Research on Traffic Safety Risk Identification of Highway Tunnels Based on Apriori

Mimi Zhang[1]

1779262717@qq.com

Nanchang Institute of Technology, Nanchang 330013, China.[1]

Abstract. The traffic accident data onto highway tunnels are statistically analyzed. The improved Apriori algorithm is applied to extract the frequent item set and set the support degree threshold. And the association rule sets of different types of traffic accident characteristic attribute are obtained, the main risk sources and correlation relationships affecting the traffic safety of highway tunnels are found out. The technical support and rationalized prevention advice are provided to improve the safety management of highway tunnel operation, strengthen the identification of accident risks in tunnels, and reduce the losses caused by tunnel accidents to traffic efficiency and personal and property safety.

Keywords: Tunnel; Association rule mining algorithm; Traffic safety risk identification

1 INTRODUCTION

In recent years, China's highway construction has made remarkable achievements. The highway tunnel has a special spatial structure, so the probability of traffic accidents is much higher than that of other sections. The traffics safety risk identification of highway tunnel has received more attention. Highway tunnel traffic safety risk identification is an important means to identify and predict tunnel traffic safety risks, which can effectively to improve the traffic safety management level (Wen, 2022)[1]. The research on tunnel traffic safety risk identification mainly focuses on the existing traffics safety identification technologies such as artificial intelligence, machine vision, pattern recognition, which can effectively to improve the traffic safety management level. However, due to the large amount of data and the high dimension of features, the existing identification technology is difficult to meet the requirements of highway tunnel traffic safety risk identification. The purpose of this study is to explore the highway tunnel traffic safety risk identification technology based on the improved Apriori algorithm, so as to better identify the highway tunnel traffic safety risk and provide technical support for traffic safety management (Jia, 2023)[2].

2 RESEARCH STATUS AT HOME AND ABROAD

The risk factor identification method of highway tunnels traffic safety is mainly based on the risk factor identification of historical accident data, such as vehicles, drivers, roads, tunnel environment and management. The identification of traffic safety risk factors of expressway tunnels is the basis of traffic safety risk analysis, and scholars at home and abroad have also

ICBBEM 2023, May 19-21, Hangzhou, People's Republic of China

DOI 10.4108/eai.19-5-2023.2334404

made many research achievements on tunnel safety risk identification. In 2012, scholars such as Wu HB (Wu, 2012)[3] proposed a quantitative analysis method, identifying that driver heart rate increment is a major factor in the high incidence of tunnel group accidents, and that driver heart rate increment is greater at the tunnel entrance than at the exit. In 2013, scholars such as Fang Shouen (Fang, 2013)[4] proposed a quantitative research method for driving behavior in tunnel groups, identifying that the driver's heart rate growth rate is the main indicator of traffic accidents in tunnel groups, with the largest heart rate increase at the entrance of the tunnel, followed by the exit. In 2020, Pervez and other (Pervez, 2020)[5] scholars showed that the illegal driving behavior of drivers is one of the important factors leading to traffic accidents in tunnel section through statistical analysis of traffic accidents in tunnel groups, and rear-end collision are the most important type of accident. In 2022, Su Baofeng and other (Su, 2022)[6] scholars designed a framework of automatic driving and road cooperation in the tunnel area based on 5G technology and created a joint early-warning mechanism for multi-source signal to the tunnel area, so as to change the traditional passive response to accidents into active prevention and avoid tunnel risks. In 2022, Liu Songrong and other (Liu, 2022)[7] scholars found that the longer the tunnel is, the more accidents will occur, but the fewer accidents will occur per unit mileage. It is speculated that the driver gradually adapts to the driving environment in the tunnel in the extra-long tunnel.

3 ANALYSIS OF TRAFFIC SAFETY CHAR ACTERISTICS OF EXPRESSWAY TUNNELS

The analysis of traffic safety characteristics of expressway tunnels is a comprehensive investigation and analysis of expressway tunnels from the perspective of traffic safety. Abnormality of traffic flows can easily lead to traffic accidents. Investigation into the safety characteristics of traffic flows is the basis of traffic safety mechanism research. Affected by the "black hole effect" and "white hole effect", the occurrence of traffic accidents is regular and repetitive. The study of tunnel brightness changes is conducive to scientific prevention of highway tunnel traffic safety risks (Koh, 2004) 8].

3.1 Tunnel Structure Analysis

The highway tunnel is composed of main structure and auxiliary structure. The main structure is an artificial permanent building built to maintain the stability of the mountain and the safety of driving, including the tunnel body lining and the tunnel portal structure. When the mountain may collapse and rock fall, the tunnel body shall be extended, or the open cut tunnel shall be built. The structure type of the tunnel portal is mainly determined by the stability, ventilation, lighting, and environmental conditions of the mountain. Auxiliary structure refers to other buildings built for operation management, water supply and drainage, power supply, ventilation, lighting, safety and other purposes except the main structure.

3.2 Analysis of Traffic Flow Characteristics

Traffic flow refers to vehicles or pedestrians moving in a certain direction. The road is composed of motorway, bicycle lane and sidewalk. On continuous roads, many vehicles and pedestrians form road traffic flow like a fluid. Traffic flow characteristic refers to the sum of quantitative or qualitative descriptions of the change rules of the flow of people and vehicles and their interrelationships under different conditions. This paper mainly analyzes three aspects: traffic volume, speed, and headway.

The traffic volume characteristic analysis generally refers to the analysis of the time distribution, spatial distribution or composition characteristics of the traffic volume. For this paper, it mainly focuses on the spatial characteristics of the tunnel, so it is analyzed from the spatial distribution of the tunnel traffic volume.

The speed characteristic analysis reflects the concentration and distribution of traffic flow operation and is the basic work of tunnel traffic safety risk analysis. Therefore, the tunnel speed characteristics are analyzed to grasp the change trend of traffic flow.

The light difference inside and outside the tunnel is large, and the entrance and exit are easy to produce black holes, black frames, white holes, and adaptation lag. The driver needs a certain amount of adaptation time to drive normally. In order to ensure that the driver has a good visual transition at the entrances and exits at both ends of the tunnel, it is necessary to fully consider the human eye adaptation and dark adaptation.

3.3 Analysis of Traffic Accident Characteristics

3.3.1 Time Characteristics

The number of accidents in the tunnel is greatly affected by time. The "black hole effect" when vehicles enter the tunnel in the daytime and the "white hole effect" when leaving the tunnel are important factors that cause traffic accidents. Figure 1 shows the statistical distribution of daytime and nighttime accidents in tunnels in a province from 2019 to 2021. It can be seen from the figure that the frequency of accidents in the daytime is significantly higher than that at night, and nearly 9 times in 2020 and 2021. There are two main reasons: first, the statistical time of the day includes the rush hour, the traffic flow is large, and the probability of tunnel accidents is higher than other periods; Second, the daytime driving "black hole effect" and "white hole effect" is stronger, the difference in brightness inside and outside the tunnel, more prone to traffic accidents.

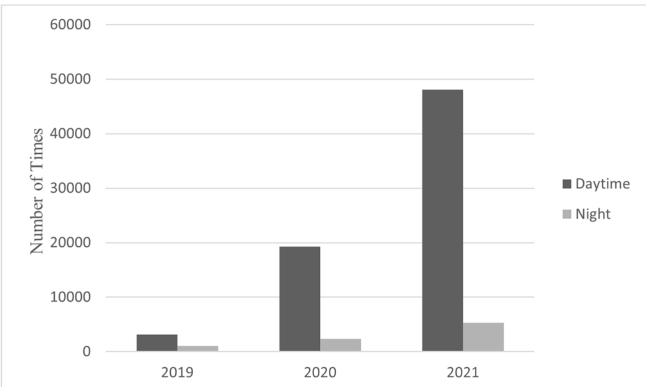

Figure 1: Accident statistics in the tunnel during the day and night

3.3.2 Accident Type Characteristics

In this paper, accidents in tunnels are classified into six categories: rear-end collision, vehicle fire, rollover, scraping, collision with objects, and others. As can be seen from Figure 2, the most frequent accidents are caused by collision with objects, including animals on the road, roadside fixtures, and objects thrown by vehicles in front of them, accounting for 61.77% of the total types of accidents. The second category of accidents is rear-end collision accidents, accounting for 33.25% of the total number of accidents. The lateral space in the lane is limited, and the driver drives at a faster speed. In an emergency, if the brake is applied too late, it may hit the tunnel wall or rear end. In addition, the black hole effect entering the tunnel and the white hole effect leaving the tunnel significantly reduce the visual effect, leading to tunnel rear-end collision.

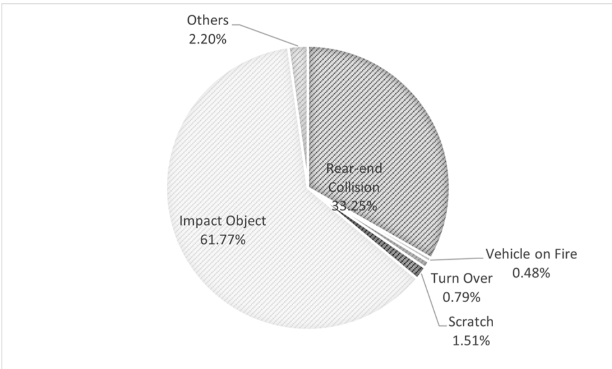

Figure 2: Distribution of accident types

4 IDENTIFICATION MODEL OF SAFETY RISK FACTORS

4.1 Apriori Association Rule Mining AlgorithmApriori

Association rule mining is one of the most active research methods in data mining proposed by Agrawal et al. in 1993 for analyzing shopping basket problems. The aim is to find the association rules between different items in the transaction database. Apriori algorithm is the earliest and most classical association rule mining algorithm. Its core idea is joining and pruning, using iterations of layer-by-layer search to find the relationship of itemsets in the database and form association rules. The concept of itemset in the algorithm is the set of items, the set including K items is K itemset, and the frequency of itemset is the number of transactions containing itemset. If an itemset satisfies the minimum support, it is called a frequent itemset. In this paper, the association rule mining algorithm of Apriori is used to identify the traffic safety risk factors of highway tunnel clusters. Association rule mining is one of the most active research methods in data mining proposed by Agrawal et al. in 1993 for analyzing shopping basket problems. Its purpose is to find the association rules between different goods in the transaction database. Apriori algorithm is the earliest and most classical association rule mining algorithm. Its core idea is joining and pruning, using iterations of layer-by-layer search to find the relationship of itemsets in the database and form association rules. The concept of itemset in the algorithm is the set of items, the set including K items is K itemset, and the frequency of itemset is the number of transactions containing itemset. If an itemset satisfies the minimum support, it is called a frequent itemset. In this paper, the association rule mining algorithm of Apriori is used to identify the traffic safety risk factors of highway tunnel clusters.

4.2 Validity Test of Association Rules

Because the rules are not necessarily valid, it is necessary to test the validity of the obtained association rules. There are two commonly used test methods: direct test and indirect verification test. The direct test method is to test the p/n significance level of each association rule. The indirect verification test method is tested by hypothetical experimental data and validation data.

4.3 An Improved Association Rule Mining Algorithm Based on Apriori Algorithm

In Apriori algorithm, if the minimum support threshold is set small, it will lead to more frequent itemsets and generate many useless association rules. The congress omitted valuable association rules. The setting of the minimum support threshold depends on the subjective judgment of scholars, so it is necessary to reduce the impact of subjective judgment. The Apriori algorithm is improved by the association rule algorithm of Top-K multidimensional and multi-level association rules. Its core is still connection and pruning. However, in pruning, the border minimum weighted support (border-wsup) is used, and the analytic hierarchy process (AHP) is used to calculate the traffic safety risk factors, determine the importance coefficient, and finally test the association rules. Let $I=\{i_1, i_2,..., i_n\}$ be the collection of all items, and $D=\{d_1, d_2,..., d_m\}$ be the transaction database.

(1)Initialization parameters: $k=1$, $L=Ø$, $C_1=Ø$.

(2)Use AHP to calculate the weight of in, and get the weight set of items $W=\{w_1, w_2,......, w_n\}$.

(3)Scan the database D, generate the first-order candidate frequency complex item set C_1, calculate the weighted support of the items in C_1, and arrange them in descending order.

(4)If K is less than the number of items in C_1, the value of border-wsup is 0; Otherwise, border-wsup is assigned to the weighted support of the Kth place in C1, and the itemset in L_1 with less support than border-wsup is deleted and stored in itemsets-array.

(5)According to L_{i-1}, Apriori algorithm is used to connect and prune to produce C_i.

(6)Generate frequent K itemset L_k.

(7) Repeat the above steps until $L_m=Ø$ and output the final association rule.

5 MODEL APPLICATION

In order to identify the traffic safety risk factors of the expressway tunnel, analyze the traffic accident data of the Jiuquling Tunnel of the Taichun-Jinjiang Expressway, use the weighted association rule algorithm to mine the association rules of the tunnel traffic safety risk factors, identify the causes of the safety risk, and study the accident mechanism, providing a strong theoretical support for the tunnel traffic safety risk analysis. Based on the above purposes, the identification of tunnel traffic safety risk factors mainly includes data conversion, weight calculation, mining association rules and other steps.

5.1 Data Handling

The attribute information of accident data consists of three parts: first, collision record data; Second, data of accident participants; The third is the data of the injured. Before the identification of traffic safety risk factors, the data needs to be processed. The continuous variables are discretized, and the classified variables are assigned values according to the categories, and the weight of traffic accident attributes is determined by the analytic hierarchy process, as shown in Table 1.

5.2 Association Rule Mining

The improved Apriori association rule mining algorithm is used to mine, analyze, and calculate the association rules of traffic safety risk factors in Jiuquling Tunnel of Taiwan-Jinzhou Expressway, and finally 100 association rules are obtained. The top 10 association rules of highway tunnel traffic safety risk factors are shown in Table 2.

5.3 Result Analysis

According to the high support association rules to represent the rules with high frequency of occurrence of corresponding frequent item sets, the following conclusions can be drawn: wet road surface is easy to make the vehicle in unsafe speed state (Yang, 2019)[9], and the support of association rules for unsafe speed is high, which means that in highway tunnels, unsafe speed is more likely to cause accidents, especially rear-end accidents; In the highway tunnel, it is easy to cause vehicle rear-end accident when going straight; Traffic accidents often occur in the daytime, because people's travel needs often occur in the daytime. To sum up, unsafe speed and weather factors are the key factors of highway tunnel traffic accidents. Secondly, the rear-end collision of straight vehicles is the main type of highway tunnel traffic accidents (Yasmin, 2014) [10].

Table 1: Weight of traffic accident attribute

Attribute Category	Attribute Name	Value Range	Weight(10^{-2})
Accident Consequence	Severity	Death S1	8.86
		Serious Injury S2	4.48
		slight Wound S3	2.77
		Property Damage Only S4	1.78
Attribute Category	Attribute Name	Value Range	Weight(10^{-2})
Accident Consequence	Injury	I1(0)	0.1
		I2(1~5)	3.6
		I3(6~10)	5.78
Accident influencing factors	Weather	Clear Day W1	1.22
		Overcast Sky W2	1.68
		Rainy Day W3	4.01

		Snowy Day W4	3.31
		Greasy Weather W5	2.69
	Accident cause	Drunken Driving C1	2.61
		Emergencies C2	3.12
		Unsafe Speed C3	2.62
		Unsafe Distance C4	3.01
		Unsafe Lane Change C5	2.89
	Collision type	Vehicle Collision T1	5.41
		Vehicle Scratch T2	2.13
		Rear End of Vehicle T3	1.66
		Impact Fixture T4	1.82
	Road surface	Dry R1	1.02
		Moist R2	6.13
		Slippery (Muddy or Oily) R3	3.87
	light	Daytime L1	1.98
		Evening L2	1.56
	sexual distinction	M	0.55
		F	0.55
	age	A1[0,18]	1.31
		A2(18,40]	1.97
		A3(40,60]	2.58
		A4(60,99]	0.77
	Vehicle status before collision	Cease CS1	2.68
		Craspedodrome CS2	2.69
		Wheel CS3	2.13
		Change Speed CS4	2.66
		Lane Change CS5	2.76

Table 2: Mining association rules with the top 10 weighted support

Order	Weighted Association Rules	Weighted Support	Confidence	Lift
1	{Collision.state=CS2, cause=C3}⇒{type=T3}	0.0212	0.68	1.75
2	{ cause=C3}⇒{type=T3}	0.0207	0.78	2.10
3	{ cause=C3}⇒{injury=I2}	0.0191	0.78	1.97
4	{ road=R2}⇒{ cause=C3}	0.0184	0.68	0.66
5	{ cause=C3}⇒{severity=T3}	0.0181	0.71	1.71
6	{ gender=M}⇒{ cause=C3}	0.0175	0.71	1.60
7	{ Collision.state=CS2}⇒{ injury=I1}	0.0155	0.81	1.55
8	{ light=L1}⇒{ cause=C3}	0.0146	0.72	1.79
9	{ cause=C3, road=R1}⇒{type=T3}	0.0141	0.95	2.71
10	{ road=R1}⇒{ Collision.state=CS2}	0.0114	0.82	1.68

6 CONCLUSIONS

This paper combines risk theory and traffic conflict theory, through investigating and analyzing the traffic flow characteristics, illumination change characteristics and traffic accident characteristics of expressway tunnels, uses mathematical statistics, association rule mining methods and traffic conflict theory to identify the traffic safety characteristics of expressway tunnels and traffic safety risk factors of expressway tunnels. Association rule mining found that the higher the conditional probability of occurrence of corresponding frequent itemsets, the higher the probability of occurrence of subsequent items. In the prevention and control of risks, the occurrence of such subsequent items should be prevented or avoided as much as possible. For highway tunnels, rear-end collision is the successor of the highest confidence, which is the key factor affecting the traffic safety of highway tunnels.

REFERENCES

[1] Wen Q.J. (2022)Research on Expected Value of Social Risk of Highway Tunnel Based on Traffic Operation State. Technology of Highway and Transport, 38:118-122. https://orcid.org/10.13607/j.cnki.gljt.2022.05.018.

[2] Jia Y.Z.(2023)Characteristics and Construction Countermeasures of Rock Burst in Tunnel of Grand Canyon Expressway. Technology and Market, 30: 72-75. https://orcid.org/ 10.3969/j.issn.1006-8554.2023.02.018.

[3] Wu H.B., Fang S.E., Liao J.H., et al.(2012)Design Consistency Evaluation Methods for Freeway Long Tunnels in China. CICTP, 2350-2361. https://orcid.org/ 10.1061/9780784412442.238.

[4] Fang S.E., Wu H.B., Liao J.H. et al. (2013) Safety Evaluation of Freeway Tunnel Groups in Mountainous Areas. Journal of Tongji University (Natural Science), 41(05):693-699.https://orcid.org/ 10.3969/j.jssn.0253-374x.2013.05.010.

[5] Pervez, A, Huang, H, Han, C, Wang, J, Lia, Y. (2020) Revisiting freeway single tunnel crash characteristics analysis: a six-zone analytic approach. Accid. Anal. Prev.142,105542.https://orcid.org/ 10.1016/j.aap.2020.105542.

[6] Su B.F., Hu J.B. (2022) Analysis and exploration of innovative technology of risk avoidance for accidents in highway tunnels. Tunnel Construction, 42(3): 363. https://orcid.org/10.3973/j.issn.2096-4498.2022.03.003.

[7] Liu S.R., Xie X.H., Ding H. et al. (2022) Statistics and Analysis on Tunnel Accident Data. Modern Tunnelling Technology, 59(SI): 691-697. https://orcid.org/ 10.13807/j.cnki.mtt.2022.S1.083.

[8] Koh J L, Shieh S F.(2004)An efficient approach for maintaining association rules based on adjusting FP-tree structures. In: International Conference on Database Systems for Advanced Applications. Springer, Berlin, Heidelberg. pp. 417-424. https://orcid.org/ 0000-0002-3223-6021.

[9] Yang Y, Yuan Z.Z, Sun D.Y, Wen X.L. (2019)Analysis of the factors influencing highway crash risk in different regional types based on improved Apriori algorithm. Advances in Transportation Studies, 49: 165-178. https://orcid.org/ 10.4399/978882552809113.

[10] Yasmin S, Eluru N; Ukkusuri, S.(2014)Alternative ordered response frameworks for examining pedestrian injury severity in New York City J Transp. Saf Secur: 6,275-300. https://doi.org/10.1080/19439962.2013.839590.