INTERNATIONAL RESEARCH JOURNAL OF ENGINEERING AND TECHNOLOGY (IRJET) E-ISSN: 2395-0056
VOLUME: 06 ISSUE: 07 | JULY 2019 WWW.IRJET.NET P-ISSN: 2395-0072

Algorithms for the Prediction of Traffic Accidents

Kavitha D N[1], Dr. Anil Kumar K M[2]

[1]*PG Student, Dept. of CS&E, JSS S&TU, Mysore, India*
[2]*Associate Professor, Dept. of CS&E, JSS S&T U, Mysore, India*

Abstract — *The study of factors influencing the traffic accidents is a important research in the field of traffic safety. Here, (for e.g.,) the traffic accidents of state region were excavated using association rule that generated various frequent item sets. The strong hidden rules in these frequent item sets usually uncover the association between factors influencing the traffic accidents, which may be used to reduce the occurrence of accidents by breaking them. The hidden rules can be used probe usual scenes of accidents and a few corresponding security improvement measures can be taken to prevent the accidents, and ultimately improve the city's traffic safety level. General speaking, association rule can produce lots of weak rules, the study designed a way to calculate minimal Support value of training parameters, and further a way to extract strong hidden rules automatically. The results of the experiments show that the proposed methods in the paper are effective. Therefore, Algorithms for prediction of traffic accidents using association rules was finally established to promote the effective application of association rule on intelligent system.*

Key Words— **data mining, traffic safety, influencing factors of accidents, association rules, strong rules;**

I. INTRODUCTION

The rapid development of urban expressways not only brings the convenience of transportation, but also causes traffic safety problems. In Karnataka there are many traffic accidents that occurred and the cities need a safer traffic environment. Exploring the impact of the factors influencing the accidents and taking corresponding effective measures to reduce the occurrence of accidents is an urgent need. In recent years, there are studies on the impact of influencing factors of traffic accidents, mainly focusing on people, cars, roads or the environment. Some studies on driver's behavior and analyzed the characteristics of the process during changing the lanes to identify dangerous driving behaviors. There are also studies on the impact of road conditions on traffic accidents, they proposed a point that high and steep roadbed will undermine the traffic safety. Also other studies focused on the impact of weather or dynamic traffic flow on accidents. However, most of these studies focused on a single factor (people, cars, roads, and the environment) on the impact of traffic accidents.

With the development of data mining technology, scholars used a variety of data mining approaches in traffic safety research. Among them, the association rule was often used to analyze the relationship between the influencing factors of traffic accidents. The strong association rules can be used to find the strong hidden network in the accident data. To get them, we could measure the importance and credibility of the rules with the two thresholds Support and Confidence. The existing association rule generally determined the model parameters (such as the minimum Support, etc.) by repeated experiments.

For the massive results excavated, it is necessary for the experts to screen useful rules according to personal expertise manually. The method is inefficient and the subjective screening process cannot be translated into an objective algorithm, so it hinders the direct application of association rule in intelligent transportation system. Here, the proposed method calculates the minimum Support in the modeling parameters, and put forward a way to extract strong rules, or automatically filtering out the weak rules. Finally, we built up an algorithm using association rules which would better promote the practical application of association rule in existing intelligent transportation system.

II. LITERATURE SURVEY

Iran is one of the countries with a high rate of traffic crash fatalities and injuries. Over the last three years, traffic crashes in Iran resulted in 24 000 people (i.e. 3 persons per hour) on the average killed and around 240 000 injured annually [2], [4]. This fact motivated a collective of authors to identify the most important factors which affect injury severity of drivers involved in traffic crashes on these roads [14]. They used the crash data from the records of the Information and Technology Department of the Iranian Traffic Police from 2006 to 2008. Each accident was described by 14 attributes; the target was injury severity with 3 levels: no-injury, injury and fatality; records included more than 169 000 drivers. The factors selection was carried out on the basis of Variable Importance Measure (VIM) which is one of the CART (Classification and Regression Tree) method outputs. The results indicated the seat belt as the most important factor associated with injury severity of traffic crashes, and not use it significantly increases the probability of being injured or killed.

Similar approach was used by authors Chang and Wang [6] to generate the CART model to identify relationships between injury severity and driver/vehicle characteristics, highway/environmental variables and crash variables. By using 2001 crash data for Taipei (Taiwan), they identify the vehicle type as the most important variable for crash severity. Also, data from central Taiwan was used in the study of authors Yau-Ren Shiau et al. [13] to identify at first the most important factors affecting more than 2 400 traffic accidents in 2011 and after that the mining methods as Fuzzy Robust Principal Component Analysis, Back propagation Neural Network and Logistic Regression was applied to generate expected classification models. The best accuracy 85.89% was obtaining by combining the first two mentioned methods.

Nayak et al. [11] analyzed the crash and road data from the Queensland Department of Transport and Main Roads (Australia) that contained more than 42 000 records referring to years 2004 – 2007. Authors used a categorization into several levels of crash proneness (some roads, due to design or condition, have higher crash rates than others) inspired by work of another collective of authors [18]. For the experiment the authors used decision trees with chi-square test, regression trees using f-test, logistic regression, neural networks and Bayesian model. Decision trees showed better performance than other tested models.

Another study was dedicated to the road safety in Dubai. In the United Arab Emirates, there are about 600 people killed in car accidents each year; road traffic accidents are the second major cause of deaths [1]. Authors used more than 1 800 000 records from the year 2008 to 2010 with 19 attributes covering accident, driver, and road/vehicle conditions. The target attribute (the accident severity degree) contained five values: fatal, serious, moderate, minor and none. Final predictive models were created within WEKA supporting tools and following algorithms: J48 decision trees, Bayesian networks, PART and Multilayer Perceptron. The most accurate model was created by Multilayer Perceptron (more than 99%) and Bayesian network gave the highest speed to build the model (0.17 seconds).

Similar algorithms (J48 and Naive Bayesian) were applied by collective of authors on data produces by Transport department of government of Hong Kong [12]. This dataset contained more than 34000 records; genetic algorithm was used for feature selection; classification experiments were performed also in WEKA mining environment and J48 provided more accurate classification model to predict the severity of injury during traffic accident.

III. METHODOLOGY

This system is a real time application which is useful for government sector to reduce the number of traffic accidents. Traffic safety represents an important part of our lives, so it is necessary to continuously improve within all possible and available opportunities and resources.

Descriptive or predictive mining applied on historical data regarding occurred accidents together with different necessary data as weather or traffic conditions creates a remarkable alternative with potentially useful and helpful outcomes for all involved stakeholders.

The system describes one possibility of how to use the collected data about traffic accidents to frequent patterns and important factors causing different types of accidents.

The system uses Apriori, Apriori TID and SFIT algorithms for analysis and the comparison between these algorithms are done. The system takes dynamic data as input and it is a generic application.

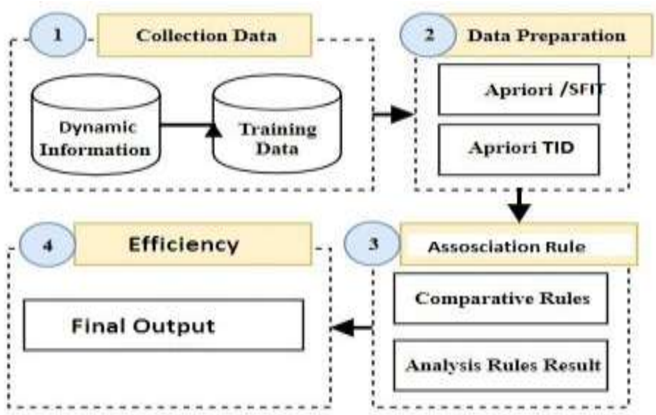

Fig- 1: System Architecture

- **Traffic Patterns Prediction (Association Rules)**

Here, system predicts accidents patterns based on the past traffic accidents details collected from various sources. Here system uses "Association Rule" to analyze previous accidents data and to extract the traffic accidents patterns. Patterns can be predicted based on road wise, city wise, date wise, month wise etc. This is the major objective of the project where we use data science to process the data, the traffic patterns predicted by the system useful to traffic departments and also publics to take some precautionary measures to reduce the number of traffic accidents.

1. **Apriori Algorithm**: Apriori algorithm is efficient and works well for any number of parameters and applied to many datasets

such as medical datasets, education dataset, agriculture dataset etc….

Algorithm Steps:

STEP 1: Scan the accident data set and determine the support(s) of each item.

STEP 2: Generate L1 (Frequent one item set).

STEP 3: Use Lk-1, join Lk-1 to get the set of candidate k - item set.

STEP 4: Scan the candidate k item set and generate the support of every candidate k – item set.

STEP 5: Add frequent item set, until C=Null Set.

STEP 6: For each item in the frequent item set generate all non empty subsets.

STEP 7: For all non empty subset determine the confidence.

If confidence is greater than or equal to the specified confidence, then add it to strong Association Rule.

2. **Apriori TID Algorithm (Apriori Transactional ID):** Apriori TID is an algorithm for discovering frequent itemsets in a transaction database. Apriori TID is an alternative implementation of the Apriori algorithm. It produces the same output as Apriori. But it uses a different mechanism for counting the support of itemsets.

Algorithm Steps:

STEP 1: Scan the accident data set and determine the support(s) of each item.

STEP 2: Generate L1 (Frequent one item set).

STEP 3: Determine candidate itemsets in Ck contained in the transaction with identifier i.e. [TID, Itemset].

STEP 4: Scan the candidate item set and generate the support of every candidate item set.

STEP 5: Add frequent item set, until C=Null Set.

STEP 6: For each item in the frequent item set generate all non empty subsets.

STEP 7: For all non empty subset determine the confidence.

If confidence is greater than or equal to the specified confidence, then add it to strong Association Rule.

3. **SFIT Algorithm (Set Operation for Frequent Itemset using Transactional database)**: This approach is a combination of Apriori algorithm and frequent itemset lattice. This algorithm is a combination of Apriori, and set operations. Principles of set operations, which are intersection and union, are used. These principles are related to frequent itemset tree. In frequent itemset tree, there are nodes holding frequent itemsets and transactions containing related itemsets. In order to construct k-itemsets, frequent (k-1)-itemsets are used. Frequent itemset union is formed for their support count and intersection operation is employed between the TID's of the itemsets.

Algorithm Steps:

STEP 1: Scan the accident data set and determine support(s) of each item.

STEP 2: Generate frequent itemsets (L1), and obtain transaction sets (TID), which includes the Itemset.

STEP 3: Generate set of candidate itemsets k from frequent itemsets and calculate support of each candidate itemset k.

STEP 4: Prune off the candidate itemsets k whose node count is lower than minimum support using their TID set.

STEP 5: Consequently, for each frequent Itemset, scan the accident data to approve the consistence of the Itemset until C=NULL SET.

STEP 6: Finally, Itemset are used to generate strong rules having minimum confidence in the frequent Itemset tree.

IV. EXPERIMENT AND RESULT

For the experiment we have used datasets of road accidents that took place in different cities. To study the performance of the algorithms, number of data sets, confidence and support threshold are used. The Tables shows that the execution time of the algorithms varies with varying support thresholds, confidence, and dataset.

TABLE 1: Comparison of Execution Time of Algorithms varying Support

Support	Constraint	Apriori	Apriori TID	SFIT
10%	**Execution Time**	600 milliseconds	203 milliseconds	100 milliseconds
20%		1125 milliseconds	230 milliseconds	120 milliseconds
30%		626 milliseconds	219 milliseconds	113 milliseconds

TABLE 2: Comparison of Execution Time of Algorithms varying Confidence

Confidence	Constraint	Apriori	Apriori TID	SFIT
60%	**Execution Time**	859 milliseconds	236 milliseconds	121 milliseconds
70%		629 milliseconds	231 milliseconds	127 milliseconds
80%		624 milliseconds	225 milliseconds	121 milliseconds

TABLE 3: Comparison of Execution Time of Algorithms varying Datasets

Constraint	Apriori	Apriori TID	SFIT
Execution Time	622 milliseconds	240 milliseconds	132 milliseconds
	859 milliseconds	236 milliseconds	121 milliseconds
	615 milliseconds	213 milliseconds	110 milliseconds

From the above tables, we find that the execution time of SFIT algorithm is less than the Apriori and Apriori TID algorithms. SFIT algorithm is more efficient than the Apriori and Apriori TID.

The efficiency of the system depends on the execution time of the algorithms. The execution time varies with varying support threshold, confidence and dataset.

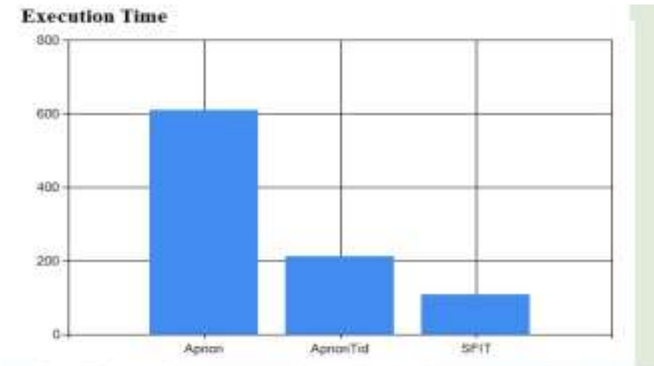

FIGURE 2: Graphical Representation of Result when Support varies

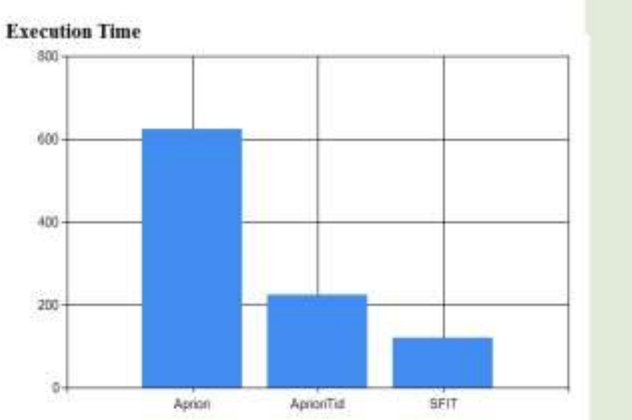

FIGURE 3: Graphical Representation of Result when Confidence varies

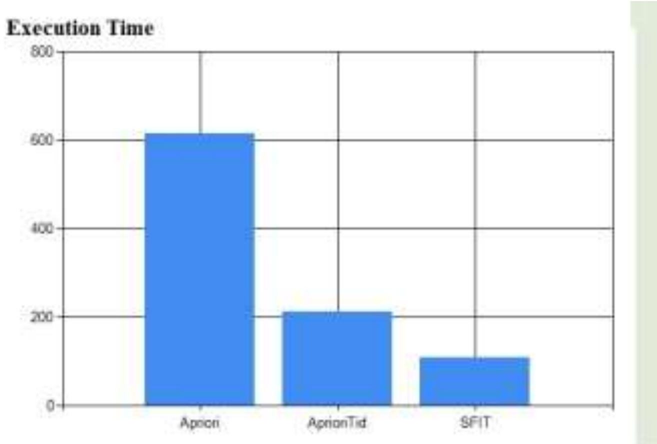

FIGURE 4: Graphical Representation of Result when dataset varies

V. CONCLUSIONS

In this paper, we used the association rules to analyze the relationship between the influencing factors of traffic accidents collected data, and proposed a minimal Support calculation method based on frequent item set. The results of the experiments show that the three methods are effective. Based on these methods, an algorithm for the prediction of traffic accidents using association rules is proposed, which helps to promote the application of association rule mining in the existing intelligent transportation systems. Through the strong rules which contain the association among influencing factors of traffic accidents, we can understand the accident scenes, so additional warning signs can be added to reduce the accidents, and ultimately improve the city's traffic safety level.

The proposed algorithms using association rules are not specific but generalized. It is not limited to traffic safety research; in fact it can be flexibly adopted in many other research fields. It is just empirically applied in the traffic safety domain and proved its effectiveness.

REFERENCES

[1] A Araar et al., "Mining road traffic accident data to improve safety in Dubai", Journal of Theoretical and Applied Information Technology, 47(3), pp. 911-927, 2013.

[2] A.T. Kashani, A. Ranjbar, A. Shariat-Mohaymany, "A Data Mining Approach to Identify Key Factors of Traffic Injury Severity", PROMET Traffic& Transportation, 23(1), pp. 11-17, 2011.

[3] EC: Mobility and Transport, Road Safety, Statistics – accidents data, online available on: http://ec.europa.eu/transport/road_safety/specialist/statistics/index_en.htm

[4] F.M.O.I. Forensic Medicine Organization of Iran; Statistical Data, Accidents, online avail-able on: http://www.lmo.ir/?siteid=1&pageid=1347.

[5] G. Nakhaeizadeh, J. Hipp, U. Güntzer, "Algorithms for Association Rule Mining &Mdash; a General Survey and Comparison", SIGKDD Explor Newsl 2, pp. 58–64, 2000.

[6] H.W. Wang, L.Y. Chang, "Analysis of traffic injury severity: An application of non-parametric classification tree techniques", Accident Analysis and Prevention, 38(5), pp. 1019-1027, 2006.

[7] J. Pendharkar, P.J. Ossenbruggen, JohnIvan "Roadway safety in rural and small urbanized areas", Accidents Analysis & Prevention, 33(4), pp. 485-498, 2001.

[8] L. Breiman, "Random Forests", Machine Learning, Vol. 45, pp. 5- 32, 2001.

[9] L. Martin, "Using data mining techniques to road safety improvement in Spanish roads", XI Congreso de Ingeniería del Transporte (CIT 2014), Procedia - Social and Behavioral Sciences 160 (2014), pp. 607–614, 2014.

[10] P. Flach et al., "On the road to knowledge: Mining 21 years of UK traffic accident reports", Data Mining and Decision Support: Aspects of Integration and Collaboration, Springer, pp. 143-155, 2003.

[11] R. Nayak, "Road Crash Proneness Prediction using Data Mining". Ailamaki, Anastasia & Amer-Yahia , Sihem (Eds.) Proceedings of the 14th International Conference on Extending Database Technology, Association for Computing Machinery (ACM), Upp-sala, Sweden, pp. 521-526, 2011.

[12] S. Vigneswaran, E.Rajamanickam, A.Arun Joseph, "Efficient Analysis of Traffic Accident Using Mining Techniques", International Journal of Software and Hardware Research in Engineering, Vol. 2, No. 3, 2014, pp. 110- 118, 2014.

[13] S. Yau-Ren Shiau, Yung-Hsiang Hung,Yu-Ting Kuo,Ching-Hsing Tsai,"The Application of Data Mining Technology to Build a Forecasting Model for Classification of Road Traffic Accidents", Mathematical Problems in Engineering, Volume 2015 (2015), pp. 1-8., 2015.