



World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

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

Sathish Kumar Sivasankaran, Priyadarshini Natarajan and Venkatesh Balsubramanian*

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

Abstract

Pedestrian safety is an important requirement for preventing road traffic injuries. Pedestrians constitute to around 22% of the total deaths occurring on the world road. 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 future. The present research used Apriori algorithm of supervised association rule mining to identify the patterns of pedestrian severity injury in urban Indian metropolitan city, Chennai. Using RADMS database of government of Tamilnadu, vehicle -pedestrian crashes were analyzed using association rule mining 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 are frequently resulting in fatal crashes among the pedestrians. Vehicle pedestrian crashes are common in sites where there is no median 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 potentially decrease the pedestrian injury related crashes.

©2018 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY.

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

1. Introduction

Road accidents continue to be serious issue in transportation systems worldwide due to huge economic and social losses it results in several developing countries like India. According to global status report on road safety by World health organization, 1.2 million people die and 50 million people get injured each year which makes road traffic accidents as the leading cause of death globally (WHO,2015). In particular, WHO reports that nearly half of the

* Corresponding author. Tel.: 91-44-2257 4117; fax: 91-44-2257 4732.

E-mail address: chanakya@iitm.ac.in

death on world roads are among the vulnerable road users which includes motorcycles, cyclists and pedestrians. Further, reports highlight that pedestrians constitute to around 22% of the total deaths occurring on the world roads and it was found that in few countries the proportion of pedestrians deaths are around two third. (Global Status report on road safety, WHO, 2015).

Walking is the 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 round 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 Year book, 2015). This high proportion of mortality and severity injury among pedestrians necessitates more investigation to identify determinants to reduce crashes in 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 in fact 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 effective and safe travel for the road users especially the pedestrians for an 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 which is preceded by the state which the road user could not cope up with the current environment. though most of the individual crashes are unique, there exists few common factors in these 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 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 data mining approach to the 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, Aria, D'Ambrosio, & Mauriello, 2011; Pande & Abdel-Aty, 2009; Weng, Zhu, Yan, & Liu, 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 analysed. In crash traffic data analysis, similar to market basket analysis crash characteristics which co exists are identified (Pandel and 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 to the crash data of Belgium implied that human and behavioral aspects have greater importance in analyzing accidents at the black spots (Geurts et al, 2003). Geurts et al, 2005 analyzed the characteristics of accidents occurring in the black spots and identified accidents on the highways and roads with separate lane during weekend and at nights frequently. Pandel and abtel aty applied association rule mining to crash data of the Florida state during 2004 and identified straight sections with vertical curves during rainy conditions are crash dominant sites. The study also realized potential of association rule mining as a decision

support tool for the traffic administrators (Pande and Abtel aty, 2009). It was applied to the Iranian railway industry to discover the pattern of accidents and underlying relationship 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 which contribute to severe crashes. Main patterns were pedestrian crashes, run-of-road crashes, run-of-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 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 to 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. 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, Apriori algorithm is applied to extract the association rules among the crash transactions. It was proposed by Agarwal et al, 1993 to discover pattern in transaction of the supermarket to find which the items are frequently brought together. Hence association rule mining is also referred as market basket analysis or frequent item set 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 $T = \{t_1, t_2, t_3, \dots\}$ each described as set of items $I = \{i_1, i_2, i_3, \dots\}$ 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 as antecedent and Y is called as consequent.

For example, consider the association rule for interpretation

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

This rule implies that in the absence of central divider when the junction is not controlled through traffic personnel or traffic signals in the highways, the severity of crash would be either fatal or grievously injured. A possible remedy for the above crash would be to install central meridian and 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 as rule support is defined as 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 done by using Eq.1. Support is usually represented as percentage of transactions or actual number of transactions occurring in the rule.

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

3.2.2 Confidence

Confidence or accuracy is 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 the antecedent support. It is calculated using the formula represented in Eq.2.

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

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 measure of number of times the consequent occurs when the antecedent is true compared to number of times consequent occurs on own. It is represented by the Eq.3.

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

$P(X \cap Y)$ is the observed frequency of the antecedents and consequence co occurrence in the rule whereas the denominator $P(X) \cdot P(Y)$ is expected frequency of antecedents and consequents. Lee at 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 at al, 2012, Montella et al., 2012).

In association rules the three parameters, support, confidence and lift are set to minimum threshold values in order 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

$$\text{Support}(X \cup Y) \geq \sigma$$

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

$$\text{Lift}(X \cup Y) \geq \varepsilon$$

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

There are no specific theories for choosing values for parameters (De Oña, López, & Abellán, 2013; Montella et al., 2011, 2012; Pande & Abdel-Aty, 2009). Hence on trial and error basis, values are set according to requirements of

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 period between January 2015 and December 2016 and finally crashes involving single pedestrian involvement were used. The final dataset contained 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 zone, 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 pedestrians vehicle crash.

Variables and Categories	Frequency of Occurrence	Percentage	Variables and Categories	Frequency of Occurrence	Percentage
Severity			Daylight	1464	42.86%
Fatal/Grievous	2050	60.01%	Street light	769	22.51%
No/Simple	1366	39.99%	And another entry	5	6
Collision Type			Unknown	792	23.19%
Head on	284	8.31%	Speed Limit		
Hit from rear	79	2.31%	30 Km/hr	3	0.09%
Hit from side	168	4.92%	35 Km/hr	16	0.47%
Hit object	46	1.35%	40 Km/hr	3359	98.33%
Hit pedestrian	2216	64.87%	50 Km/hr	37	1.08%
Others	616	18.03%	60 Km/hr	1	0.03%
Skidding	7	0.20%	Traffic Movement		
Central Divider			One-way	51	1.49%
No	2202	64.46%	Two-way	3365	98.51%
Yes	1214	35.54%	No. of Lanes		
Road Category			1	2861	83.75%
Highway	2344	68.62%	2	517	15.13%
Not highway	305	8.93%	Greater than 2	38	1.11%
Unknown	767	22.45%	Road Works		0.00%
Road Condition			No	3348	98.01%
Good	3411	99.85%	Yes	68	1.99%
Poor	5	0.15%	Accident Cause		
Light Condition			Alcohol abuse	9	0.26%
Darkness	391	11.45%	Animal involved in accident	3	0.09%

Variables and Categories	Frequency of Occurrence	Percentage	Variables and Categories	Frequency of Occurrence	Percentage
Changing lane without due care	65	1.90%	Unknown	889	26.02%
Dangerous overtaking	70	2.05%	Hit and Run		
Driving against flow of traffic	56	1.64%	No	2614	76.52%
High speed	2646	77.46%	Unknown	288	8.43%
Inattentive turn	30	0.88%	Yes	514	15.05%
Injured in accidents	79	2.31%	Junction Control		
No details entered	42	1.23%	Flashing signal	3	0.09%
Non-respect of rights of way rules	416	12.18%	Give way sign	9	0.26%
Weather Condition			No control	201	5.88%
Cloudy	35	1.02%	Not at junction	2538	74.30%
Fine	3368	98.59%	Police officer	22	0.64%
Rainy	13	0.38%	Stop sign	25	0.73%
License Type			Traffic signals	618	18.09%
Full	2454	71.84%	Vehicle Type		
No license	545	15.95%	Bus	133	3.89%
Unknown	417	12.21%	HGV	114	3.34%
Driver Gender			Human power vehicle	3	0.09%
Female	46	1.35%	LMV	1006	29.45%
Male	2192	64.17%	Motor cycle	1577	46.17%
Unknown	1178	34.48%	Unknown	583	17.07%
Season			Driver Age		
Autumn	879	25.73%	<18	75	2.20%
Spring	885	25.91%	>65	26	0.76%
Summer	885	25.91%	18-24	842	24.65%
Winter	767	22.45%	25-34	795	23.27%
Chennai Zone			35-44	447	13.09%
East	608	17.80%	45-54	258	7.55%
North	597	17.48%	55-64	111	3.25%
South	1502	43.97%	Pedestrian Age		
West	709	20.76%	<18	198	5.80%
Time			>65	425	12.44%
Early Morning	335	9.81%	18-24	151	4.42%
Evening	722	21.14%	25-34	734	21.49%
Midnight	209	6.12%	35-44	431	12.62%
Morning	705	20.64%	45-54	495	14.49%
Night	831	24.33%	55-64	524	15.34%
Noon	614	17.97%	Unknown	458	13.41%
Pedestrian Gender			Day Type		
Female	801	23.45%	Weekday	2451	71.75%
Male	1726	50.53%	Weekend	965	28.25%

From the four zones in Chennai (north, south, west and east) 3416 data points were reported. 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 speed limit of 40 km/hr, 77.46% of accidents is 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 non skewed 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, 'arules' package of R software was used for association rule generation (Hahsler, Grün, & Hornik, 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, *Road condition= good*, *Weather conditions = Fine*, *Traffic movement= two way*, *Speed limit = 40* and *Road work = no* are found to be the top five frequent items in the dataset considered.

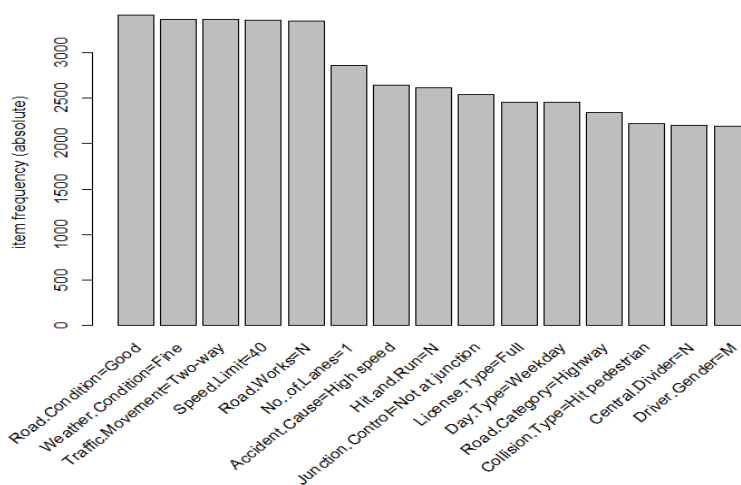


Fig. 1 Item. Frequency plot for top 15 variables.

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

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 difficult to interpret whereas setting high support values results in fewer rules which might miss out important rules. After a

significant number of trial and error the minimum values were set for the present study. Setting up a minimum support value of 10% 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 with the pedestrian severity consequents were extracted by generating rules 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.

4.3 Pedestrian Severity crashes

In the crash database of the Chennai city, 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.

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, 10% was set as minimum support value 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 lift value greater than 1.

The top 10 rules generated by keeping pedestrian fatal crashes in consequents is listed in the Table 3. The top most 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 vehicle pedestrian crashes occurred on sites where road works are not carried out in fine weather condition during the autumn season by drivers who possess valid driving license when they drive at high speed whereas the speed limit on such roads were 40 kmph,77% of such crashes resulted in fatal/grievous injuries; the proportion of such crashes were 1.28 times the proportion of crashes in the complete dataset. Similarly other rules could also be interpreted. Also we find most of causes for 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. 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 127. Among these rules all rules had lift value greater than 1.

The top 10 rules generated by keeping pedestrian fatal crashes in consequents is listed in the Table 4. The top most 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 fatal/grievous crashes (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 is 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 fatal/grievous 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 causes for the top 10 rules were related to environmental and road conditions.

Table 4.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 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 Figure 2 and 3.

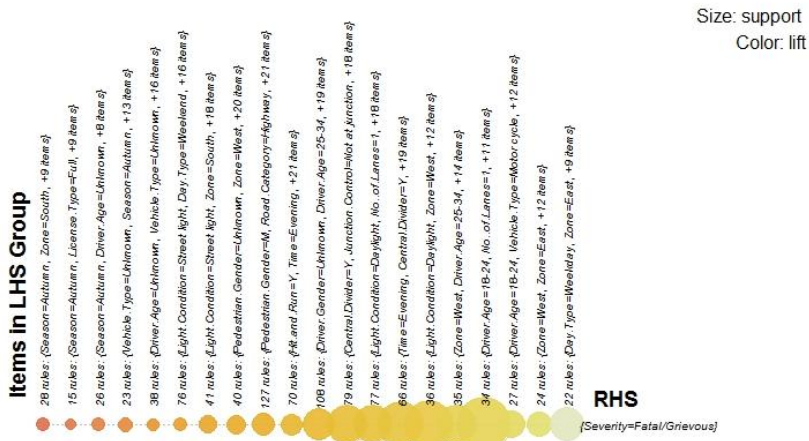


Figure 2. Grouped balloon plot of 992 rules generated for pedestrian fatal/grievous crashes

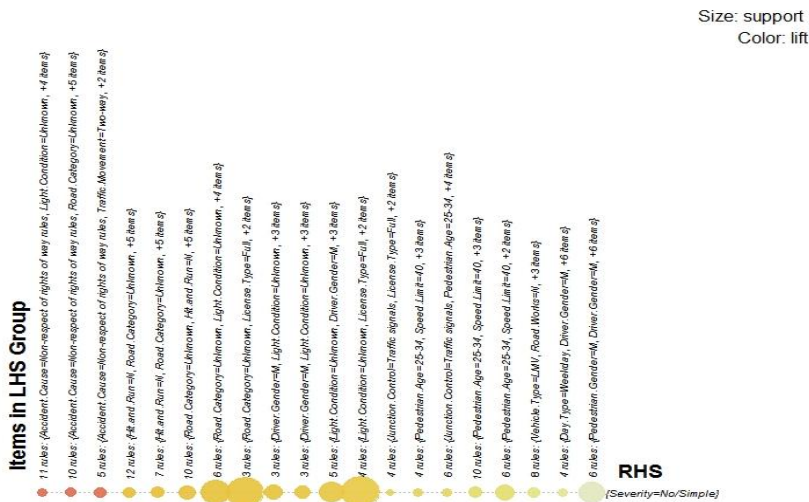


Figure 2 Grouped balloon plot of 127 rules generated for pedestrian simple injury/property damage only crashes

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 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 in nature. Further association rules identifies significant rules with specific details which is not possible in case of conventional statistical models.

Following are few of the insights which are found out using the association rule mining approach in the present study. The results of the present study provides key risk factors which can be considered.

- 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 at 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 the high risk locations for the pedestrian crashes which are often found to be fatal. To mitigate accidents, speed breakers should be installed on 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 on 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 the pedestrians.(For eg, more cautious while crossing the roads, obeying traffic and pedestrian signals, legal crossings while crossing the streets).

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

Acknowledgements

The authors would like to thank all the members and officials of the Government of Tamilnadu for having facilitated this study. The constant support was extended by the management of Institute of road transport and the transport department in the completion of this work.

References

- Abellán, J., López, G., & De Oña, J. (2013). Analysis of traffic accident severity using decision rules via decision trees. *Expert Systems with Applications*, 40(15), 6047-6054.
- Das, S., & Sun, X. (2014). 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).
- Geurts, K., Thomas, I., & Wets, G. (2005). Understanding spatial concentrations of road accidents using frequent item sets. *Accident Analysis & Prevention*, 37(4), 787-799.
- Geurts, K., Wets, G., Brijs, T., & Vanhoof, K. (2003). Profiling of high-frequency accident locations by use of association rules. *Transportation Research Record: Journal of the Transportation Research Board*, (1840), 123-130.
- Hahsler, M., & Chelluboina, S. (2011). Visualizing association rules: Introduction to the R-extension package arulesViz. *R project module*, 223-238.
- Hahsler, M., Buchta, C., Gruen, B., Hornik, K., & Hahsler, M. M. (2018). Package ‘arules’.
- Hahsler, M., Grün, B., & Hornik, K. (2005). A computational environment for mining association rules and frequent item sets
- Kumar, S., & Toshniwal, D. (2016). A data mining approach to characterize road accident locations. *Journal of Modern Transportation*, 24(1), 62-72.
- Li, Y., Yamamoto, T., & Zhang, G. (2018). Understanding factors associated with misclassification of fatigue-related accidents in police record. *Journal of safety research*, 64, 155-162.
- Lopez, D., Glickman, M. E., Soumerai, S. B., & Hemenway, D. (2017). Identifying factors related to a hit-and-run after a vehicle-bicycle collision. *Journal of Transport & Health*.
- Mirabadi, A., & Sharifian, S. (2010). Application of association rules in Iranian Railways (RAI) accident data analysis. *Safety Science*, 48(10), 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), 1451-1463.

- Montella, A., Aria, M., D'Ambrosio, A., & Mauriello, F. (2011). Data-mining techniques for exploratory analysis of pedestrian crashes. *Transportation Research Record: Journal of the Transportation Research Board*, (2237), 107-116.
- Pande, A., & 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), 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., & Liu, Z. (2016). Investigation of work zone crash casualty patterns using association rules. *Accident Analysis & Prevention*, 92, 43-52
- World Health Organization, Global status report on Road Safety, 2015.
- Lee, C., Song, B., & 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), 9534-9542.
- Agrawal, R., Imieliński, T., & 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.