

Predictive approach of Association Rule Model for Analysis Road Traffic Accidents

Pooja Dixit¹, Ram Singh Kushwaha²

¹Research Scholar, Bundelkhand University Jhansi, UP.

²Assistant Professor, Bundelkhand University, Jhansi, UP.

Abstract

Nowadays, with the increasing use of private vehicles, the number of traffic road accidents is increasing day by day. In this study we have used data mining techniques to analyze traffic accidents. One of the most prominent issues of road traffic safety analysis is the association rule Mining technique. In this paper we have used apriori algorithm to find some rules of high lift, high support, and confidence. The factors of road accident have been analyzed by the algorithm and have tried to present a model to reduce the side effects of road accidents. The main objective of this paper is to build a strong, robust association model that help to find the frequent datasets pattern from the datasets.

Keywords: Association Rules, Apriori Algorithm, Data Mining Data Visualization Traffic Accident Analysis.

Introduction

Traffic accident is an important topic in the traffic engineering sector According to the 2018 global report, the number of road accident deaths has reached 1.35 million annually and the number of injuries and disabilities has reached about 50 million, apart from which the economic impact is also notable. Over a period of 1 year the cost of medical care and its productivity has decreased and deaths from injuries and motor vehicle truck accidents have also increased by more than 63 billion, thus it is necessary searching for work related to road accidents. And to reduce the mortality and accidents in it, the analysis of cause of accidents and some new research and models is needed to help reduce the number of deaths and accidents in road accidents. Some important data mining techniques are used to assist manufacturers in finding rules and creating new policies and road rules with some passive patterns. Association rule mining is an important technique among the basic algorithms of data mining. And is a very important and growing area of research used to find all the rules in a database that meet some minimum support and lack of minimum confidence [1].

That is why it is very important to continue the search for causes of traffic road accident which will help us reduce the number of road accident deaths, However the research that has been done on road accidents in recent years is not enough and also not effective because if

you look at the road accident records of the last 5 years, then the reduction in deaths and accidents has almost stabilized to regain the good progress rate of the road. It may be necessary to find the top states and cities in the pace of security development, giving road and traffic safety administrators a completely new approach. Any influential political, legislative resolution and influential investment requires in-depth investigation and strong or impressive analysis methods. as traditional statistical analyzes are based on testing samples over entire small-scale, The conclusions are therefore based on the analysts' prior assumptions. Data mining and knowledge discovery both take a different technique to analysis the data [8]. Data mining is a technique with the help of which we can confirm the present and discover some new and interesting hypothesis.

This is why it is used to discover obscure and unregulated facts and patterns. Research has studied the behavior and utility of data mining methods for road and traffic safety, as data mining is largely used in versatile sets. Data analysis is done on the basis of Exploratory and descriptive methods are emphasized in this study. The main objective of this study is to use association and item set mining techniques as well as to obtain novel and unheard and interesting facts from a strong clustering technique and road accident data with reasonable and expected patterns [2]. Among the basic techniques of data mining, Association Rule Mining is an innovative technology that is increasingly used in the field of research and is useful for finding all kinds of rules in the database. Apriori algorithm [3] is a powerful algorithm to mine a continuous item set according to the rules of data mining technique, which is used to derive the rules of two different steps means to obtain frequent item sets. Minimum support is used. Association Rule Mining is a single unit technique. Association rule mining is an important technique to detect frequent items and patterns from large databases. In this research paper, Association Rule Mining is working on traffic accident data sets to find some useful rules, besides high support and high lift of rules were found which help to get more clear interpretation of rules with network Can.

The major contributions of this research paper can be summarized as follows

1. By applying association rule mining, those characteristics and factors for traffic road accident are found which is the main reason for the accident.
2. Using network visualization techniques in association rule mining can easily illustrate traffic accident cars with the help of finding accident causes and providing valuable insights into reducing them [1].

Related Works

Different data mining techniques have been used in the literature survey to analyze the main causes of traffic road accidents that including data mining models, classification techniques based on decision tree, clustering techniques, association rule mining is. The main factors of road accident have been used to find patterns of road accidents related to these techniques and to find out the cause of accidents. The main patterns of road accidents include the age of

the driver, time of road accident, weather at the time of accident, type of vehicles in most accidents, types of accidents and accidents in areas where accidents are more severe, etc.. the severity of Road accidents can be categorized mainly as fatal major, minor or none major incidents [4].

According to Li et Al. [5], they used the data mining algorithms like Apriori Algorithm, Naive Bayes Classifier and Clustering Algorithm is applied for analysis for traffic road accident. The authors also analyzed the main attributes of accidents in their research like lighting conditions, weather type, road type or surface of accident. Chong et. al. [6] applied the severity of injury caused during traffic accidents through its model using machine learning algorithms. In addition, he trained the Hybrid Learning Approach, SVM, Decision Tree and considered implementing the Hybrid Model Decision Tree and Neural Network Algorithms and Neural Network.

According to the Ona et. al., [7], Analysis of an accident on a rural highway in Spain presented an analysis of accidents in which they classified the accident according to the severity of the injury caused by the accident. in [8] research paper, in Abu Dhabi A model was presented to estimate the severity of the traffic accidents. In addition, Kumar and Toshniwal [9] proposed to implement association rule mining algorithm to analyze the accident data in traffic road accident, which contain many factors of accident like the age of the victim, gender, day, month, region, road type and area.

Atnafu and kaur [10] examined the data mining tools used in it to predict the level of road accident severity. According to them the technique were surveyed under the categories of Data Mining Techniques, Classification, Association Rule Mining and Data Mining Tools. The head of National Highways Authority of India (NHAI) used for the data section also analyzed the problem of mining accumulated data with the use of association rule mining. They also found patterns related to it based on item combinations. the main purpose of this research paper is to reduce its computational time and memory requirement with the highest composition. Here the main reason for using Association Rule Mining Technique is discover the hidden patterns based on the enormous combination of the individual and conditional frequencies

Methodology

➤ Association Rules

Association rule mining is the most popular algorithm of data mining, which is used to extract potential patterns from large data sets. Association rule mining is an adaptable technique among other data mining techniques because it does not depend on a particular function and dependent variable. This algorithm is used to reduce the causes of accidental accidents directly and to make quick decisions. For example, if consider some accidents rules {weather = rainy, light = dark, time = 22-24}) {accident = rollover}. Here this rule is

indicating that if the time in the rainy day is between 22 to 24 and there is more darkness on the road, then it is more likely to roll-over in a road accident. Thus we can conclude that if the road lights are properly arranged then the possibility of traffic road accident can be reduced. In this paper, the Frequent Pattern of Accidents is extracted using the Apriori algorithm for association rule [4].

➤ Apriori Algorithm

The Apriori Algorithm was proposed by Agarwal et. al., who follows the rules of association rule mining according to transaction of data. it is a popular algorithm used in Market Basket Analysis.

Apriori algorithm, an algorithm for data mining that is used to search very large databases [11]. It is a general simple priority-based algorithm that contains the most of all items that are taken in a frequent use [12]. It is based on the concept of Support Confident and Lift, which is as follows:

$$\text{Support (Item } I) = \frac{(\text{Total Number of transactions with Item } I)}{(\text{Total number of transactions})}$$

$$\text{Confidence (Item } I_1 \rightarrow \text{Item } I_2) = \frac{\text{No. of transactions with } I_1 \text{ and } I_2}{\text{No. of transactions with } I_1}$$

$$\text{Lift (Item } I_1 \rightarrow \text{Item } I_2) = \frac{\text{Confidence}(\text{Item } I_1 \rightarrow I_2)}{\text{Supprt}(\text{Item } I_2)}$$

There are a few steps to work on the Apriori algorithm which are as follows.

- Step 1:** The minimum support and confidence defined to apply for association rule.
- Step 2:** The second step is to select all subsets in the transaction, with higher support greater than the minimum support.
- Step 3:** In this type, a sub-set with maximum confidence is selected from the rules.
- Step 4:** in the step, the lift is sorted in decreasing order according to the association rule mining.
- Step 5:** And in this last step, the rules are visualized with confidence and support.

Model Design for Analysis the Factor of Traffic Accident

In this section, we create a model to analyze the factors of road accident analysis which are explained in various phases. Figure 1 analyzes the factors of traffic accident. First step is to define the problem of traffic road accidents and used for analysis. This paper define this problem as an analysis of the factors of traffic accidents and their rules. The second step is for

the analysis of accident data. For this study, we have used the traffic road accident data set. The next phase which is the third phase, in this we follow the preprocessing step based on the properties of the collected data. and In the fourth step, we analyze the data using some statistical method and Apriori algorithm [5].

Traffic Accident Factor Analysis

In this section, we analyze the main factors of traffic road accident using different approaches, according to the model formulated in section IV.

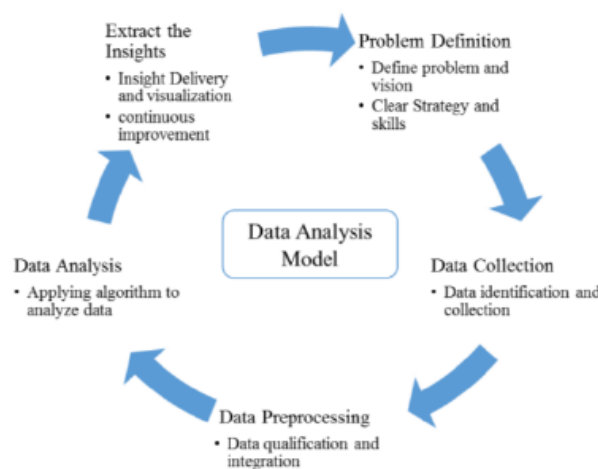


Figure 1.A data analysis model

➤ Data Cleaning

In this section we describe the data cleaning step, for that we have first described the data taken in the analysis and then applied that data preprocessing step [13].

➤ Data Transformation

Data transformation has implemented by using time factor along with age and intoxication. Some of the data analysis steps below are used to transform data using the market basket analysis technique. At this stage The time in this phase refers to the time of the traffic road accident, which is categorized in Three-hour intervals. The age is classified into the three classes young class, middle age class and the elderly class, in which the young are less than 35, the middle age group in between 35 to 60, and the elderly are classified in the category of over 60. The intoxication attribute is classified with Yes and No. The value of the intoxication characteristic is determined on the bases of consuming alcohol and drugs.

➤ Feature Extraction

Feature Extraction Technique is used to extract new features based on existing features. According the Time of Road Accident We have obtained two features One feature is day or

other is night. Accident Type which derived three attributes Hit Vehicle, Hit Sidewalk, Run Over. Cause of accident which attributes are Alcohol Influence, Sudden deviation, Did not leave enough space, Lack of estimation for road users. age of victim are classified into Youth, Middle age group and elder group.

Results and Discussion

From the data set it can be seen and understood that the main cause of road accidents is hit collision between vehicles and over speed. The main causes of accident are driving due to the effects of alcohol and lack of adequate space between vehicles and Sudden deviations are included, besides accidents occurring during different time intervals are also the main cause which is shown in Fig. 1.

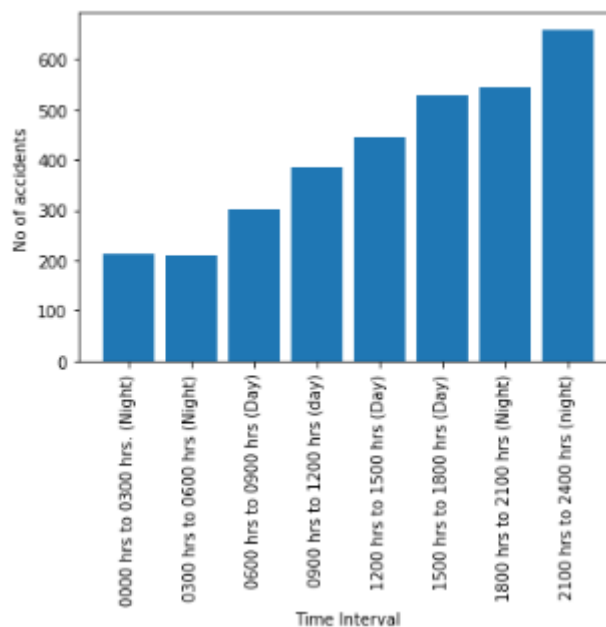


Figure 2. Number of accidents per time interval

It is clear from Figure 2 that road accidents occur more at night than at day. Using the Apriori algorithm, we are able to identify the frequent patterns related to the accident and the age and characteristics of the victim. By doing this, we can find out the main cause of the accident, possibly Table 1 showing frequent patterns in the context of the accident [4].

Table 1. Most Common Patterns of Accidents

S. No	Accident Cause	Accident Type	Inoxicated	Age
1	Did Not leave Enough Space	Hit Vehicle	Yes	Youth
2	Lack of estimation for road users	Run Over	Yes	Youth
3	Sudden Deviation	Hit Vehicle	No	Middle Aged
4	Did not leave enough space	Hit Sidewalk	Yes	Youth
5	Alcohol Influence	Run Over	Yes	Youth

It is becoming clear from the table1, that most of the accident involved youth. The main reason for the collision of vehicles has been found to be not having enough space between vehicles. as young and middle class people do not pay attention on leaving enough space in the middle of the vehicle.

This picture is clear from figure 2 is that maximum road accidents occurred in between 10:00 pm and 12:00 p.m. Characteristics of accidents at the time of extreme accident are tabulated shown in Table 2.

Table 2. Most Common accidents patterns during peak accident time

S. No	Accident Cause	Accident Type	Inoxicated	Time
1	Did not leave enough space	Run Over	Yes	Night
2	Alcohol Influence	Run Over	Yes	Night
3	Did not leave enough space	Hit Vehicle	Yes	Night
4	Lack of estimation for road users	Run Over	No	Day
5	Sudden deviation	Hit Vehicle	No	Day
6	Did not leave enough space	Hit Sidewalk	Yes	Night

From Table 2, it is becoming clear that maximum accidents in road accidents are caused in night which include most of the youth. And the main reason for this has been found to consume alcohol among the youth. This makes it clear that despite the age of the driver, the effect of alcohol is an important reason for a road accident.

Table 3. Association rules with high Support

S. No	Rules	Support	High Confidence	Lift
1	Run Over -> Lack of estimation for road users-> Day	0.56	0.593	3.50
2	Hit Vehicle-> Sudden deviation -> Day	0.76	0.613	3.59
3	Run Over-> Lack of estimation for road users-> night	0.65	0.667	3.75
4	Hit Vehicle-> Sudden deviation-> Night	0.67	0.671	3.19
5	Hit vehicle-> Did not leave enough space -> Day	0.71	0.489	3.29
6	Run Over-> Did not leave enough space-> Day	0.54	0.518	3.96
7	Run over-> Alcohol Influence-> Night	0.65	0.733	3.54
8	Hit sidewalk-> Alcohol Influence -> Night	0.56	0.529	3.09
9	Hit Vehicle -> Sudden deviation -> Night	0.78	0.760	3.98

To find out the interesting rule from the Road Accident data, we applied the association algorithm which is implemented in Python script and found the patterns from the data sets. For doing this we set minimum support and minimum confidence 0.003 and 0.5 respectively. And min lift is set by with 3. Table 3 shown the Association rules with high Support, Confidence and lift.

Conclusion

Road accident deaths are one of the leading causes of total deaths worldwide. By analyzing various features associated with road accidents, the major factors of accident have been identified by this study. Data about accident-related patterns have been collected by the Apriori algorithm. This algorithm is used to find out the data related to the road accidents with respect to the characteristics such as type of accident, cause, age of the person, road type. After analyzing the traffic road accident data, it can be inferred that among the major causes of the accident is the increasing intoxication among the youth, due to which the rate of traffic road accident is increasing, in addition to that it has been observed that as compared to the day, In the night the accident rate is high.

References

1. Mingchen Feng, Jinchang Ren, Yue Xi “Association Rule Mining for Road Traffic Accident Analysis: A Case Study from UK” February 2020, DOI: 10.1007/978-3-030-39431-8_50
2. Sami Ayramo, Pasi Pirtala, Janne Kauttonen, Kashif Naveed, Tommi Kärkkäinen “Mining road traffic Accidents” Reports of the Department of Mathematical Information Technology Series C. Software and Computational Engineering No. C. 2/2009
3. Bhandari, A., Gupta, A., Das, D.:” Improved apriori algorithm using frequent pattern tree for real time applications in data mining” *Procedia Comput. Sci.* 46, 644-651 (2015).
4. Maya John, Hadil Shaiba “Apriori-Based Algorithm for Dubai Road Accident Analysis”, 16th International Learning & Technology Conference 2019, *Procedia Computer Science* 163 (2019) 218-227.
5. L. Li, S. Sherestha and G. Hu, "Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques," in *Proceeding of the Software Engineering Research, Management and Applications*, London, July 2017.
6. M. Chong, A. Abraham and M. Paprzycki, "Traffic Accident Analysis Using Machine Learning Paradigms," *Information*, vol.29, pp.89-98, 2005.
7. "J. Oña, R. Mujalli and F.J. Calvo, "Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks," *Accident Analysis and Prevention*, vol.43, no.1, pp.402-411, Jan. 2011.
8. M. Taamneh, S. Alkheder and S. Taamneh, "Datamining techniques for traffic accident modeling and prediction in the United Arab Emirates," *Journal of Transportation Safety & Security*, vol.9, no.2, pp.146- 166, 2016.

9. S. Kumar and D. Toshniwal, "Analysing road accident data using association rule mining," in Computing, Communication and Security (ICCCS), 2015 International Conference on, 2015, pp. 1-6.
10. B. Atnafu and G. Kaur, "Survey on Analysis and Prediction of Road Traffic Accident Severity Levels using Data Mining Techniques in Maharashtra, India," International Journal of Current Engineering and Technology, vol. 7, pp. 1974-1978, 2017.
11. Lantz, Brett,(2015) "Finding Patterns - Market Basket Analysis Using Association Rules" Machine Learning with R, Packt Publishing.
12. Dietrich, David, Barry Heller and Beibei Yang (2015) "Advanced Analytical Theory and Methods: Association Rules", in Carol Long and Kelly Talbot (eds) Data Science and Big Data Analytics: Discovering, Anlayzing, Visualizing and Presenting Data, John Wiley and Sons.
13. Sun-Young Ihm,Young-Ho Park "Analysis of Traffic Accident using Association Rule Model" Journal of Multimedia Information System VOL. 5, NO. 2, June 2018 (pp. 111-114): ISSN 2383-7632(Online) <http://dx.doi.org/10.9717/JMIS.2018.5.2.11>.