping accident contributory factors using association rules[J]. Journal of Transportation Safety & Security, 2019, 11(1): 36-57. https://doi.org/10.1080/19439962.2017.1341440
[10] XU, J.H, LI, Y., ZHANG, Y.T. An algorithm to mine the association rules among causes of accident severity at intersections[J]. Journal of Safety and Environment, 2022, 22(3): 1412-1420. DOI:10.13637/j.issn.1009-6094.2021.1553.
[11] YUAN, Z.Z., LOU, C., YANG, Y. Analysis of highway traffic accidents causes under time differences[J]. JOURNAL OF BEIJING JIAOTONG UNIVERSITY, 2021, 45(3): 1-7. DOI:10.11860/j.issn.1673-0291.20200120.