VOL. 14, NO. 12, JUNE 2019 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

PREDICTION OF ROAD ACCIDENT LOCATIONS IN ROAD ACCIDENT DATABASE BY MINING SPATIO-TEMPORAL ASSOCIATION RULES

Arun Prasath N and M. Punithavalli
Department of Computer Application, Bharathiar University, Coimbatore, India
E-Mail: arunphd25@gmail.com

ABSTRACT

According to the World Health Organization (WHO), road accidents are regarded as one of the leading causes of death. The trend of a road accident can change in future as it is hard to predict the rate at which road accidents are taking place. The road accident leads to an unacceptable loss in terms of property, health and other economic factors. There are instances where road accidents occurred more frequently at a specific location. Some of the road accident features influence road accident to occur frequently. So, it is essential to identify the correlation in various attributes of road accident for predicting road accident. Data mining techniques are widely used to find the correlation in various attributes of the large database. A data mining approach was proposed to characterize road accident locations. In this approach, the Apriori algorithm was applied to characterize locations by generating rules. The Apriori algorithm has high space and time complexity problem and it is also costlier process owing to a large database. In this paper, Frequent Pattern-growth (FP-growth) is introduced for road accident prediction. In FP-growth, the larger feature space is condensed into smaller sub-spaces so that the costly repeated scans are avoided. Then the attributes with high confident values are trained by a decision tree classifier called as J48. It trains and classifies the data as critical and non-critical accident type. Hence by using FP-growth space and time complexity of association rule mining based road accident prediction is reduced and its accuracy is improved by using J48 classifier.

Keywords: road accident location prediction, association rule mining, apriori algorithm, frequent pattern growth algorithm.

INTRODUCTION

Globally road accident [7] is the eighth leading cause of death. Even though there is a commandable improvement in the road and vehicle safety with added ingredients of high class technology standards, still the total number of road accidents has been increased over the last decades. Therefore, finding out ways of minimizing the severity and frequency of traffic accident is of utmost importance. Moreover, road accident leads to a large number of fatalities, injuries and economic loses. Factors affecting accident severity and frequency are mainly related to accident characteristics, atmospheric factors, driver characteristics, vehicle characteristics and highway characteristics.

Accident data forms enormous database that covers different accident attributes. Many analytical methods have been used to analyze the accident database. Data mining technique [2] is one of the recent methods in this domain. Data mining uses various tools to analyze accident data including database technology, statistics, pattern recognition, data visualization, high-performance computing, information retrieval and neural networks. These models can determine the interactions between the variables which would be established directly using ordinary statistical modeling techniques.

A data mining approach [5] was proposed which characterize locations of road accidents. K-means algorithm was applied on the collected road accident data which groups the road accident location into three categories they are low-frequency, moderate-frequency, and high-frequency accident locations. It clusters the road accident data based on the threshold values owing to this effect Apriori which is an association rule mining technique, used to reveal different factors associated with road accidents at a different location with varying accident frequencies. However, time and space consumption of Apriori is high as well as it is costlier process too especially when there is a large number of patterns.

In this paper, an efficient association rule mining technique called Frequent Pattern (FP-growth) is introduced to improve the prediction of road accidents. The larger feature space of accident data is reduced into smaller sub-spaces so that the costly repeated scans are avoided. After the prediction of road accident by using FP-growth, the attributes with high confident values are trained by a decision tree classifier called J48 which classifies the accident type as critical and non-critical.

LITERATURE SURVEY

In [6], a deep learning approach was proposed for analysis of traffic accident risk prediction. In this approach, the spatial and temporal patterns of traffic accident frequency were analyzed in the collected big traffic accident data. From the analysis, spatiotemporal correlations of traffic accidents were obtained. Moreover, based on these patterns a high accurate deep learning model related to are current neural network was proposed for the prediction of traffic accident risk. However, this approach cannot provide the road level accident risk prediction.

A novel methodology [3] was presented to predict road accident occurrence. This methodology uses a combination of hierarchical multivariate Poisson-lognormal regression analysis, gamma-updating and Bayesian inference algorithm. The hierarchical multivariate Poisson-lognormal regression analysis is used to analyse the data and the gamma updating updates the data about causalities in the accidents. The Bayesian

inference algorithm was supported by Bayesian Probabilistic Networks (BPNs) which predicts the occurrence of road accidents. However, the efficiency of this methodology is low.

The severity of road accidents were analyzed using decision trees (DTs) [1]. Here, the DT method was used to analyse the severity of traffic accident. DT was constructed with the attributes of the traffic accident. The attribute with high information forms a root node of the DT and based on the splitting criteria the branches of the DT formed. The Decision Rules (DRs) were generated by traversing the DT from the root node to the leaf node. The DRs were used to identify safety problems. A specific method was proposed to extract all the knowledge from a particular dataset. Different DTs were built by varying the root node which gets every possible set of DRs from each tree. However, the space complexity of this method is high.

A traffic accident model [4] was developed using Artificial Neural Network (ANN) to predict the road traffic accident in Jordan. Initially, data such as a number of registered vehicles, a total length of paved roads, population and the gross domestic product were collected and those were split into three sets such a straining data, validation data, and testing data. The training data was trained by using ANN. The hidden layer of ANN uses Tan-sigmoid transfer function and predict the road traffic accident in the output layer of ANN. However, ANN required greater computational resources.

A model system [10] was presented to predict the duration and severity of traffic accidents. The severity of traffic accidents were predicted by Ordered Probit model and the duration of the traffic accidents would be predicted by Hazard model. By using these models important aspects such as number of injuries, number of fatalities, damage to the properties and accident duration were predicted. The important influences of related variables were identified. In this approach the characteristics of the passenger, pedestrian, driver and traffic condition have potential effects on accident duration and severity which were not considered.

A Fuzzy Neural Network Model (FNNM) [9] was established to predict the road accident frequencies. In FNNM, an optimal number of fuzzy sets for input variables of training data was determined by using k-means clustering. The mean silhouette value was considered as criteria to find out the number of clusters. Then the fuzzy inference system was generated after trying both sub-clustering method and grid partition method. Through this process, effective fuzzy rules were determined. Then, the back propagation method was used to train the model. Finally, the optimal model was tested. However, the consumption time for the generation of fuzzy inference system is high.

A new combined model based on the Induced Ordered Weighted Geometric Average (IOWGA) operator [8] was proposed for traffic accident prediction. This prediction model was the combination of GM (1, 1) model and the Verhulst model with changeable weight coefficients of every single model. In addition to this, the combined model was based on the Optimal Weighted (OW) method which was also presented for traffic accident prediction. The selection of best suitable single method with respect to the real application problem is not addressed in the combined model.

PROPOSED METHODOLOGY

In this section, the proposed methodology for road accident prediction using FP-growth and J48 classifier are described in detail. FP-growth is used to categorize the road accidents by generating the association rules. After the prediction of road accidents, a J48 classifier is applied to find the critical and non-critical accident types.

Frequent Pattern Growth based Road Accident Prediction

Association rules represent an effective and convenient way to identify certain dependencies between attributes in a database. A set of accident data items is denoted as $A = \{A_1, A_2, \dots A_m\}$ and a database of transactions $D = \{t_1, t_2, \dots t_n\}$ where $t = \{A_{i1}, A_{i2}, \dots A_{ik}\}$ and $A_{ij} \in A$, an association rule is an implication of the form $S, T \subset A$ are set of accident data items called itemsets and $S \cap T = \emptyset$. FP-growth is one of the very effective algorithms used to mine association rules. This algorithm generates frequent itemsets and it creates a huge amount of candidate itemsets like Apriori algorithm. The accident database is pre-processed by scanning the whole database. It determines the frequency of all items in the database thereafter infrequent items are discarded from each transaction. After the removal of infrequent itemsets, the items are sorted based on the frequency of itemsets and then FP-Tree is constructed. It is a highly compact representation of the original database. The following algorithm is used to construct FP-Tree.

Construction of FP-Tree Algorithm

Input: Accident database and a minimum support threshold γ

Output: FP-Tree

1. Scan the accident database once.
2. Collect the set of frequent items $freq$ and their supports.
   //Support is the ratio of the number of records contains $S \cup T$ to the total number of records in the database.
3. Remove items which are lesser than γ from each transaction of the database.
4. Sort the frequent items in a descending order of their support values and make it as the list of frequent items $List$.
5. Create a root node of FP-Tree $Tree$ and label it as null for each transaction in the database.
6. According to the order of $List$, select and sort the frequent item list in the transaction.

7. Let the sorted frequent item list in the transaction be$[m|M]$
   // m is the first element and M is the remaining list.
8. Call ins_tree$([m|M], Tree)$
   //Procedure ins_tree$([m|M], Tree)$
9. if $Tree$ has achild N such that $N.item-name = m.item-name$
10. N++
11. else
12. Create a new node N
13. Count (N) = 1
14. N's parent link be linked to $Tree$
15. N's node-link be linked to the nodes with the same item-name via the node-link structure
16. End if
17. if M is non-empty, call insert_tree(M, N)

By using the above algorithm, FP-tree is constructed and then all frequent itemsets are generated using the following FP-growth algorithm.

FP-growth Algorithm

Input: FP-tree, accident database and a minimum support threshold γ

Output: Complete set of frequent patterns
1. Call FP-growth $(FP-tree, null)$
   //FP-growth $(Tree, \alpha)$
2. if $Tree$ contains a single path M then for each combination (β) of the nodes in the path do
3. Generate pattern $\beta \cup \alpha$ with support = minimum support of nodes in β.
4. else for each a_i in the header of Tree (in reverse order) do
5. Generate pattern $\beta = a_i \cup \alpha$ with support $=a_i.support$
6. Construct $\beta's$ conditional pattern base and then $\beta's$ conditional FP-tree $Tree\ \beta$
7. if $Tree\ \beta \neq \emptyset$
8. then call FP-growth $(Tree\ \beta, \beta)$
9. End if
10. End if

The above FP-growth algorithm is accomplished by traversing from the bottom node of FP-tree to root node. It Checks each level of node which has single path while traversing at each level of the tree. If the node does not have a single path, at first conditional pattern bases of that node is determined, then all infrequent items are discarded from the determined pattern bases and finally, FP-tree is constructed. This process is continued till a single path is found. If a single path is determined, all combinations of the path including the candidate node is determined. These combinations are the desired frequent itemsets of the desired node. After getting frequent itemsets, association rules are generated whose confidence value is greater than a threshold value. In case of road accident data, an association rule identifies the attribute value which is responsible for accident occurrence.

J48 Classification algorithm

From the FP-growth based association rule mining, the road accidents are categorized as high-frequency accident location, moderate-frequency accident location and low-frequency accident locations. After the categorization, it is paramount to find the critical and non-critical accident type. The attributes with high confident values are given as input to the J48 classifier. J48 classifier is used to train the road accident data. The J48 algorithm develops its decision tree depending on the information of the theoretical attribute values of the present training data. Every attribute of road accident data estimate the gain value and this calculation process is continued till the prediction of critical or non-critical type of accident process is completed. A feature which gives a lot of information regarding the data instances is classified as a root node. This node contains the maximal information gain. The information gain of an attribute $Atr$and a sample S is calculated as follows:

$$Entropy\ (P) = -\sum_{i=1}^{n} p_i log_2 p_i \quad (1)$$

where, $P = (p_1, p_2, \ldots p_n)$ is a probability distribution and n is the number of attributes in road accident data.

$$\text{Information gain (S, A)} = \text{Entropy (S)} - \sum_{i=1}^{n} s_j \text{Entropy}\left(s_j\right) \quad (2)$$

where, s_j denotes the set of all possible values for attribute A. The process of J48 is given as follows:

J48 Algorithm

Input: training data

Output: Leaf node (critical or non critical type)
1. Create root node $node$
2. if (Tbelongs to same category C) //T is the instance of road accident data
3. leaf node = $node$
4. Mark $node$ as class C
5. Return $node$
6. End if
7. For i=1to n
8. Calculate information gain of each attribute A_i using equations (1) and (2).
9. End for
10. $test_A$= testing attribute
11. for (each T in splitting of T)
12. if (T is empty)
13. child of $node$ is a leaf node

14. else
15. child of $node = dtree\ T$
16. End if
17. End for
18. Calculate classification error rate of node $node$
19. return $node$

RESULT AND DISCUSSIONS

In this section, the effectiveness of proposed FP-growth with J48 based road accident prediction method is tested in terms of accuracy, precision, and recall. For the experimental purpose, the data about the road traffic accident is collected from traffic police in the area of Coimbatore. The data collected from the period of1st January 2013 to 30th November 2013. It consists of different attributes such as date and time of the accident, road name, maintained by whom i.e. whether C&M, NH or municipality, number of accidents by means of severity (fatal, grievous, minor injury, only damage to the property), nature of the accident, reason as stated in FIR and place name.

Accuracy

Accuracy is defined as the proportion of the true outcomes (both true positives and true negatives) among the sum of cases observed. It is calculated as follows:

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}$$

where, if the class label is positive and the road accident prediction outcome is positive then it is True Positive.

If the class label is negative and the road accident prediction outcome is negative then it is True Negative.

If the class label is negative and the road accident prediction outcome is positive then it is False Positive.

If the class label is positive and the road accident prediction outcome is negative then it is False Negative.

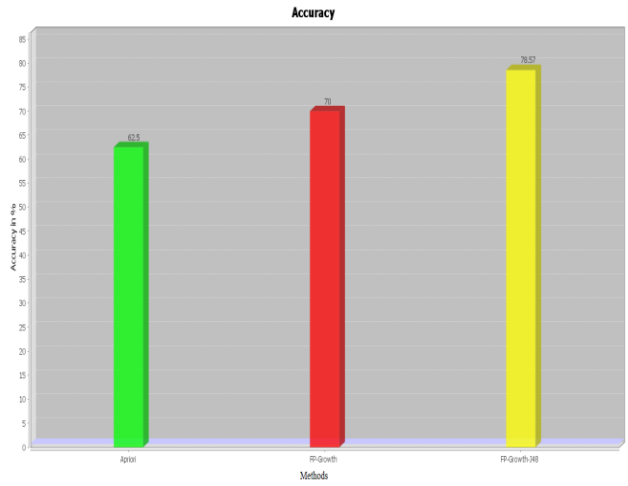

Figure-1. Comparison of Accuracy.

Figure-1 shows the comparison of accuracy between existing Apriori-based road accident prediction and proposed FP-growth with J48 based road accident prediction method. The different road accident prediction methods are taken in X-axis and the accuracy is taken in Y axis. The accuracy of the proposed FP-growth-J48 is 25.7% greater than Apriori and 12.2% greater than FP-growth based road accident prediction. From this, it is clear that the proposed FP-growth with J48 based road accident prediction method has high accuracy than any other methods.

Precision

Precision is defined as the computation of exactness or quality, and is measured as follows:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

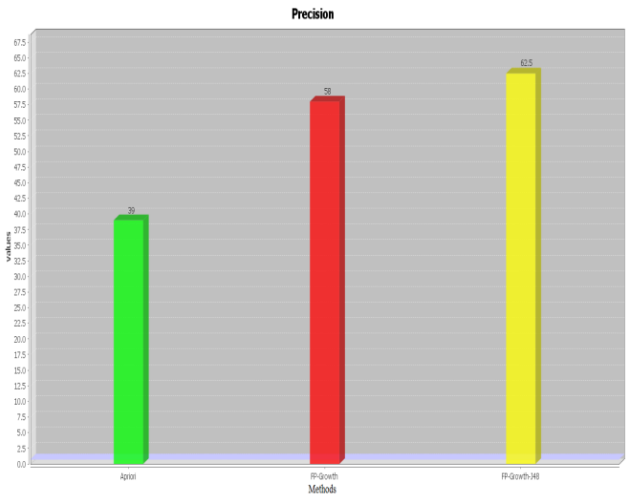

Figure-2. Comparison of Precision.

Figure-2 shows the comparison of precision between existing Apriori-based road accident prediction and proposed FP-growth with J48 based road accident prediction method. The different road accident prediction methods are taken in X-axis and the precision is taken in Y-axis. The precision of the proposed FP-growth-J48 is 60.2% greater than Apriori and 7.8% greater than FP-growth based road accident prediction. From this, it is clear that the proposed FP-growth with J48 based road accident prediction method has high precision than any other methods.

Recall

Recall is defined as the number of true positives divided through the sum of elements which successfully belong to the positive class. It is denoted as follows:

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

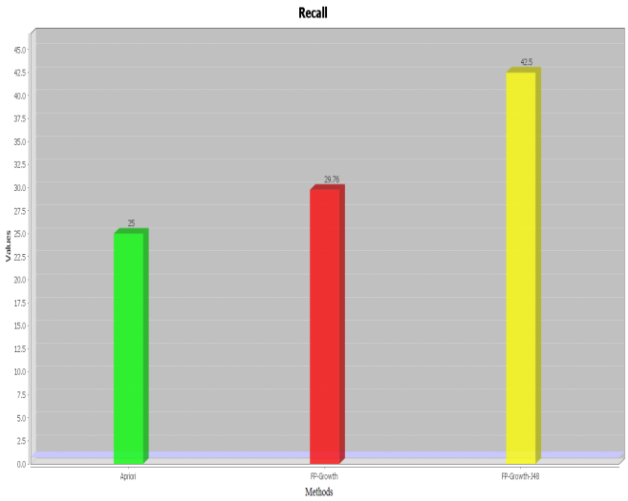

Figure-3. Comparison of Recall.

Figure-3 shows the comparison of recall between existing Apriori-based road accident prediction and proposed FP-growth and FP-growth with J48 based road accident prediction method. The different road accident prediction methods are taken in X-axis and the recall is taken in Y-axis. The recall of the proposed FP-growth-J48 is 70% greater than Apriori and 42.8% greater than FP-growth based road accident prediction. From this, it is clear that the proposed FP-growth with J48 based road accident prediction method has high recall than any other methods.

CONCLUSIONS

In this paper, a data mining technique based on road accident prediction is improved by using an efficient association rule mining technique called FP-growth and J48 classifier. FP-growth categorizes the collected road accident data by generating association rules. The road accident data are categorized as high-frequency accident location, moderate-frequency accident location and low-frequency accident location. Further it is more important to know whether the accident is critical accident type or non-critical accident type. It is achieved by training the attributes of road accident data which has high confident value by using theJ48 classifier. Hence the proposed FP-growth with J48 classifier reduces the high space and time complexity as well as it provides high accuracy for prediction of the road accident. The experimental result proves that the proposed road accident prediction method has high accuracy, precision, and recall.

REFERENCES

[1] Abellán J., López G. and De OñA J. 2013. Analysis of traffic accident severity using decision rules via decision trees. Expert Systems with Applications. 40(15): 6047-6054.

[2] Alkheder S., Taamneh M. and Taamneh S. 2017. Severity prediction of traffic accident using an artificial neural network. Journal of Forecasting. 36(1): 100-108.

[3] Deublein M., Schubert M., Adey B. T., Köhler J. and Faber, M. H. 2013. Prediction of road accidents: A Bayesian hierarchical approach. Accident Analysis & Prevention. 51: 274-291.

[4] Jadaan K. S., Al-Fayyad M. and Gammoh H. F. 2014. Prediction of road traffic accidents in jordan using artificial neural network (ANN). Journal of Traffic and Logistics Engineering. 2(2): 92-94.

[5] Kumar S. and Toshniwal, D. 2016. A data mining approach to characterize road accident locations. Journal of Modern Transportation. 24(1): 62-72.

[6] Ren H., Song Y., Wang J., Hu Y. and Lei J. 2018. A Deep Learning Approach to the Citywide Traffic Accident Risk Prediction. 1-6.

[7] Ryder B., Gahr B., Egolf P., Dahlinger A. and Wortmann F. 2017. Preventing traffic accidents with in-vehicle decision support systems-The impact of accident hotspot warnings on driver behaviour. Decision support systems. 99: 64-74.

[8] Zheng J. and Wu X. 2015. Prediction of Road Traffic Accidents Using a Combined Model Based on IOWGA Operator. Periodica Polytechnica Transportation Engineering. 43(3): 146-153.

[9] Zheng L. and Meng X. 2011. An approach to predict road accident frequencies: application of fuzzy neural network. In 3rd International Conference on Road Safety and Simulation. 1-16.

[10] Zong F., Zhang H., Xu H., Zhu X. and Wang L. 2013. Predicting severity and duration of road traffic accident. Mathematical Problems in Engineering. 2013: 1-9.