© 2018 IJSRST | Volume 4 | Issue 8 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

Analysis of Road Accident Locations Using DBSCAN Algorithm

Kavita Agrawal, Vimala Manohara Ruth, Nandini Y, Sravani K
Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India

ABSTRACT

Road and traffic accidents are one of the important problems in India. However, as accidents are unpredictable and can occur in any type of situation, there is no guarantee that this trend will sustain in future also. Therefore, the identification of different geographical locations where most of the accidents have occurred and determining the various characteristics related to road accidents at these locations will help to understand the different circumstances of accident occurrence. In this paper we apply data mining algorithms on accident dataset as an attempt to address this problem. Association rules were discovered by Apriori algorithm, clusters containing accident prone areas were formed by DBSCAN Algorithm, an android application using fragments was developed to alert the user when he/she enters cluster regions i.e. accident prone areas (ARP) by sending user popup message.

Keywords : Association Rule Mining, DBSCAN algorithm, Apriori Algorithm

I. INTRODUCTION

In recent years, because of too much travel speed of road traffic, the accidents have been increasing on yearly basis. Road accidents happen quite frequently and they claim too many lives every year. Sometimes, it is found that road accident occurrences are more frequent at certain specific locations. The analysis of these locations can help in identifying certain road accident features that make a road accident to occur frequently in these locations.

Data mining uses many different techniques and algorithms to discover the relationship in large amount of data. It is considered one of the most important tool. Association rule mining algorithm is a popular methodology to identify the significant relations between the data stored in large dataset and also plays a very important role in frequent item set mining. A classical association rule mining method is the Apriori Algorithm whose main task is to find frequent item sets, which is the method we use to analyse data.

Clustering is a process of making a group of abstract objects into classes of similar objects. One of the common clustering method is the DBSCAN algorithm is used. It is a density based clustering algorithm: given a set of points in space, it groups together points with many nearby neighbours.

Using Map Activity we created maps in android studio. GPS of the user is tracked using user's latitude and longitude. Using Marker, position of the user is marked. Clusters (Accident prone areas) are also marked in application.

We used an accident dataset ,refer to fig 1 which is provided by Traffic police control office. It contains details of accidents that occurred between 2010 and 2017.The dataset contains 11596 records.

Location	Latitude	Longitude
near MM Hospital, Asifnagar Road	17.39338	78.44462
opposite to HDFC Bank Malakpet	17.37227	78.50961
At Mettuguda Signal, Mettuguda, Secunderabad	17.43615	78.51876
At Mettuguda Signal, Mettuguda, Secunderabad	17.43615	78.51876

Fig 1: Dataset

II. RELATED WORKS

Liling Li andSharadShrestha [1] applied statistics analysis and data mining algorithms on the fatal accident dataset. It is an attempt to give road safety suggestions. They discovered Association rules by Apriori algorithm, classification model was built by Naive Bayes classifier, and clusters were formed by K-means clustering algorithm. The relationship between fatal rate and other attributes including light condition, collision manner, surface condition, weather, and drunk driver were investigated. They used the Fatal Accidents Dataset that contains all fatal accidents on public roads in 2007.

Sachin Kumar and DurgaToshniwal[2] used data mining techniques to identify high-frequency accident locations and further analyzing them to identify various factors that affect road accidents at those locations. They first divided the accident locations into k groups based on the accident frequency counts using k-means clustering algorithm. Then association rule mining algorithm is applied on these to reveal the correlation between different attributes in the accident data and understand the characteristics of these locations.

Amira A. El Tayeb, VikasPareek, AbdelazizAraar [3]used association rule mining algorithms to deliver rules from larger dataset. First data preprocessing is applied on collected records. They collected 1887 traffic records from Dubai police authority then they applied association rules to find frequent pattern sets. They used Data mining software, Weka tool to get these association rules.

DeepthiJayan.K and B.Ganeshkumar [4] an effort was made to identify the accident hotspots, it was an GIS based implementation for Kannur District, Kerala. For this they used the 2006, 2007, 2008 road accident data set. The data set included attributes like date, location, type of vehicle involved, number of persons injured or died. To identify the accident-prone areas Density function available in the spatial analyst extension of the Arc GIS software was applied. Map scanning, Geo referencing, Digitizing were used.

Dipo T. Akomolafe, and Akinbola Olutayo [5] used data mining tool to predict the likely occurrence of accident on highways, the likely cause of the accident and accident prone locations was proposed using Lagos –Ibadan highway as a case study. WEKA software was used to analyze accident data gathered along this road. The results showed that causes of accidents, specific time/condition that could trigger accident and accident prone areas could be effectively identified. Decision Tree was proposed to model data from road accident database to determine causes of accidents and accident prone locations using historical data collected from Ibadan-Lagos express road as reference point. Attribute importance analysis was carried out to rank the attribute by significance using information gain. Finally, correlation based feature subset selection (cfs) and consistency subset selection (COE) filter algorithm were used to rank and select the attribute that are most useful.

Ali Moslah Aljofey and Khalil Alwagih [6] analysed times of accident frequencies for highway locations. The proposal framework consists of clustering technique and classification trees. The k-means algorithm is applied to a set of frequencies of highway locations accidents within 24 hours to find out when and where accidents occur frequently. These frequencies were extracted from 358,448 accident records in Britain between 2013 and 2015. As a result of clustering technique, four clusters were ranked in descending order according to the accidents rate for

location within the cluster. After that, the decision tree (DT) algorithm is applied to the resulting clusters to extract the decision rules as the cluster name represents the class value for all tuples contained.

Geurts K, Wets G, and Brijs T, Vanhoof K [7] identified and profiled black spots and black zones (geographical locations with high concentrations of traffic accidents) in terms of accident-related data and location characteristics must provide new insights into the complexity and causes of road accidents, which, in turn, provide valuable input for governmental actions. Association rules were used to identify accident-related circumstances that frequently occur together. Furthermore, these patterns were analyzed and compared with frequently occurring accident-related characteristics at low-frequency accident locations. The strength of this approach lies with the identification of relevant variables that make a strong contribution toward obtaining a better understanding of accident circumstances and the discerning of descriptive accident patterns from more discriminating accident circumstances to profile black spots and black zones. This data-mining algorithm is particularly useful in the context of large data sets for road accidents. The results showed that human and behavioural aspects are of great importance in the analysis of frequently occurring accident patterns. These factors play an important role in identifying traffic safety problems in general. However, the accident characteristics that were the most discriminating between high-frequency and low-frequency accident locations are mainly related to infrastructure and location.

Divya Bansal, Lekha Bhambu [8] elaborates upon the use of association rule mining in extracting patterns that occur frequently within a dataset and shows the implementation of the Apriori algorithm in mining association rules from a dataset containing crimes data concerning women. As for this WEKA tool is used for extracting results. For this one dataset is taken from UCI repository and other data is collected manually from the session court of sirsa to collect data on heart melting crimes against women. The main motive to use UCI is to first check the proper working of dataset and then apply Apriori on real dataset against crimes on women which extracts hidden information that what age group is responsible for this and to find where the real culprit is hiding. faster than Predictive Apriori Algorithm.

Kim[9] developed a log-linear model to clarify the role of driver characteristics and behaviours in the causal sequence leading to more severe injuries. They found that driver behaviours of alcohol or drug use and lack of seat belt use greatly increase the odds of more severe crashes and injuries.

Bedard [10] applied a multivariate logistic regression to determine the independent contribution of driver, crash, and vehicle characteristics to drivers' fatality risk. It was found that increasing seatbelt use, reducing speed, and reducing the number and severity of driver side impacts might prevent fatalities.

Abdel-Aty [11] used the Fatality Analysis Reporting System (FARS) crash databases covering the period of 1975-2000 to analyze the effect of the increasing number of light truck vehicle (LTV) registrations on fatal angle collision trends . They investigated the number of annual fatalities that result from angle collisions as well as collision configuration (car-car, car-LTV, LTV-car, and LTV-LTV).

Ossiander [12] used Poisson regression to analyse the association between the fatal crash rate (fatal crashes per vehicle mile travelled) and the speed limit increase and found that the speed limit increase was associated with a higher fatal crash rate and more deaths on freeways .

III. METHODOLOGY

The approach we took for our study follows:

A. Apriori Algorithm

In case of road accident data, an association rule can identify the various attribute values which are responsible for an accident occurrence.The data we considered for association rule mining is assumed data. An arff file is created considering the attributes that are shown in table 1.

After applying apriori algorithm with minimum support=0.2 and minimum confidence=0.9 in weka, association rules were generated. The best 10 rules are shown in Fig 2.We could see that more no. of accidents are occurring due to 2-wheelers.Strong rules with high lift value disclose that curves at markets are main locations where accidents have occurred.

Attribute Name	Type	Values
Accident Category	Nominal	2-wheeler, 3-wheeler, 4-wheeler, pedestrian-hit, multi-vehicular-incident
Location_No	Nominal	A, B, C, D, E, F, G, H, I, J
Lighting on road	Nominal	daylight, streetlight, nolight
Roadway feature	Nominal	curve, slope, intersection
Accident Severity	Binary	critical, non-critical
Area Around	Nominal	hospital, hill, market, agriculture-land
Road Type	Nominal	highway, non-highway

Table 1: Attributes

Apriori Algorithm:

Step1:Scan the data set and find the support(s) of each item.
Step2:Generate length (k+1) candidate item sets from length k frequent item sets to generate the set of candidate k - item set.
Step3:Scan the candidate k item set and generate the support of each candidate k - item set.
Step4:Add to frequent item set, until C=Null Set.
Step5:For each item in the frequent item set generate all non empty subsets.
Step6:For each non empty subset determine the confidence. If confidence is greater than or equal to this specified confidence. Then add to Strong Association Rule

- Min_support=0.2
- Min_conf=0.9

B. Clustering

To find out which areas are accident prone areas, DBSCAN algorithm is applied. It requires two values. One is epsilon value, which specifies how close points should be to each other to be considered a part of cluster and other is minPts, which specifies how many neighbours a point should have to be included into a cluster.

DBSCAN is performed based on two attributes (latitude and longitude). The result contains clusters(each cluster contain group of accident areas). These clusters are marked on map as shown in fig 3.

DBSCAN Algorithm:

Let $X = \{x_1, x_2, x_3, ..., x_n\}$ be the set of data points. DBSCAN requires two parameters: ε (eps) and the minimum number of points required to form a cluster (minPts).
Step1: Start with an arbitrary starting point that has not been visited.

Step2: Extract the neighbourhood of this point using ε (All points which are within the ε distance are neighbourhood).

Step3: If there are sufficient neighbourhood around this point then clustering process starts and point is marked as visited else this point is labelled as noise (Later this point can become the part of the cluster).
Step4: If a point is found to be a part of the cluster then its ε neighbourhood is also the part of the cluster.
Step5: A new unvisited point is retrieved and processed, leading to the discovery of a further cluster or noise.

Step6: This process continues until all points are marked as visited.

C.Android Application

After clustering is performed an android application is developed that sends a popup message to the user when he/she enters accident prone area. Map activity is used to get maps.
Steps performed to send a message to the user follows:

- Creating a map (using map activity)
- Getting GPS of the user ,see in fig 4 (using location listeners)
- Display Accident Prone Areas in map (which are obtained from clustering see in fig 5)
- Finding out user entering that region (using latitudes and longitudes)
- Popup message (send pop up message when user enters cluster area)

IV. RESULTS

```
1. Area_around=market 21 ==> Road_type=non-highway 21
2. Area_around=hill 20 ==> Road_type=non-highway 20
3. Roadway_feature=curve Area_around=market 18 ==> Road
4. Accident_severity=critical Area_around=market 16 ==>
5. Accident_category=2-wheeler Area_around=market 14 ==
6. Accident_category=2-wheeler Area_around=hill 13 ==>
7. Accident_severity=critical Area_around=hill 13 ==> R
8. Roadway_feature=curve Accident_severity=critical Are
9. Accident_category=2-wheeler Roadway_feature=curve Ar
```

Fig 2: Association rules

The above screenshot contains 10 best rules found in weka by applying apriori algorithm.

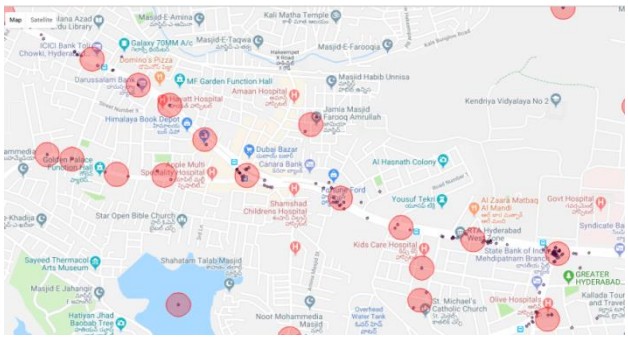

Fig 3: Clusters

In the above figure black points are accident areas and red circles are clusters formed.

Fig 4: User location

The above figure contains blue mark which is user location. The above figure contains blue mark which is user location

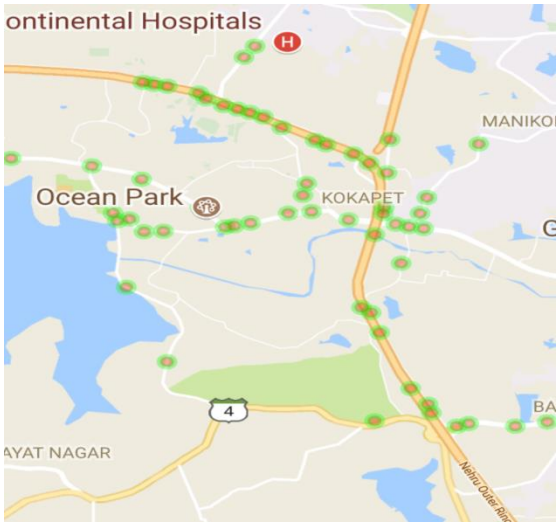

Fig 5: clusters marked in app

The above figure contains accident prone areas .These are obtained from clustering.

V. CONCLUSION

Various data mining techniques such as clustering, association rule mining are widely used in the system to identify reasons that affect the severity of road accidents.Association rule mining is a very popular technique that can be used to identify the relationship among different sets of attributes that frequently occur together when an accident takes place. In our system, we applied association rule mining algorithm on different groups of accident locations. The rules generated for every group exposed the various factors associated with road accidents in these locations. The results obtained from DBSCAN clustering gives the cluster of accident areas. The user is notified when he/she enters accident area. Thus accidents can be decreased to an extent.

VI. REFERENCES

[1]. Liling Li, SharadShrestha, Gongzhu Hu," Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques",IEEE,Vol.10, 2017, No.11-09.

[2]. Sachin Kumar, DurgaToshniwal,"A Data mining approach to characterize road accident locations", J. Mod. Transport,Vol.24, 2016, No.1,62-72.

[3]. Amira A. El Tayeb, VikasPareek, AbdelazizAraar" Applying Association Rules Mining Algorithms for Traffic Accidents in Dubai",IJSCE,Vol.5,2015.

[4]. DeepthiJayan.k, B.Ganesh Kumar "Identification of Accident Hot Spots: A Gis Based Implementation for Kannur District, Kerala " IJGG, Vol.1, 2010.

[5]. DipoT.Akomolafe, Akinbola Olutayo," Using Data Mining Technique to Predict Cause of Accident and Accident Prone Locations on Highways", American Journal of Database Theory and Application ,Vol:2012, No.1(3): 26-38

[6]. Geurts K, Wets G, Brijs T, Vanhoof K," Pro?ling of high frequency accident locations by use of association rules". Transport Research Vol.10,2013

[7]. Ali Moslah Aljofey and Khalil Alwagih "Analysis of Accident Times for Highway Locations Using K-Means Clustering and Decision Rules Extracted from Decision Trees." ISSN,Vol .7,2018.

[8]. Divya Bansal and Lekha Bhambhu. Execution of Apriori algorithm of data mining directed towards tumultuous crimes concerning women"". International Journal of Advanced Research in Computer Science and Software Engineering, 3(9), September 2013.

[9]. Kim, K., Nitz, L., Richardson, J., & Li, L., Personal and Behavioral Predictors of Automobile Crash and Injury Severity"" Accident Analysis and Prevention, Vol. 27, No. 4, 1995, pp. 469-481.

[10]. Bedard, M., Guyatt, G. H., Stones, M. J., & Hireds, J. P., "The Independent Contribution of Driver, Crash, and Vehicle Characteristics to Driver Fatalities". Accident analysis and Prevention, Vol. 34, 2002, pp. 717-727.

[11]. Abdel-Aty, M., & Abdelwahab, H., "Analysis and Prediction of Traffic Fatalities Resulting From Angle Collisions Including the Effect of Vehicles-Configuration and Compatibility" Accident Analysis and Prevention, 2003.

[12]. Ossiander, E. M., & Cummings, P.,"Freeway speed limits and Traffic Fatalities" in Journal of Computer Applications, Vol II, No.4, Oct - Dec 2009 Page 37 Washington State. Accident Analysis and Prevention, Vol. 34, 2002, pp. 13-18.