Investigating Factors Associated with Hit-and-run crashes in Indian Metropolitan City Using Association rules

Sathish kumar Sivasankaran and Venkatesh Balasubramanian

RBG lab, Department of Engineering Design, IIT Madras, Chennai, India

Abstract

Hit-and-run crashes are a serious road safety problem threatening the world especially in developing economies like India. It is considered as crime worldwide and stringent laws prevail to punish the offending drivers. In spite of these, rate of hit-and-run crashes in India is higher compared to developed nations and states such as Singapore (1.83%), California (8.1%), Guangdong (7.7%) and shanghai (4.45%) in china. The recent figures according to the Ministry of Road Transport and Highways (MoRTH), government of India shows that hit-and-run crashes accounts for 11.6% (55,942) of all crashes and number of people died due to hit-and-run crashes was 15.2% (22,962) of total persons killed in the country during 2016.

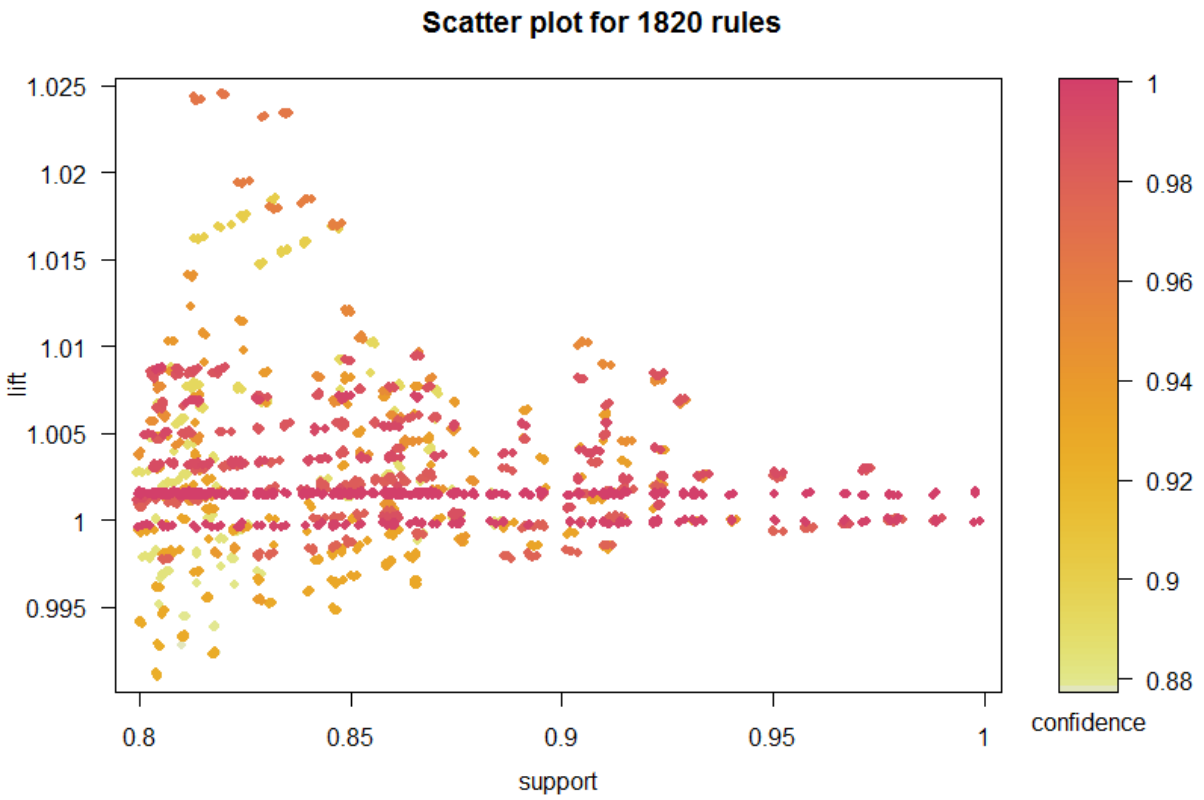
Association rule mining is a popular data mining approach which does not require dependent variable and has the capability in finding complex relationships occurring in large datasets. However, we find research studies relating to associated data mining approach in traffic safety is limited. Apriori algorithm is applied to explore the association rule for hit-and-run crashes.

Hit-and-run crash data for the Chennai metropolitan city was collected from RADMS (Road Accident Data Management System) database. Each crash report includes four types of information (i) Driver and victim information (ii) Vehicle and road information (iii) Environmental characteristics and (iv) Crash information. The dataset includes 670 hit-and-run crashes that occurred in Chennai metropolitan city between January 2015 and December 2016.

Apriori algorithm was applied using the 'arules' package of R software. To identify reasonable and accurate rules, minimum threshold values for the support and confidence was set to 0.8 and 0.5 respectively. This has resulted in generation of 1820 rules. It is very much significant to investigate high support and high lift rules as shown in Figure.1. Higher support indicates higher proportions of times those factors contribute to crashes. In the present case, support >0.9 is considered to be high support rules. Investigating the top 25 high support rules, we find that most of the rules are associated with road characteristics and environmental characteristics. Higher the value of lift represents stronger association between antecedent and the consequent we find 16 rules from the top left corner have high lift values. From the list of rules, all the consequents direct towards Road category = Highways which proves that highways are the prominent sites for the drivers to flee in case of hit-and-run crashes. Among these high lift rules, most of the antecedents contain crash factors, road characteristics and environmental factors.

Crash severity is an important determinant for the drivers decision to stay or flee from the spot of accidents. Top 25 rules which are associated in causing fatal/ grievous injury severity were investigated based on their lift values. In most of the rules, road characteristics contributed much towards fatal/ grievous injuries. Among other factors, environmental characteristics (weather conditions) alone have significantly associated with hit and run crashes in many rules. However, driver and vehicle factors never appeared in any of rules generated.

Figure 1. Scatter plot showing high support and high lift rules.



Network visualization helps in finding the relationship between the antecedent and the consequents for the obtained rules. The Gephi network visualization produced a directed graph with 103 nodes and 1107 edges. Larger sized nodes in the network represent the characteristics or factors which are important in identifying hit-and-run crashes. We find around 10 larger sized nodes and the characteristics of the larger sized nodes were (i) Weather condition= Fine (ii) Severity= Fatal/ Grievous Injury, (iii) Road conditions= Good, (iv) Junction control= No Control, (v) No of Lanes=single, (vi)Traffic movement=Two-way, (vii) Road category= Highway, (viii) License= Valid License, (ix)Primary Cause= Inappropriate speed} and (x) Landmark= Public place. These 10 large sized nodes act as decisive factors for the drivers to flee resulting in hit-and-run crashes.

References

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).

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

Tay, R., Barua, U., & Kattan, L. (2009). Factors contributing to hit-and-run in fatal crashes. *Accident Analysis & Prevention*, *41*(2), 227-233.

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.