ADVANCED TRAFFIC CONGESTION AVOIDANCE ALGORITHMS

[1] Sagar Penumala, [2] Dr. G. Prasanna Lakshmi
[1]M. Tech Student, [2]Professor
[1,2]Department of Computer Science and Systems Engineering
[1,2]Andhra University College of Engineering, Visakhapatnam, AP, India
sagarpenumala@gmail.com

Abstract

Roadway traffic safety is a major concern for Road Transport agencies as well as ordinary citizens. There are a lot of vehicles driving on the roadway every day, and traffic accidents could happen at any time anywhere. Some accident involves a fatality, means people die in that accident. Today, traffic safety is one of the main priorities of governments. In order to give Tips to driving safe and preventing accidents on the road, careful analysis of roadway traffic data is critical to find out variables that are closely related to fatal accidents. Considering the importance of the topic, identifying the factors of road accidents has become the main aim to reduce the damage caused by traffic accidents. This paper proposing a combination of clustering (Eclat) association rule mining(Clara) algorithms to find the rules that are mostly contributing the road accidents. The main aim of the proposed system is to find the most important patterns in road accident data.

Keywords: *Association rule mining, Data mining, Clustering, Clara, Eclat.*

1. Introduction

Road accidents are uncertain and unpredictable. Analysis of Road accidents requires knowledge about the factors affecting them. Road and traffic accidents are defined by a set of variables which are mostly discrete nature. The major problem in the analysis of accident data is its heterogeneity. Heterogeneity must be considered during the analysis of data otherwise, some relationship between data may remain hidden. In order to give safe driving suggestions, careful analysis of the roadway traffic data is critical to find out variables that are closely related to fatal accidents. However, any damage resulting from road accidents is always unacceptable in terms of health, property damage and other economic factors. Sometimes, it is found that road accident occurrences are more frequent at certain specific locations. The analysis of these locations can help in identifying precise road accident features that make road accident to occur frequently in these locations

A variety of research has been done on data collected through police records covering a limited portion of highways. The analysis of such data can only reveal information regarding that portion only, but accidents are scattered not only on highways but also on local roads. In order to give safe driving Tipsthe, careful analysis of roadway traffic data is critical to find out variables that are closely related to fatal accidents. The aim of the project is to find the rules that are mostly contributing to road accidents using Association(Eclat) and Clustering(Clara) methods.

2.Literature Review

Researchers have proposed many models to analyze the road accidents data some of them are:

Sachin Kumar, Durga Toshniwal proposed a framework that used K-modes clustering technique as a preliminary task for segmentation of 11,574 road accidents on the road

network of Dehradun (India) between 2009 and 2014. Association rule mining is used to identify the various circumstances that are associated with the related incidences of an accident for both the entire data set (EDS) and the clusters identified by K-modes clustering algorithm [1][3]. Taamneh and Madhar Alkheder proposed Data mining techniques to establish classifier to predict the injury severity of any new accident with hands-on accuracy, based on 5,973 traffic accident records in Abu Dhabi over a 6-year period from 2008 to 2013. Additionally, the research aimed to create a new set of rules that can be used by the United Arab Emirates (UAE) Traffic Agencies to identify the main factors that contribute to accident severity [2][3][8][9]. The new relationship between fatal rate and other attributes including collision manner, weather, surface condition, light condition, and drunk driver were investigated [3]. Self-organization map (SOM) is used for finding a no. of pattern to analysis the road accident data which help to find the prediction of accident reasons and improve the accuracy of analysis compare to k-means clustering algorithm [4].

Muhyiddin ISLAM and Kunnawee KANITPONG address an in-depth study through crash investigation and reconstruction which has not yet been practiced in Thailand to identify the contributory factors in road crashes by the concerned authorities. This research attempts to establish the linkage between the causes and consequences with event classification of an investigated case by highlighting the dynamic driving situation with initial traveling speed, pre and post impact speed of the vehicles involved to describe the crash scenario [5]. Maninder Singh and Amrit Kaur proposed a method to evaluate a set of variable that contributes to the degree of accident severity in traffic crashes [6]. E. Suganya and S. Vijayarani analyze the road accidents using classification algorithms namely linear regression, logistic regression, decision tree, Support Vector Machine, Naive Bayes, KNN, Random Forest and Gradient Boosting Algorithm [10]. Jamshid Sodikov proposed a paper to show how to extract meaningful data from the raw database and visualize it. The results revealed that hour wise, day wise, month wise and year wise plots allowed observing how road traffic accidents change in timescale. Two types of road traffic accident mainly occurred, such as type 1 (collision) and type 5 (collision with the pedestrian) [11].

3. Proposed Methodology

Data Collection

Determining the main factors associated with road traffic accidents is one of the main objectives of the accident data analysis. The proposed system implementing algorithms on UK road accidents data. The dataset contains 32 columns of 140037 records, which describes different contributing factors of road accidents. The attributes of the data are mostly categorical in nature

Data preprocessing

Data preprocessing is the primary task to be done prior to analysis to get the data ready for analysis. As good data can only provide good results, data preprocessing becomes necessary prior to analysis. In data preprocessing, the proposed system performs data cleaning, data imputation, data normalization, and transformation. Data cleaning process removes null values and redundant attributes from the dataset. Removing outliers from the dataset is a very important task as they will affect the outcome proposed model.

Implementation

Clara is one of the partitioning type clustering method which can be efficiently used for large datasets. K-means and K-medoids work efficiently for smaller data sets but does not scale well for larger datasets. Clara draws multiple samples of the data set, applies PAM on each sample and return its best clustering as the output. Clara clustering algorithm is applied to the processed dataset to group similar factors. The proposed method applies the eclat algorithm both on the original dataset and clustered data to compare the patterns in road accidents.

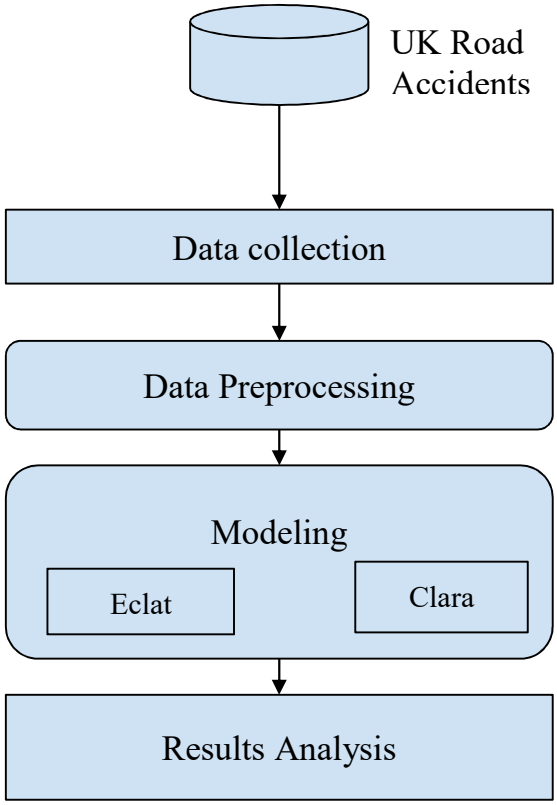

Fig.1 System Architecture

4. Results

Association rule mining is applied to original and clustered data to compare the rules that are involved in road accidents. A number of rules generated before clustering is given in table [1].

Table [1]. No of rules generated by the algorithms

Algorithm	No of Rules
Apriori	9864
Eclat	2089

The tables clearly indicating that Eclat is producing less no of rules in finding hidden patterns.

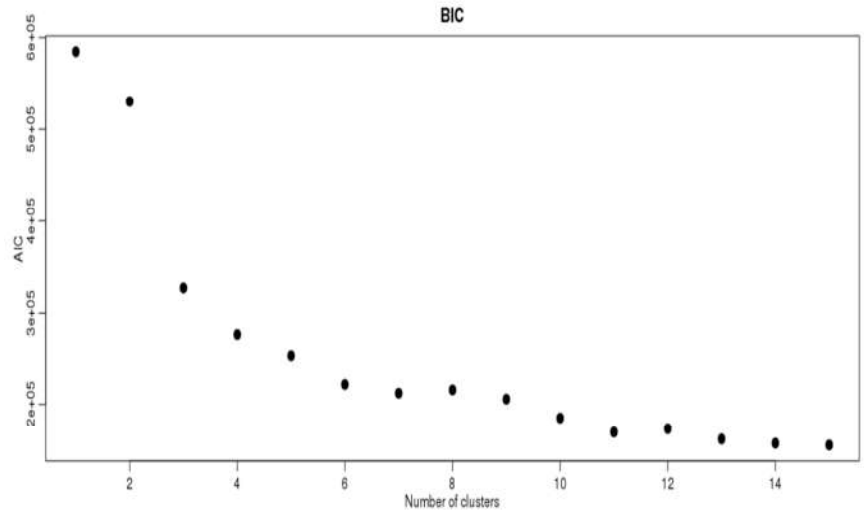

Fig.2. Plot for finding no of clusters using BIC

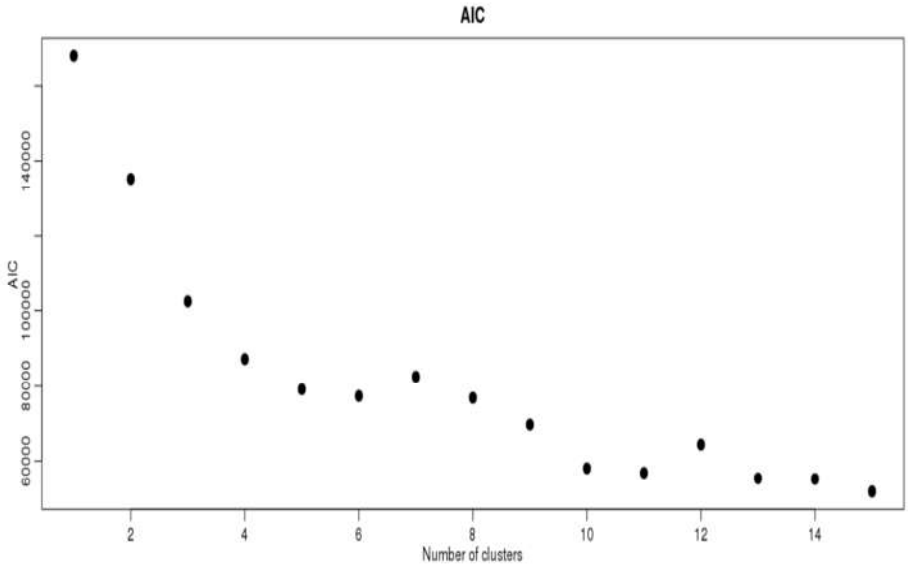

Fig.3. Plot for finding no of clusters using AIC

The number of rules generated after clustering is given in table [2].

TABLE [2]. Important rules generated after clustering

Algorithm	No of Rules
Apriori	35
Eclat	19

The study of the tables indicating that eclat is outperforming the apriori algorithm while finding the patterns in accidents data. The tables indicating that clustering improving the generating rules by removing the noisy data.

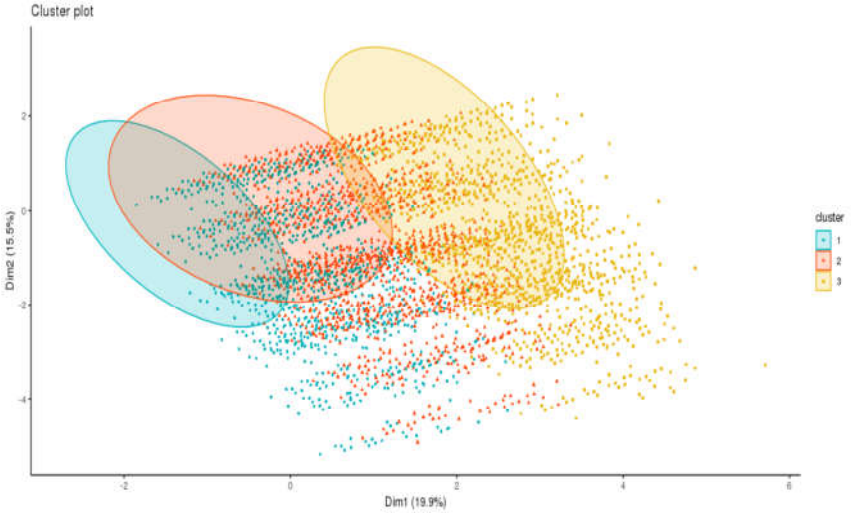

Fig.4. Clustering using clara

5. Conclusion

The proposed system implemented the Eclat and Clara clustering methods to find the most affecting rules that are contributing the road accidents. The result of the study indicating that a combination of clustering and association rule mining algorithms providing the best results in finding the patterns involved in road accidents data.

6. References

[1] Sachin Kumar, Durga Toshniwal, "A data mining framework to analyze road accident data", 2015.

[2] Taamneh, MadharAlkheder, "Data-mining techniques for traffic accident modeling and prediction in the United Arab Emirates", 2017.

[3] Living Li, Sharad Shrestha, "Analysis of road traffic fatal accidents using data mining techniques" in 15th International Conference on Software Engineering Research, Management, and Applications,2017.

[4] Apeksha V. Sakhare, Prajakta S. Kasbe, "A review on road accident data analysis using data mining techniques" in International Conference on Innovations in Information, Embedded and Communication Systems, 2017.

[5] Muhyiddin ISLAM, KunnaweeKANITPONG, "Identification of factors in road accidents through in-depth accident analysis", 2008.

[6] Maninder Singh, Amrit Kaur "A Review on Road Accident in Traffic System using Data Mining Techniques", 2013.

[7] E. Suganya, S. Vijayarani, "Analysis of road accidents in India using data mining classification algorithms", 2017.

[8] Sachin Kumar, Durga Toshniwal "Analyzing road accident data using association rule mining", 2015.

[9] Ayushi Jain, Garima Ahuja, "Data mining approach to analyzing road accidents in India", 2016.

[10] Suraj D, Sandeep Kumar S, "A Survey on Analyses of Factors Related to Road Accidents Using Data Mining Techniques", 2018.

[11] Jamshid Sodikov, "Road Traffic Accident Data Analysis and Visualization in R",2018.