Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 48 (2020) 3496–3507

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research – WCTR 2019, Mumbai, 26-30 May 2019

Identifying Patterns of Pedestrian Crashes in Urban Metropolitan Roads in India using Association Rule Mining

Sathish Kumar Sivasankaran, Priyadarshini Natarajan and Venkatesh Balasubramanian*

RBG Labs, Department of Engineering Design, IIT Madras, Chennai-36.TamilNadu, India.

Abstract

Pedestrian safety is an important component of efforts to prevent road traffic injuries. Pedestrians constitute to around 22% of the total deaths occurring on the world roads. According to the recent report by the Ministry of Road Transport and Highways (MoRTH), Government of India, the number of pedestrian-related deaths was 15,746 (10.5%) of total persons killed in the country during the year 2016. This high proportion of mortality and severity injury among pedestrians necessitates more investigation to identify determinants to reduce crashes in the future. The present research used the Apriori algorithm of supervised association rule mining to identify the patterns of pedestrian severity injury in urban Indian metropolitan city, Chennai. Using the RADMS database of the government of Tamilnadu, vehicle-pedestrian crashes were analyzed for recent two years between 2015 and 2016. The results highlight the fact that middle-aged pedestrians are more vulnerable to road traffic crashes. Exceeding speed limits than the posted speed, especially in the highways, results in fatal crashes among pedestrians. Vehicle-pedestrian crashes are frequent at sites where there are no median separators when drivers do not respect the right of way rules. The findings of the present study will help the traffic safety professionals to understand patterns of crashes and take necessary countermeasures to decrease pedestrian injury-related crashes potentially.

Keywords: Road traffic accidents; Association rule; Data mining; Pedestrian crashes; Injury severity;

1. Introduction

Road accidents continue to be a serious issue in transportation systems worldwide due to substantial economic and social losses in developing countries like India. According to a global status report on road safety by World Health Organization, 1.2 million people die, and 20- 50 million people get injured each year, which makes road traffic

* Corresponding author. Tel.: 91-44-2257 4117; fax: 91-44-2257 4732.
E-mail address: chanakya@iitm.ac.in

2352-1465
10.1016/j.trpro.2020.08.102

accidents as the leading cause of death globally (WHO, 2015). In particular, WHO reports that nearly half of the death on world roads are among the vulnerable road users, which include motorcycles, cyclists, and pedestrians. Further, WHO reports highlight that pedestrians constitute around 22% of the total deaths occurring on the world roads, and it was found that in a few countries, the proportion of pedestrian deaths is around two-third. (Global Status report on road safety, WHO, 2015).

Walking is a common mode of transport in all societies around the world. Every trip begins with walking and ends with walking. Pedestrian is any person who is walking at least for part of his journey. Millions of people are injured in traffic-related crashes all around the world while walking, whereas some of them are permanently disabled. In India, like many other developing nations, pedestrian-related deaths are high. According to the recent report by the Ministry of Road Transport and Highways (MoRTH), Government of India, the number of pedestrian-related deaths was 15,746 (10.5%) of total persons killed in the country during the year 2016 whereas it was 13,894 (9.5%), the previous year (Road Transport Yearbook, 2015). This high proportion of mortality and severity injury among pedestrians necessitates more investigation to identify determinants to reduce crashes in the future.

Pedestrian safety is an important component of efforts to prevent road traffic injuries. Further, pedestrian-related collisions should not be accepted as inevitable because they are predictable and also preventable. (Global Status report on road safety, WHO, 2015). In the recent century, more and more policies are being made to encourage active and safe travel for road users, especially pedestrians for sustainable and efficient transportation systems. The objective of the present study is to identify potential factors and the underlying pattern between the crash characteristics which contribute to Pedestrian crashes in the urban metropolitan city, Chennai.

Road traffic crashes can be considered as series of directly or indirectly associated events. It is a rare and random event that is preceded by the state in which the road user could not cope up with the current environment. Though most of the individual crashes are unique, there exist few common factors in those individual crashes. So from this perspective, data mining methods prove to efficient in providing valuable insights by identifying significant patterns from large datasets. It also identifies a complex relationship between variables, and there's no requirement for assigning variables as dependent or independent, which is required in other data mining techniques such as Logistic Regression and Classification and Regression Trees (CART).

2.Literature Review

Application of the data mining approach to road safety research is limited and diversified into several areas. Among the data mining approaches, logistic regression, probabilistic models, CART (Classification and Regression Trees), artificial neural networks, correspondence analysis, association rules, text mining, and cluster analysis are the most popular techniques. It has been applied in diversified fields such as construction industry, railway crash analysis, industrial and operations management, marketing, oil and gas companies, marine applications. However, applications in road traffic data analysis are limited (Das & Sun, 2014; Montella et al., 2011; Pande & Abdel-Aty, 2009; Weng et al., 2016).

Association rule mining is the descriptive data mining technique (Agarwal et al., 2003), which gained importance in the traffic crash data analysis recently. In traditional market basket analysis, frequently brought items in supermarkets are analyzed. In crash traffic data analysis, crash characteristics that co-exist are identified (Pandel & Abtel-Aty, 2009). Apriori algorithm of association rule mining is used in the present study due to its simplicity in understanding and because of straightforward computations.

Association rule mining was applied earlier to the crash data of Belgium, which showed that human and behavioral aspects have greater importance in analyzing accidents at the black spots (Geurts et al., 2003). Geurts et al., 2005 investigated the characteristics of crashes occurring in the black spots and identified accidents are common on the highways and roads with separate lane during the weekend and at nights. Pandel & Abtel-Aty, 2009 applied association rule mining to crash data of the Florida state during 2004 and identified straight sections with vertical curves during rainy conditions are dominant crash sites. The study also realized the potential of association rule mining as a decision support tool for traffic administrators (Pande and Abtel-Aty, 2009). Association rule mining was applied to the Iranian railway industry to discover the pattern of accidents and underlying relationships to develop rules and regulations to prevent rail accidents (Mirabadi and Sharifian, 2010). Lopez et al. analyzed specific problems of rural highways in the province of Granada province in Spain. The generated association rules highlighted patterns that contribute to severe crashes. Main patterns include pedestrian crashes, run-off-road crashes, run-off-road crashes involving PTW crashes at night without street lights (Lopez et al., 2011). Das and Sun, 2014 investigated the pattern of traffic crashes under rainy weather through the association rule mining approach and found that single-vehicle run-off crashes frequently occur in rainy weather. Kumar and Toshniwal, 2016 used association rule mining to the crash data from the Indian state of Uttarkhand. K-means clustering algorithm was applied to classify accident areas based on the frequency followed by association rules to characterize the accident locations. Yanyan et al., 2017 utilized association rule to analyze factors contributing to misclassification of fatigue-related accidents reported in police records. In their study, association rule mining was applied first to identify potential factors followed by logistic regression on identified factors that hinder police officers in classifying fatigue-related accidents.

3.Methodology

3.1 Association rules

Association rules enumerate interesting transactions among interacting variables. In traffic data analysis, association rule mining is applied to the crash data to obtain rules for identifying accident patterns. In crash data analysis, the Apriori algorithm is applied to extract the association rules among the crash transactions. Agarwal et al.,1993 proposed the algorithm to discover a pattern in the transactions of the supermarket to find which items are frequently brought together. Hence association rule mining is also referred to as market basket analysis or frequent itemset mining/ itemsets mining in more general terms. Rules generation is a discovery process where transactions that occur mutually in the given dataset are identified.

Consider a set of transactions T = $\{t_1, t_2, t_3...\}$ each described as set of items I = $\{i_1, i_2, i_3...\}$ in the dataset. An association rule is defined by the expression $X \rightarrow Y$, where X, Y are item sets where $X = \emptyset$, $Y = \emptyset$, and $X \cap Y = \emptyset$. Within the association rule, X is called antecedent, and Y is called as consequent.

For example, consider the association rule for interpretation

$$\{\text{Central\_divider} = \text{No}, \text{Junction control} = \text{No control}, \\ \text{Road category} = \text{Highway}\} \rightarrow \{\, severity = Fatal\ or\ grievous\ injury\}$$

This rule implies that in the absence of central divider when the junction is not controlled through traffic personnel or traffic signals at the highways, the nature of the severity of the crash would be either fatal or grievously injured. A possible remedy for the above crash would be to install central medians and install traffic controlled devices in highways, which would mitigate the severity of traffic accidents.

3.2 Interesting Rule Mining

Three measures of significance are commonly used in association rule mining: Support, Confidence, and Lift.

3.2.1. Support

Support, sometimes referred to as rule support, is defined as a number of times the rules occur in the transaction data to the total number of transactions in the dataset. The formula for calculating support is shown in

Eq.1. Support is usually represented as a percentage of transactions or the actual number of transactions occurring in the rule.

$$Support = \frac{P(X \cap Y)}{N} \quad (1)$$

3.2.2 Confidence

Confidence or accuracy is the percentage of transactions in which the antecedents *X* and the consequents *Y* are true to the total number of rules in which antecedents *X* are true. It can also be defined as total support to antecedent support. It is calculated using the formula represented in Eq.2.

$$Confidence = \frac{P(X \cap Y)}{P(X)} \quad (2)$$

3.2.3. Lift

Lift is defined as the confidence of the rule to the baseline confidence of the consequent, or it can be expressed as a measure of a number of times the consequent occurs when the antecedent is true compared to the number of times consequent occurs on own. The Eq.3 represents it.

$$Lift = \frac{P(X \cap Y)}{P(X).P(Y)} \quad (3)$$

$P(X \cap Y)$ is the observed frequency of the antecedents and consequence co-occurrence in the rule, whereas the denominator $P(X).P(Y)$ is the expected frequency of antecedents and consequents. Lee et al., 2012 estimates the inferences from the lift values as follows: Lift value <1, indicates negative independence between *X* and *Y*, value Lift = 1 indicates independence and Lift value > 1 indicates positive independence (Lee et al., 2012, Montella et al., 2012).

In association rules, the three parameters, support, confidence, and lift, are set to minimum threshold values to reduce the number of rules identified by the algorithms and also to speed up the process of rules generation. The common parameters which are set to reduce the number of rules are

$$Support(X \cup Y) \geq \sigma$$

$$Confidence(X \cup Y) \geq \delta \text{ and}$$

$$Lift(X \cup Y) \geq \varepsilon$$

Where σ, δ, and ε are the minimum support, confidence, and Lift, respectively.

There are no specific theories for choosing values for parameters (Deona et al.,2013; Montella et al., 2011, 2012; Pande & Abdel-Aty, 2009). Hence on trial and error basis, values are set according to requirements of the number of rules that has to be generated. Setting lower values for support may increase the number of frequent itemsets, whereas setting high values may reduce the number of rules generated.

4.Results and Discussions

4.1. Data

Vehicle-Pedestrian crash data was compiled from a comprehensive road safety accident reporting database RADMS established by Tamil Nadu in 2009. RADMS is a software package developed to collect and analyze road traffic accidents by taking inputs from police, highways, and transport departments. The raw data contained detailed information about the crashes, which was unsorted containing 7592 crashes belonging to the period between January

2015 and December 2016, and finally, crashes involving single pedestrian-vehicle involvement were alone retrieved for the Chennai city. The final dataset had around 3416 crashes to which association rule mining was applied to explore the crash pattern data. The key variables were classified into four categories: Driver- and pedestrian-related (driver's gender, license status, driver's age, pedestrian's age, and pedestrian's gender); Environment-related (weather conditions, light conditions, Chennai zonal regions, season, time, day of the week, vehicle type), Crash related (severity, accident cause, collision type, hit-and-run), Road related (central divider, road category, road condition, speed limit, traffic movement, road works, number of lanes and junction control).

Table 1. Summary of variables relating to pedestrian-vehicle crashes.

Variables and Categories	Frequency of Occurrence	Percentage
Severity		
Fatal/Grievous	2050	60.01%
No/Simple	1366	39.99%
Collision Type		
Head-on	284	8.31%
Hit from rear	79	2.31%
Hit from side	168	4.92%
Hit object	46	1.35%
Hit pedestrian	2216	64.87%
Others	616	18.03%
Skidding	7	0.20%
Central Divider		
No	2202	64.46%
Yes	1214	35.54%
Road Category		
Highway	2344	68.62%
Not highway	305	8.93%
Unknown	767	22.45%
Road Condition		
Good	3411	99.85%
Poor	5	0.15%
Light Condition		
Darkness	391	11.45%
Daylight	1464	42.86%
Street light	769	22.51%
Unknown	792	23.19%
Speed Limit		
30 Km/hr	3	0.09%
35 Km/hr	16	0.47%
40 Km/hr	3359	98.33%
50 Km/hr	37	1.08%
60 Km/hr	1	0.03%
Traffic Movement		
One-way	51	1.49%
Two-way	3365	98.51%
No. of Lanes		
1	2861	83.75%
2	517	15.13%
Greater than 2	38	1.11%
Road Works		0.00%
No	3348	98.01%
Yes	68	1.99%
Accident Cause		
Alcohol abuse	9	0.26%
Animal involved in the accident	3	0.09%
Changing lane without due care	65	1.90%
Dangerous overtaking	70	2.05%
Driving against flow of traffic	56	1.64%
High speed	2646	77.46%
Inattentive turn	30	0.88%
Injured in accidents	79	2.31%
No details entered	42	1.23%
Non-respect of rights of way rules	416	12.18%
Weather Condition		
Cloudy	35	1.02%
Fine	3368	98.59%
Rainy	13	0.38%
License Type		
Full	2454	71.84%
No license	545	15.95%
Unknown	417	12.21%
Driver Gender		
Female	46	1.35%

Variables and Categories	Frequency of Occurrence	Percentage
Male	2192	64.17%
Unknown	1178	34.48%
Season		
Autumn	879	25.73%
Spring	885	25.91%
Summer	885	25.91%
Winter	767	22.45%
Chennai Zone		
East	608	17.80%
North	597	17.48%
South	1502	43.97%
West	709	20.76%
Time		
Early Morning	335	9.81%
Evening	722	21.14%
Midnight	209	6.12%
Morning	705	20.64%
Night	831	24.33%
Noon	614	17.97%
Pedestrian Gender		
Female	801	23.45%
Male	1726	50.53%
Unknown	889	26.02%
Hit and Run		
No	2614	76.52%
Unknown	288	8.43%
Yes	514	15.05%
Junction Control		
Flashing signal	3	0.09%
Give way sign	9	0.26%
No control	201	5.88%

Variables and Categories	Frequency of Occurrence	Percentage
Not at junction	2538	74.30%
Police officer	22	0.64%
Stop sign	25	0.73%
Traffic signals	618	18.09%
Vehicle Type		
Bus	133	3.89%
HGV	114	3.34%
Human power vehicle	3	0.09%
LMV	1006	29.45%
Motor cycle	1577	46.17%
Unknown	583	17.07%
Driver Age		
<18	75	2.20%
>65	26	0.76%
18-24	842	24.65%
25-34	795	23.27%
35-44	447	13.09%
45-54	258	7.55%
55-64	111	3.25%
Pedestrian Age		
<18	198	5.80%
>65	425	12.44%
18-24	151	4.42%
25-34	734	21.49%
35-44	431	12.62%
45-54	495	14.49%
55-64	524	15.34%
Unknown	458	13.41%
Day Type		
Weekday	2451	71.75%
Weekend	965	28.25%

From the preliminary analysis, the variables road condition, speed limit, traffic movement, number of lanes, road works, accident cause, weather condition, junction Control were highly skewed. For example, it could be seen that 99.85% of the accidents occurred in good road conditions, 98.33% of the accidents occurred on roads with a speed limit of 40 km/hr, 77.46% of accidents are due to high speed, 98.51% on a two-way traffic movement, 83.75% on a single lane road, 98.01% with no road works, 98.59% on fine weather conditions and 74.3% of accidents were not at junctions. The nonskewed variables were accident severity, central divider, light conditions, season, accident zone, time, vehicle type, driver, and pedestrian age. The missing values from the selected variables were replaced with the value "unknown" in light condition, road category, license type, driver gender, pedestrian gender, hit and run cases, vehicle type, driver age, and pedestrian age.

4.2. Generation of rules

In the present study, the 'arules' package of R software was used for association rule generation (Hahsler et al., 2007; Hahsler et al., 2008; Hahsler et al., 2018). On preliminary investigation, the dataset included 3416 pedestrian crashes. The frequency plot for the items generated by the association rule mining is shown in Fig.1. From the frequency plot, we find that *Road condition= good, Weather conditions = Fine, Traffic movement= two way, Speed limit = 40, and Road work = no* were the top five frequent items in the dataset.

For significant findings, settings were calibrated to find the meaningful rule by minimum support and confidence values. Also, setting minimum support values may result in more number of rules, which would be challenging to interpret, whereas setting high support values results in fewer rules which might miss out on important rules. After a significant number of trial and error, the minimum values were set for the present study. Ten percent of minimum support means no items or set of items will be considered as frequent if it does not occur in at least 341 crashes (10% of 3416 crashes). This selection is not random, as in this study, for getting intuitive information for each case by optimizing the support and confidence values using supervised association mining (See Table 2).

The association rules were generated considering the nature of the severity of the pedestrians since the objective is to mitigate the severity of the crashes. Hence rules with the pedestrian severity consequents were extracted by using the *apriori* algorithm. Rules which have shared consequents were sorted out based on the decreasing lift values. Table 2 provides the frequency of rules generated for the different cases along with the statistics of support, confidence, and lift.

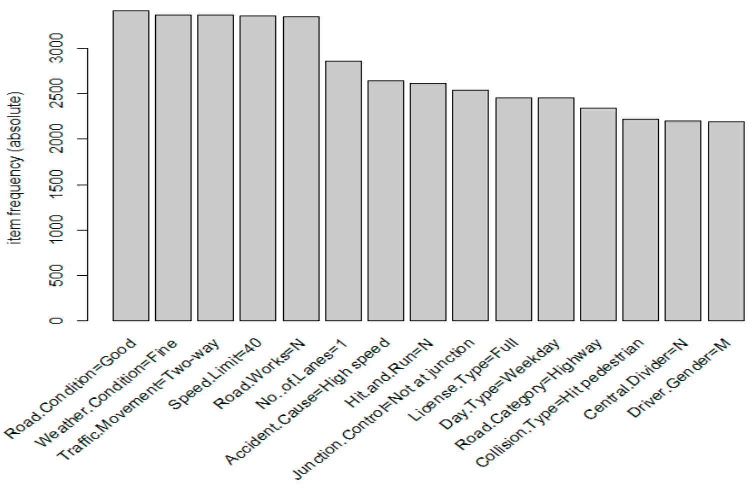

Fig. 1 Item. Frequency plot for the top 15 variables.

4.3 Pedestrian Severity crashes

In the crash database, the severity is classified under two categories: Fatal/Grievous injuries and Simple injury/Property Damage Only(PDO). From the descriptive statistics table, it is clear that pedestrian fatal/grievous injury is significantly higher than the simple injuries/PDO. So two different cases were considered for further analysis based on injury severity: Case1: Fatal/grievous injury and Case 2: Simple injury/PDO.

Table 2. Summary chart of association rule mining

case	consequence	Rules(all)	Rules(Lift>1)	Support (Mean)	Confidence (Mean)	Lift (Mean)
1	Pedestrian severity =Fatal/Grievous	1014	992	0.193	0.673	1.121
2	Pedestrian severity = Simple Injury/PDO	127	127	0.113	0.656	1.641

4.3.1 Fatal/grievous injury pedestrian crashes

The association rules for the fatal/grievous injury crashes as consequents were extracted from the generated rules. After several trial and error, the minimum support value was set to 10%, and the confidence was set to 50%. The number of rules generated after removing the redundant rules was 1014. Among these rules, 992 rules had a lift value greater than 1.

The top 10 rules generated by keeping fatal pedestrian crashes in consequents are listed in Table 3. The topmost rule with the highest lift value is {Speed limit=40, Road works=N, Accident cause=High speed, Weather condition=Fine, License type=Full, Season=Autumn} which is highly associated with fatal/grievous crashes (Support =0.12, Confidence =0.77 and lift =1.28). The explanation for this rule is: 12% of the crashes occurred at sites where road works are not being carried out and crashes occurred in fine weather condition during the autumn season by drivers who possess valid driving license who drove the vehicle at high speed though the speed limit on such roads was 40 kmph, 77% of such crashes resulted in fatal/grievous injuries; the proportion of such crashes were 1.28 times the proportion of other crashes in the complete dataset. Similarly, other rules could also be interpreted. Also, it was found that most of the causes of the top 10 rules were related to environmental and road conditions.

Table 3. First 10 association rules for Pedestrian fatal/Grievous crashes (case 1)

Rules	Antecedents	Support	Confidence	Lift
1	{Speed limit=40, Road works=N, Accident cause=High speed, Weather condition=Fine, License type=Full, Season=Autumn}	0.12	0.77	1.28
2	{Road works=N, Accident cause=High speed, Weather condition=Fine, License type=Full, Season=Autumn}	0.12	0.77	1.28
3	{Speed limit=40, Road works=N, Accident cause=High speed, License type=Full, Season=Autumn}	0.12	0.76	1.27
4	{Road works=N, Accident cause=High speed, License type=Full, Season=Autumn}	0.12	0.76	1.27
5	{Road category=Highway, Speed limit=40, Road works=N, Weather condition=Fine, License type=Full, Season=Autumn}	0.11	0.76	1.27
6	{Speed limit=40, No of Lanes=1, Road works=N, Weather condition=Fine, License type=Full, Season=Autumn}	0.13	0.76	1.27
7	{Speed limit=40, Road works=N, Weather condition=Fine, License type=Full, Season=Autumn}	0.13	0.76	1.27
8	{Collision type=Hit pedestrian Speed limit=40, Accident cause=High speed, Weather condition=Fine, License type=Full, Season=Autumn}	0.11	0.76	1.27
9	{No of Lanes=1, Road works=N, Weather condition=Fine, License type=Full, Season=Autumn}	0.13	0.76	1.27
10	{Road works=N, Weather condition=Fine, License type=Full, Season=Autumn}	0.13	0.76	1.27

4.3.2 Simple injury/Property Damage Only pedestrian crashes

The association rules for the simple injury/property damage only injury crashes as consequents were extracted from the generated rules. The number of rules generated after removing the redundant rules was 127. Among these rules, all rules had a lift value greater than 1.

The top 10 rules generated by keeping simple injures pedestrian crashes in consequents are listed in Table 4. The topmost rule with the highest lift value is {Central divider=N, Light condition=Unknown, Accident cause=Non-respect of rights of way rules, Weather Condition=Fine, License type=Full} which is highly associated with simple injuries/property damages (Support =0.10, Confidence =0.91 and lift =2.28). The explanation for this rule is:10% of the vehicle-pedestrian crashes occurred on sites where there are no median separators in the unknown light conditions in fine weather condition by drivers who possess valid driving license when they do not respect the right of way rules, 91% of such crashes resulted in simple injuries; the proportion of such crashes were 2.28 times the proportion of crashes in the complete dataset. Similarly, other rules could also be interpreted. Also, we find most of the causes for the top 10 rules were related to environmental and road conditions.

Table 3. First 10 association rules for Pedestrian Simple injury/Property damage only crashes (case 2)

Rules	Antecedents	Support	Confidence	Lift
1	{Central divider=N, Light condition=Unknown, Accident cause=Non-respect of rights of way rules, Weather Condition=Fine, License type=Full}	0.10	0.91	2.28
2	{Central divider=N, Light condition=Unknown, Accident cause=Non-respect of rights of way rules, Weather Condition=Fine}	0.11	0.91	2.28
3	{Central divider=N, Accident cause=Non-respect of rights of way rules, Weather Condition=Fine, License type=Full}	0.10	0.91	2.28
4	{Central divider=N, Road category=Unknown, Accident cause=Non-respect of rights of way rules, Weather Condition=Fine}	0.10	0.91	2.27
5	{Light condition=Unknown, Accident cause=Non-respect of rights of way rules, Weather Condition=Fine, License type=Full}	0.11	0.91	2.27
6	{Central divider=N, Light condition=Unknown, Accident cause=Non-respect of rights of way rules, License type=Full}	0.11	0.91	2.27
7	{Light condition=Unknown, Accident cause=Non-respect of rights of way rules, Weather Condition=Fine}	0.11	0.91	2.26
8	{Central divider=N, Light condition=Unknown, Accident cause=Non-respect of rights of way rules}	0.11	0.90	2.26
9	{Road category=Unknown, Accident cause=Non-respect of rights of way rules, Weather Condition=Fine, License type=Full}	0.11	0.90	2.26
10	{Central divider=N, Accident cause=Non-respect of rights of way rules, License type=Full}	0.11	0.90	2.26

Balloon plots provide better visualization for a large number of generated by the association rules. In the balloon plot, the size represents the support values, and the intensity of the color represents the lift values. The balloon plot for the considered two cases is provided in Figures 2 and 3.

5. Conclusion and Future Work

The present paper used data mining approach to analyze the vehicle-pedestrian crashes to extract knowledge for improving pedestrian safety. The unique advantage of using the non-parametric method by association rule mining is that it does not limit the distribution assumption of variables and associations, whereas, in parametric modeling, the explanatory variables should be independent. Further association rules identify significant rules with specific details, which are not possible in the case of conventional statistical models.

The following are a few of the insights which are found out using the association rule mining approach in the present study. The results of this study can provide key factors associated with being found at fault.

- Vehicle-pedestrian crashes occurred on sites where there is no median when they do not respect the right of way rules. To prevent this, median separators should be installed on a priority basis on high-risk zones of accidents. Educational training programs and campaigns should be made mandatory to obey the traffic rules.
- Highways are identified as high-risk locations for pedestrian crashes, which are often found to be fatal. To mitigate accidents, speed breakers should be installed on a priority basis, along with the frequent inspection by the traffic officials.

- Vehicle-pedestrian crashes are severe or even fatal when people drive at high speed, crossing the posted speed limits. To prevent this, drivers should be educated, and speed limits should be posted in appropriate locations. Even pedestrians are encouraged to use retro-reflective clothing, especially during the night times.
- Middle-aged pedestrians (25-54 years) are more vulnerable to road traffic crashes and are significant in numbers. Safety campaigns must be arranged to make pedestrians follow proper traffic safety rules. Education campaigns and programs are the need of the hour to enforce traffic rules and ordinance for pedestrians. (E.g., more cautious while crossing the roads, obeying traffic and pedestrian signals, legal crossings while crossing the streets).

The present paper provides a novel approach to identify the key contributing factors from the set of pedestrian-related crashes. Also, the severity of the crashes was analyzed, which provides better knowledge and helps the city planners and decision-makers to improve pedestrian safety.

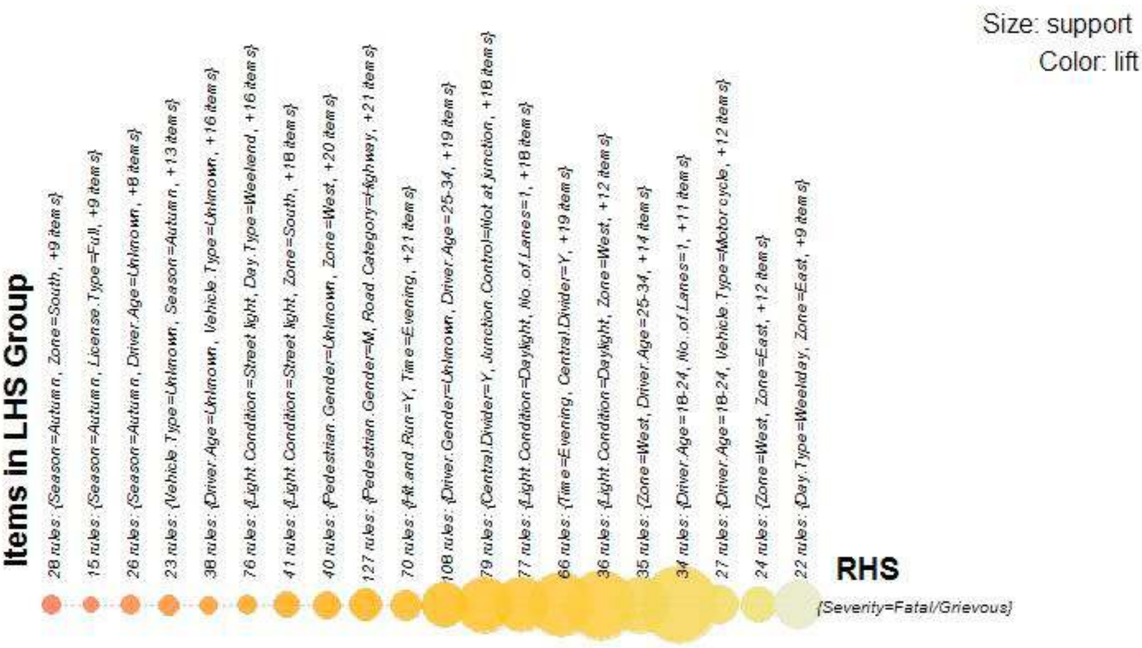

Figure 2. Grouped balloon plot of 992 rules generated for pedestrian fatal/grievous crashes

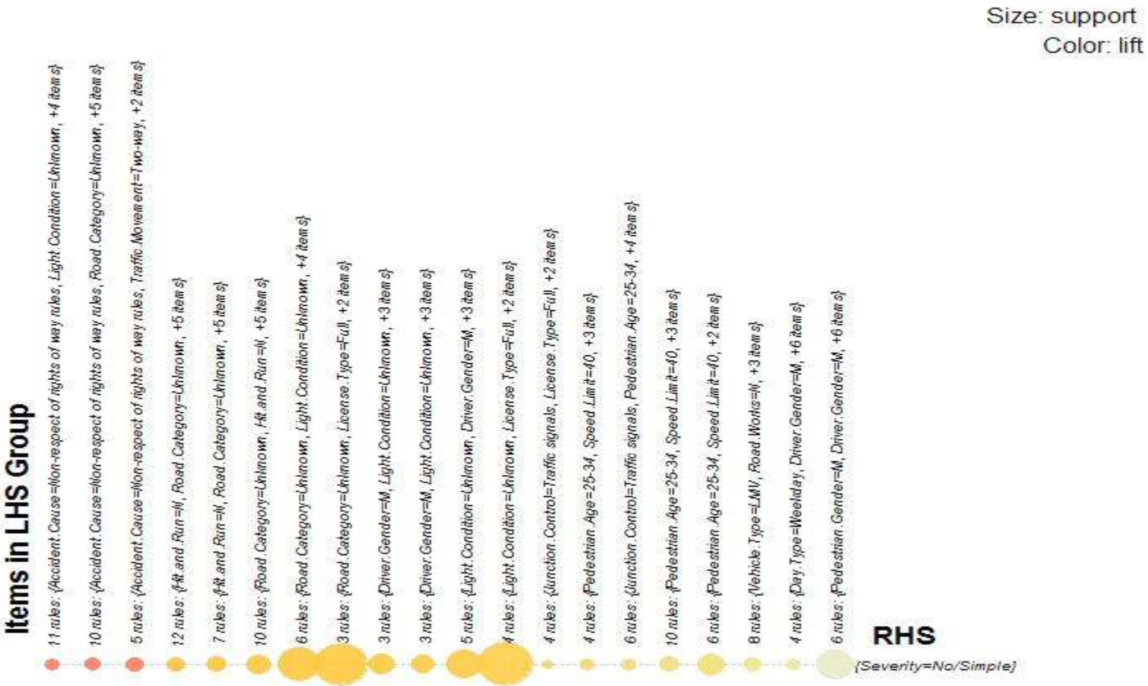

Figure 3. Grouped balloon plot of 127 rules generated for pedestrian simple injury/property damage only crashes

Acknowledgments

The authors wish to thank RADMS and TN police for providing the data. This research did not receive any specific research grants from the funding agencies in public, commercial, or not-for-profit sectors.

References

Abellán, J., López, G., and De OñA, J., 2013. Analysis of traffic accident severity using decision rules via decision trees. Expert Systems with Applications, 40(15), pp.6047-6054.

Das, S. and Sun, X., 2014, January. Investigating the pattern of traffic crashes under rainy weather by association rules in data mining. In Transportation Research Board 93rd Annual Meeting (No. 14-1540). Transportation Research Board Washington DC.

Geurts, K., Thomas, I. and Wets, G., 2005. Understanding spatial concentrations of road accidents using frequent item sets. Accident Analysis & Prevention, 37(4), pp.787-799.

Geurts, K., Wets, G., Brijs, T., and Vanhoof, K., 2003. Profiling of high-frequency accident locations by use of association rules. Transportation research record, 1840(1), pp.123-130.

Hahsler, M. and Chelluboina, S., 2011. Visualizing association rules: Introduction to the R-extension package arulesViz. R project module, pp.223-238.

Hahsler, M., Buchta, C., Gruen, B., Hornik, K., & Hahsler, M. M. (2018). Package 'arules'.

Hornik, K., Grün, B. and Hahsler, M., 2005. arules-A computational environment for mining association rules and frequent item sets. Journal of Statistical Software, 14(15), pp.1-25.

Kumar, S. and Toshniwal, D., 2016. A data mining approach to characterize road accident locations. Journal of Modern Transportation, 24(1), pp.62-72.

Li, Y., Yamamoto, T. and Zhang, G., 2018. Understanding factors associated with misclassification of fatigue-related accidents in police record. Journal of safety research, 64, pp.155-162.

Lopez, D., Glickman, M.E., Soumerai, S.B. and Hemenway, D., 2018. Identifying factors related to a hit-and-run after a vehicle-bicycle collision. Journal of Transport & Health, 8, pp.299-306.

Mirabadi, A. and Sharifian, S., 2010. Application of association rules in Iranian Railways (RAI) accident data analysis. Safety Science, 48(10), pp.1427-1435.

Montella, A., 2011. Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types. Accident Analysis & Prevention, 43(4), pp.1451-1463.

Montella, A., Aria, M., D'Ambrosio, A., and Mauriello, F., 2011. Data-mining techniques for exploratory analysis of pedestrian crashes. Transportation research record, 2237(1), pp.107-116.

Pande, A. and Abdel-Aty, M., 2009. Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. Safety science, 47(1), pp.145-154.

Road accidents in India Report 2012-2016,Transport Research Wing, Ministry of Road Transport and Highways, Government of India.

Weng, J., Zhu, J.Z., Yan, X. and Liu, Z., 2016. Investigation of work zone crash casualty patterns using association rules. Accident Analysis & Prevention, 92, pp.43-52.

World Health Organization, Global status report on Road Safety,2015.

Lee, C., Song, B. and Park, Y., 2012. Design of convergent product concepts based on functionality: An association rule mining and decision tree approach. Expert Systems with Applications, 39(10), pp.9534-9542.

Agrawal, R., Imieliński, T., and Swami, A., 1993, June. Mining association rules between sets of items in large databases. In Acm sigmod record (Vol. 22, No. 2, pp. 207-216). ACM.