© 2018 IJSRSET | Volume 4 | Issue 1 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology

Analysis of Roadway Traffic using Data Mining Techniques : A Review

Nisha A. Solanke, Prof A. D. Gotmare
M. Tech (Computer Science & Engineering) Bapurao Deshmukh College of Engineering, Sevagram, Maharashtra, India

ABSTRACT
Roadway traffic safety is a major concern for transportation governing agencies as well as ordinary citizens. Data Mining is taking out of hidden patterns from huge database. It is commonly used in a marketing, surveillance, fraud detection and scientific discovery. In data mining, machine learning is mainly focused as research which is automatically learnt to recognize complex patterns and make intelligent decisions based on data. Globalization has affected many countries. There has been a drastic increase in the economic activities and consumption level, leading to expansion of travel and transportation. The increase in the vehicles, traffic lead to road accidents. Considering the importance of the road safety, government is trying to identify the causes of road accidents to reduce the accidents level. The exponential increase in the accidents data is making it difficult to analyze the constraints causing the road accidents. The paper describes how to mine frequent patterns causing road accidents from collected data set. We find associations among road accidents and predict the type of accidents for existing as well as for new roads. We make use of association and classification rules to discover the patterns between road accidents and as well as predict road accidents for new roads.

Keywords: Data Mining, Association Rule, Classification Rule, Apriori Algorithm, Naïve Bayes Algorithm

I. INTRODUCTION

There are a lot of vehicles driving on the roadway every day, and traffic accidents could happen at any time anywhere. Some accident involves fatality, means people die in that accident. As human being, we all want to avoid accident and stay safe. To find out how to drive safer, data mining technique could be applied on the traffic accident dataset to find out some valuable information, thus give driving suggestion. Accidents happened due to the negligence of driving vehicle on the roads. There are various reasons responsible for the accident like abandon of traffic rules but road conditions and the traffic are considered the one of prime cause of fatality and causality across the globe. These accidents occur due to dynamic design and development of automobile industries. A traffic crash happens due certain reasons like smashes of two vehicles on road, walking person, animal, or any other natural obstacles. It could result in injury, property damage, and death. Traffic accident analysis required study of the various factor affecting behind them.

In survey it's seen that approximate 1.2 million death and 50 million injuries estimated worldwide every year. The approximate estimation of causality and injuries due to poor road infrastructure is a big challenge before the living beings. The order to deal with the problem, in computational science, we can adopt data mining model for different scenario. In any vehicle accident, it studies about the driver's behavior, road infrastructure and possibilities of weather forecast that could be somewhere connected with different accident incidents. The main problem in the study and analysis of accident data is its mix heterogeneous environment and data segmentation which is used widely to overcome accident problem. [2,5,7]

Data Mining is a computational technique to deal with large and complex data set and these data sets can be of normal, nominal and mixed. It is quite easy to use in variety of domain belong to science and management; also, it could be used in fraud identification and many more scientific cases as well as in accident severity problem. Partition of objects in a group of clusters or in

a homogeneous set is a fundamental operation of data mining.

Clustering is a method to partition objects in a similar group. The k-means algorithm having a good efficiency for clustering large data sets but restricted in forming clusters for real word data while working only on numerical data because it helps in reducing the cost function by altering the meaning of the clusters [1,3].

II. LITERATURE REVIEW

In the growing countries in the globe, the motorist, are facing road accidents due to poor management in traffic seeing the common leading cause of injury in body and mortality. Data mining techniques could be used to resolve these issues. In survey, numerous researchers contributed and discussed about various techniques of data mining, few important in the context of our problem are shared in this review paper.

Gower et, al., (1971) in the "Execution of Apriori algorithm of data mining directed towards tumultuous" showed the importance of similarity coefficient and Gowda et, al., and Anderberg et, al., share dissimilarity measures that specify the standard mechanism of hierarchical clustering methods work with numeric and categorical values. But conversion of categorical data with the numeric dataset which will not produce meaningful result when categorical domains are not in order.

Ralambondrainy (1995) in the "International Journal of Soft Computing and Engineering" introduced k-means algorithm approach using data mining to cluster categorical data which convert multiple category attributes into binary numeric attributes. But in data mining these attributes are in hundreds and thousands that compulsory make increment in computation as well as in the space costs of the k-means.

Zhexue Huang (1998),in the "Accidents Analysis and Prevention" proposed two algorithms which is extension of K-means algorithm. This extended k-means based algorithm includes categorical domain with numeric and categorical values. The k-mean algorithm uses a simple matching dissimilarity measure to deal with categorical objects where k-means algorithm extended replaces the means of clusters with modes, and uses a frequency-based method to update modes in the clustering process to minimise the clustering cost function.

Sachin et, al., (2015), in the "Journal of Computer Applications" proposed a framework for Dehradun, India road accident (11,574) happened during 2009 and 2014 by using K-modes clustering technique and association rule mining. The analysis of result using combination of these technique conclude that the result will be more effective if no segmentation has been performed prior to generate association rules [2].

In the world health organization [8], India is taking leading edge with 1,05,000 traffic deaths in a year, with comparison to the china with over 96,000 deaths on road. The survey was executed with approximate 178 countries. As per the survey results, it shown that approximate more than 300 Indians causality on roads every day. There are more than two million people have casualty from a traffic accident. The survey is taken from the report of data collection for 2008.

S. Krishnaveni, (2011),in the "Analysis and Visualization of road accidents" work with some of classification models to predict the injuries happened in traffic accident in Nigeria's and compared Naive Bayes Bayesian classifier [3]. This research is employed on the artificial neural networks based approach while the decision trees data analysis can be used to works on reduction of massacre on the highways. The data was classified in continuous and categorical data where continuous data analysed using artificial neural networks technique and the categorical data, using decision trees technique. The results reveal that decision tree approach outperformed the ANN with a lower error rate and higher accuracy rate. This research based on three most important causes of accident due to tyre burst, loss of control and over speeding. This study used traffic accident records from 1995 to 2000, a total number of 417,670 cases. They applied them to an actual data set obtained from the National Automotive Sampling System (NASS) General Estimates System (GES). Experiment results reveal that in all the cases the decision tree outperforms the neural network. This research analysis also shows that the three most important factors in fatal injury are: driver's seat belt usage, light condition of the roadway, and driver's alcohol usage. [4]

K. Jayasudha, (2009), in the "Mining Association Rules Between Sets of Items in Large Databases" shown the effective use of association rule to investigate the accident issue. She also put efforts that systematic deployment of patters and rules shows the positive impact and it helps in understanding the case of fatality in accidents using decision support system. [9].

K. Geetha, (2015), in the "Fast Algorithm For Mining Association Rules in Large Databases" this study works on traffic accident data of tamilnadu city. The main aim of this study is to reduce the number of road accidents. The traffic accident data is managed in form of text or numerical formats in unsorted manner [5].

Sachin Kumar et, al., (2016) in the "Mining road traffic accidents data to improve safety in dubai" suggest to apply k-means algorithm and ARM technique to solve traffic accident severity problem. Author divide the different accidental prone location with three different categories which are high, moderate and low frequency to extract the hidden information behind the data set and take some preventive action according to accident location [6].

Miao Chong. et. al.,in the "Random Forests" also proposed the efficient use of ANN and DT prove good result, in support they have used GES automobile accident data from 1995 to 2000, by studying the analysis performance of different data mining technique a significant result visible in support of fatality case study. Direct decision based approach outperforms the direct NN approach in all cases. Author discussed in this theory, if speed limit factor is well known then accident can be controlled [10].

III. PROPOSED WORK

The proposed work is planned and carried out in the following manner:

Roadway traffic safety is a major concern for transportation governing agencies as well as ordinary citizens so, for that purpose we are introducing an analytic tool in which Genetic algorithm will be used for classification. We are taking dataset of a country and analyzing that database month wise for one year, To find out which states are similar to each other considering fatal rate, and which states are safer or more risky to drive, clustering algorithm was performed on the fatal accidents dataset. Before applying the algorithms, the tuples with missing value in chosen attributes were removed. The proposed work is planned to be carried out in the following manner.

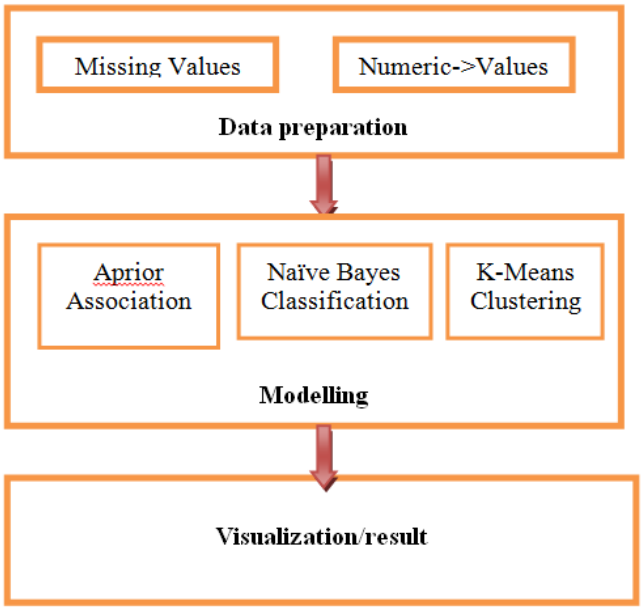

Figure 1. Architectural Design

Data preparation was performed before each model construction. All records with missing value (usually represented by 99 in the dataset) in the chosen attributes were removed. All numerical values were converted to nominal value according to the data dictionary in attached user guide.

In the modeling We first calculated several statistics from the dataset to show the basic characteristics of the fatal accidents. We then applied association rule mining, clustering, and Naiive Bayse classification to find relationships among the attributes and the patterns.

Clustering

Clustering is a process of collection of objects which are similar between them while dissimilar objects belong to other clusters. A clustering technique is used to obtain a partition of N objects using a suitable measure such as resemblance function as a distance measure'd'.

K-means Algorithm for Clustering

K-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. The algorithm has a loose relationship to the k-nearest neighbour classifier, a popular machine learning technique for classification

that is often confused with k-means because of the k in the name. One can apply the 1-nearest neighbour classifier on the cluster centres obtained by k-means to classify new data into the existing clusters.

Association Rule

To find out how to drive safer, data mining technique could be applied on the traffic accident dataset to find out some valuable information, thus give driving suggestion.

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 in information technology in the previous decades.

Association rule mining algorithm is a popular methodology to identify the significant relations between the data stored in large database and also plays a very important role in frequent itemset mining. A classical association rule mining method is the Apriori algorithm who main task is to find frequent itemsetss, which is the method we use to analyze the roadway traffic data.

Naive Bayes Classification

Classification in data mining methodology aims at constructing a model (classifier) from a training data set that can be used to classify records of unknown class labels. The Naive Bayes technique is one of the very basic probability-based methods for classification that is based on the Bayes' hypothesis with the presumption of independence between each pair of variables.

Naive Bayes classifier was built on the cleaned data. The Naive Bayes Classifier shows that the fatal rate does not strongly depend on the given attributes, although they are considered feature in comparison to other attributes in the dataset.

The main source of dataset is www.data.gov.in which government website in that state wise dataset of accidents we are taken of the year 2014-2015, number of files are present in the form of excel format and each file contain state wise data based on various types of parameters.

IV. OBJECTIVES

The main objectives of the study are listed below:

1. To collect and process the dataset
2. To apply Apriori Algorithm for deciding association
3. To apply Naive Bayes Classification model and K-means clustering algorithm to generate clusters
4. To carry out statistical analysis of the generated result
5. To obtain cluster and provide safety measures based on statistical, association, classification model.

V. CONCLUSION

In this paper, we have collected multiple researchers' works together in single document as review and discussed about the contribution towards impact of road and traffic accident on human life and society. This survey highlights the number of approaches used to avoid the accident happened in various countries and cities. The paper also discussing about various data mining techniques which is proved supporting to resolve traffic accident severity problem and conclude which one could be optimal technique in road traffic accident scenario. The brief survey will also help us to find better mining technique in this kind of problem.

VI. REFERENCES

[1]. Zhexue Huang, "Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values", Data Mining and Knowledge Discovery 2, 283–304 (1998).

[2]. Sachin Kumar and Durga Toshniwal, "A data mining framework to analyse road accident data", Journal of Big Data (2015) 2:26 DOI 10.1186/s40537-015-0035-y.

[3]. S. Krishnaveni ans Dr. M. Hemalatha, "A perspective analysis of Traffic Accident Using Data Mining Techniques", International Journal of Computer Application.

[4]. Olutayo V.A and Eludire A.A, "Traffic Accident Analysis Using Decision Trees and Neural Networks", I.J. Information Technology and Computer Science, 2014, 02, 22-28 Published Online January 2014 in MECS (http://www.mecs-press.org/) DOI: 10.5815/ijitcs. 2014.02.03.

[5]. K. Geetha and C. Vaishnavi, "Analysis on Traffic Accident Injury Level Using Classification", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 2, February 2015, ISSN: 2277 128X.
[6]. Sachin Kumar and Durga Toshniwal, "A data mining approach to characterize road accident locations", J. Mod. Transport. (2016) 24(1):62–72 DOI 10.1007/s40534-016-0095-5.
[7]. Tibebe Beshah, Shawndra Hill, "Mining Road Traffic Accident Data to Improve Safety: Role of Road- elated Factors on Accident Severity in Ethiopia"
[8]. Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993. [9] K. Jayasudha and Dr. C. Chandrasekar, "An overview of Data Mining in Road Traffic and Accident Analysis", Journal of Computer Applications, Vol – II, No.4, Oct – Dec 2009.
[9]. Miao Chong, Ajith Abraham and Marcin Paprzycki, "Traffic Accident Analysis Using Machine Learning Paradigms", Informatica 29 (2005) 89–98.
[10]. M. Sowmya and Dr.P. Ponmuthuramalingam, "Analyzing the Road Traffic and Accidents with Classification Techniques", International Journal of Computer Trends and Technology (IJCTT) – volume 5 number 4 –Nov 2013.
[11]. Amira A El Tayeb, Vikas Pareek, and Abdelaziz Araar.
[12]. Applying association rules mining algorithms for traffic accidents in dubai. International Journal of Soft Computingand Engineering, September 2015.
[13]. KMA Solaiman, Md Mustafizur Rahman, and Nashid Shahriar. Avra Bangladesh collection, analysis & visualization of road accident data in Bangladesh. In Proceedings of InternationalConference on Informatics, Electronics & Vision, pages 1-6.IEEE, 2013.
[14]. Eric M Ossiander and Peter Cummings. Freeway speed limits and traffic fatalities in washington state. Accident Analysis & Prevention, 34(1):13-18, 2002.