BOUNDED MEMORY BASED FREQUENT PATTERN GROWTH APPROACH WITH DEEP NEURAL NETWORK AND DECISION TREE FOR ROAD ACCIDENT PREDICTION

Arun Prasath N

PhD Research Scholar, Department of Computer Applications,
Bharathiar University, Coimbatore 641046, India
arunphd25@gmail.com

Dr.M.Punithavalli

Professor, Department of Computer Applications, Bharathiar University,
Coimbatore 641046, India

Abstract - Frequent Pattern-growth (FP-growth) was introduced to categorize the accident locations. After the categorization of accident locations, accidents were classified by using J48. FP-growth has poor spatial and temporal locality issues. This is because of the construction process of FP-tree and its access behaviour of the mining algorithm. In this article, the issues of FP-growth are solved by introducing Bounded FP-growth. It uses only a bounded portion of the primary memory using specialized memory management. When the tree grows out of the allocated memory, it is forced to be partially saved on secondary memory. The secondary memory is accessed in a block-by-block basis so that both temporal and spatial localities of FP-growth are optimized. After the memory management process, FP-growth is applied to categorize the accident locations. The attributes which have high support values are processed in Deep Neural Network J48 (DNNJ48) for classification of accident type. The analysis of experimental results proved that the proposed Bounded FP-growth-Deep Neural Network with J48 classifier (BFP-growth-DNNJ48) achieved higher accuracy (94%) when compared with existing methods like AdaBoost-SO and TASP-CNN.

*Keywords***:** Road accident prediction, Apriori, Frequent Pattern-growth, Neural Network, decision tree, J48.

1. Introduction

The expeditious growth of urbanization has contributed to the rise of vehicles in modern society, which in turn the reason for increase in the number of accidents. In the Global road safety status report, 1.25 million deaths in road accidents were recorded. In turn, road accidents are the leading cause to many fatalities, causalities and economic losses. In order to address these problems, various road accident prediction methods [Labib *et al.* (2019)] has been developed for various purposes such as to estimate the expected accident frequency for various road entities and identify factors related to the occurrence of an accident. Most of the road accident prediction methods are based on statistical records or by doing some surveys based on interviews. But a satisfactory and unerring approach for road accident prediction cannot be found by this method.

Data mining is a scientific approach that can be applied to get a better result for road accident prediction. The data mining techniques can find the correlation in various attributes of road accidents to detect the severity of road accidents. Also, the identification of geographical location is very essential to understand the different circumstances of accident occurrence. A data mining approach [Kumar and Toshniwal (2016)] was proposed to characterize the locations of road accidents. Initially, road accident data were collected and then k-means clustering was applied on the data to cluster it as low frequency, moderate frequency and high frequency accident locations. The k-means clustering method clusters the data based on the threshold value.

Association rule mining technique named Apriori was used to uncover various factors related to road accidents. But, the Apriori consumes high space and time to characterize the road accident locations and also it is a costlier process especially when there is a large number of a pattern in the database. In order to overcome this problem, Frequent Pattern- growth (FP-growth) which is an association rule mining technique [ArunPrasath and Punithavalli (2019)]was introduced to find the confidence values of the attributes, the high confidence attributes only selected to construct the tree J48 classifier. Once tree is constructed from training data, J48 classifier classifies the accident type of test data as critical accident type and non-critical accident type. The FP-growth used in this approach has spatial and temporal locality issues. The spatial locality issues happened due to the nodes of any

prefix path of the tree may not be placed adjacent to each other in the memory while constructing the FP-tree. The temporal locality issues happened in the event that part of the FP-tree is kept in auxiliary memory when the mining of association rule from Fp-tree.

In this article, bounded FP-growth and Deep Neural Network with J48 (DNNJ48) are introduced to handle the spatio-temporal locality issues in FP-growth to enhance the classification accuracy of J48. The bounded FP-growth handles the spatio-temporal locality issues using specialized memory management in which the FP-tree is constructed in the available primary memory. If the data in the FP-tree exceeds the primary memory, a part of FP-tree will be forced to save in the secondary memory. The secondary memory is accessed in a block-by-block basis and the memory for the nodes in FP-tree is allocated by traversing it in a depth-first manner. Hence, the spatio-temporal locality issues in FP-growth are solved by bounded FP-growth. DNNJ48 is intrinsically interpretable as it is a tree. It is realized by Neural Networks (NN) where the categorized accident data by FP-growth is trained with gradient descent to classify the data as critical and non-critical.

2. Literature Survey

[Shanthi and Ramani (2011)] introduced data mining algorithms to classify the collision patterns in road accidents. Initially, the data cleaning process was applied in the collected data to remove the incomplete, noisy and inconsistent data. Then, Multi valued Oblivious Decision Tree (MODTree) filtering; feature ranking, Correlation-based Feature Selection (CFS), Mutual Information Feature Selector (MIFS) and Fast Correlation Based Filter (FCBF) feature selection algorithms were applied on pre-processed data to select the most appropriate features. Finally, the selected features were given as input to the Naïve Bayes, C4.5, Classification and Regression Trees (C&RT), RndTree, Decision List, rule induction and random tree to derive the classification rule for vehicle collision pattern detection. The decision tree-based classifiers required more space to store the data.

[Krishnaveni and Hemalatha (2011)] used some classification algorithms such as PART, J48, Random Forest (RF) and AdaBoostM1 and Naïve Bayes to predict the severity of injury occurred during traffic accidents. This contains labels of road classification, seriousness, casualty sex, types of collision, vehicle class of driver or passenger casualty, natural light, driver sex, degree of injury driver age and severity of accident for injury severity predictions. The injury severities categorized based on vehicle, casualty and accident. However, the prediction accuracy of the classification algorithms is low.

[Zheng and Meng (2011)] introduced a Fuzzy Neural Network Model (FNNM) for the prediction of road accident frequencies. Initially, k-means clustering was used to get the optimal number of fuzzy sets for input variables of training data. Then, the sub clustering and grid partition method were applied to generate the fuzzy inference system. It generates fuzzy rules using back propagation method. However, the consumption time to generate a fuzzy inference system is high

[Deublein et al. (2013)] presented a Bayesian hierarchical approach to predict the frequency of injury severities of road users and injury accidents on roads. Initially, the road accident risk models were formulated as risk indicating parameters to fragment the homogenous roads using Bayesian Probabilistic Networks (BPN). Then, hierarchical multivariate regression analyses and data mining techniques were combined to develop networks for the estimation of earlier inferences and adapt the BPN to the available data. The correlation between risk indicating parameters and the response variables were described using a multivariate Poisson- lognormal regression model which used to predict the frequency of injury accidents. The spatial correlation between the successive road-sections also influences the accident rates.

[An et al. (2014)] proposed a temporal data mining based system to identify road accidents. A cell transmission model was developed to represent traffic flow state in terms of ternary time series model. The linear drift due to changes in time series was dealt by implementing Discrete Fourier Transform (DST). After that, the similarity between the historical crash data and time-series data was calculated using Euclidean distance and based on the measures of similarity the road accidents were detected. The Euclidean distance performed well only when the dataset includes compact or isolated clusters.

[Mohamed(2014)] introduced a Multi-class Support Vector Machine (M-SVM) to predict the reasons for traffic road accidents. Data collected from the police department in Dubai was pre-processed by filling the missing data and then the class imbalance problem was solved. The most discriminative features were selected and it was given as input to M-SVM to predict the causes of traffic road accidents. The proper selection of kernel function in M-SVM is more difficult.

[Zheng and Wu (2015)] proposed a model based on the Induced Ordered Weighted Geometric Average (IOWGA) operator for prediction of the traffic accident. It combined the results of GM(1,1) as well as the Verhulst model by modifying the weight coefficients of each model through Optimized Weight (OW) method. The selection of the best suitable single methods with respect to the real application problem is not addressed in the combined model.

[Theofilatos *et al.* (2016)] proposed a rare-events modeling approach to predict road accidents. Initially, accident occurrence was explored in this approach by developing rare-events logistic regression models. Then, the method of stratified sampling was applied and three models were developed. These models determined the logarithm of average speed and lower speed. From the calculation of speed, the rare-event modeling approach predicted road accidents. The rare-events logit models with replications could be used to develop a totally consistent approach for road accident prediction.

[Ozbayoglu *et al.* (2016)] proposed a preliminary real-time autonomous detection system to detect the accidents based on the computer intelligence model. The collected data from the sensors were integrated with the event logs and then the most significant features were selected for the accident detection model. The selected features were processed in feed- forward neural network, regression tree and neighbour model to predict the occurrence of the accident. The false alarms of preliminary real-time autonomous detection system are considerably high.

[Jain *et al.* (2016)] proposed a data mining approach to analyze road accidents in India. K-means clustering algorithm was used in the proposed data mining approach to group the similar area. The results of K-means clustering were processed in a decision tree to analyze the road accidents. The random selection of k-value in K-means clustering greatly influences the performance of the data mining approach.

[Zheng *et al.* (2019)] proposed a deep learning approach based on Convolutional Neural Network (CNN) for prediction of traffic accident severity. Instead o directly using accident feature into CNN, this method found the feature relationship among the traffic accident features and the weight of feature was assigned based on relationship degree. The highly related features were fused and transformed as combined feature. However, the extraction of relationship among features for sparse dataset is unnecessary. So this method only provided better results for datasets which consist of more related features.

[Zhao *et al.* (2019)] proposed AdaBoost with Synthetic minority oversampling technique and One-hot encoding (AdaBoost-SO) method to predict the risk of the vehicle accidents. This method was started with filling the missing values and balancing the number of samples in each category. Then, road safety data characteristics such as driver's gender, vehicle type, vehicle age and driver's age were used in AdaBoost-SO to predict the risk of vehicle accidents. The convergence speed was slow if the numerical difference between multiple features is large.

3. Proposed Methodology

In this section, the locality issues of FP-growth are described. Followed by describing locality issues, the proposed bounded FP-growth and DNNJ48 is described in detail for road accident detection. A set of accident data items is represented as $A = \{A_1, A_2, \ldots A_m\}$ and a database of transactions $D = \{t_1, t_2, \ldots t_n\}$ where $t = \{A_{i1}, A_{i2}, \ldots A_{ik}\}$ and $A_{ij} \in A$, an association rule is an implication of the form S, $T \subset A$ is set of accident data items called item sets and $S \cap T = \emptyset$. A sample transaction database is given in Table 1.

Table 1. Transaction Database.

Transaction ID	Accident Data Items
1	A, C, D
2	C, D, E
3	A, B, E
4	A, B, C, E
5	B, E, F, G
6	A, C, D, E, G
7	B, C, D, E
8	A, B, C, E, H
9	A, C, D, E, H
10	A, B, C, E, H

The transaction database D is pre-processed by analyzing the whole database and finds the frequency of each data item in D to remove the infrequent items based on the minimum support threshold. Here, the minimum support threshold is set as 4. After the removal of infrequent items, the new transaction database is given in Table 2. Then, the data items are sorted items based on its frequency. The FP-tree is constructed by scanning the D. The constructed FP-tree for the new transaction database is shown in Fig. 1. Then, the association rules are generated by FP-growth. The FP-tree has poor spatial and temporal locality problems which is describe in the next section.

Table 2. New Transaction Database

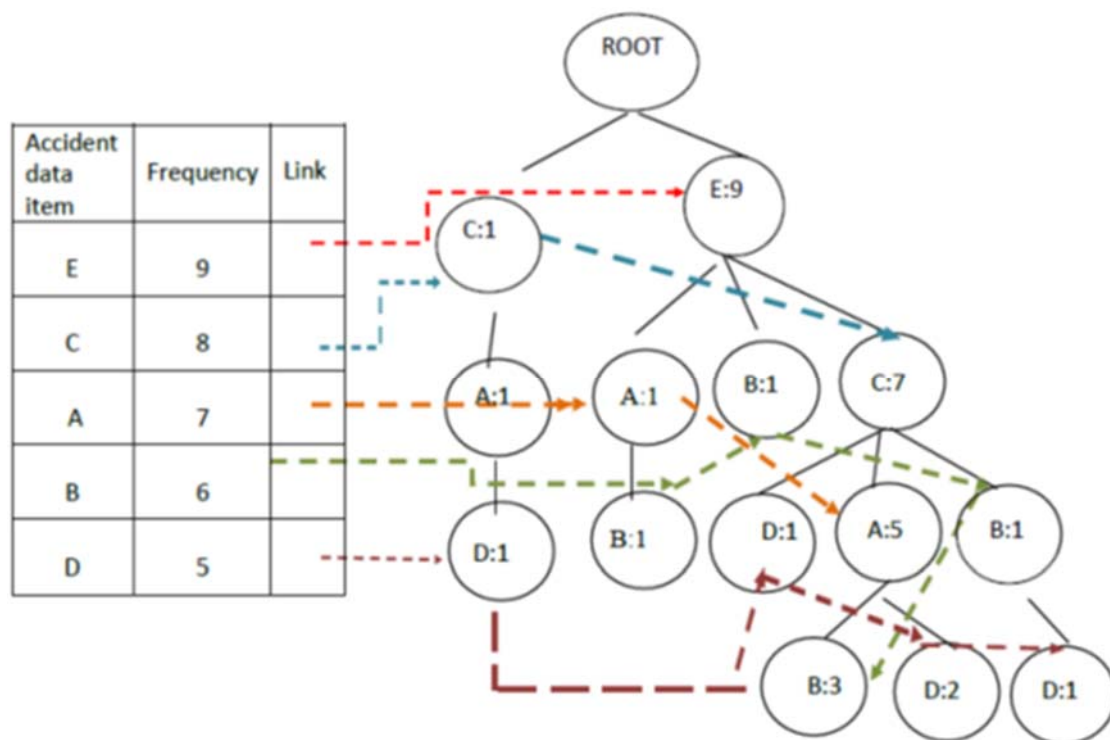| Transaction ID	Accident Data Items
1	C, A, D
2	E, C, D
3	E, A, B
4	E, C, A, B
5	E, B
6	E, C, A, D
7	E, C, B, D
8	E, C, A, B
9	E, C, A, D
10	E, C, A , B

Fig. 1. Construction of FP-tree

3.1. *Locality issues of FP-growth*

The building process of FP-tree and its access behaviour of the miming algorithm of FP- growth is the reason for its poor spatio and temporal localities.

Poor Spatial Locality

The child node is attached to a branch during the construction of FP-tree and the prefix path property of the FP-tree is preserved. In FP-tree, the prefix path is created and their nodes are positioned in the Numbering and spacing linear memory depends on the order by which the transaction is read and added to the file. The nodes of any prefix path of the tree may not be positioned adjacent to each other because of the construction process of FP-tree. As a result, there may be a chance of frequent cache missing when the FP-growth tries to extract prefix paths during the construction process of FP-tree or by traversing the FP-tree. The whole FP-tree could not get into the primary memory due to excessive data. Part of FP-tree is positioned in secondary memory by Virtual Memory Management (VMM) system. The VMM is like a memory manager to manage the memory requests made by the system and its applications.

The multiple disc swaps and primary memory may be revealed during the mining process owing to the frequent page faults resulting from the poor spatial locality of the tree nodes. Therefore, the categorization of accident data through the mining process takes exponentially longer time if the VMM cannot locate the whole tree in primary memory.

Poor Temporal Locality

When the mining algorithm tries to build conditional pattern bases and the conditional FP- tree, the FP-tree must be in the primary memory. This condition leads to a poor temporal locality problem when a part of FP-tree is positioned in secondary memory by VMM. Because the FP-growth may load the same accident data block several times back-and-forth from the secondary memory to primary memory. The FP-tree with node creation and node placement in memory is shown in Fig. 2.
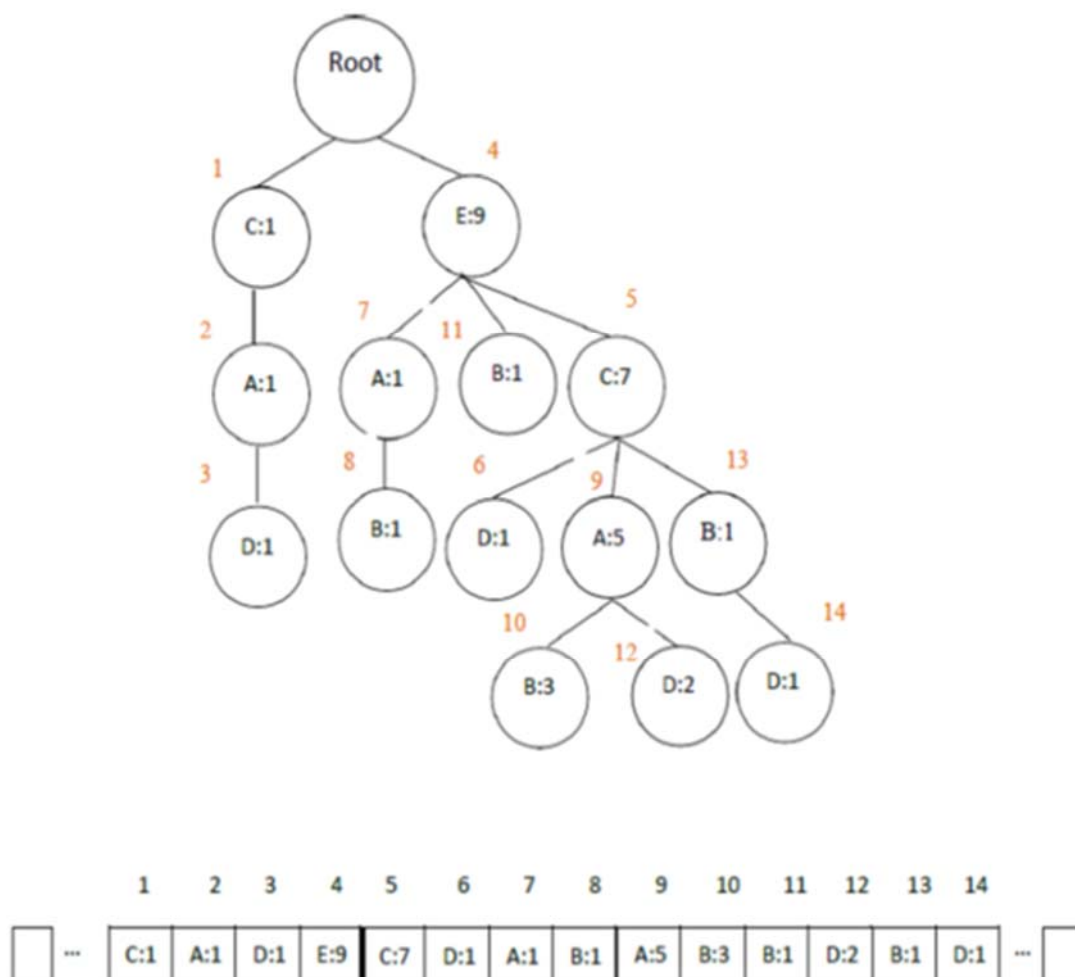

Fig. 2. FP-tree with node creation and node placement in memory

3.2. *Spatio-Temporal locality problem in FP-growth using BFP-growth to categorize the accident locations*

In BFP-growth, the spatial locality of FP-growth is improved by reorganizing the FP- tree nodes in a depth-first manner which is shown in Fig. 3. The FP-growth has a poor temporal locality problem because header table of FP-tree is read multiple times. In order to solve this problem, the BFP-growth collects the frequency for determining conditional pattern bases and then builds a conditional FP-tree from the collected pattern bases. The system would slow down massively due to the thrashing process. Thrashing makes little or no progress, generally, due to the exhaustiveness of memory the process become very constrained to perform accident locations categorization process. The VMM stores part of FP-tree in secondary memory which will be a cause to constant page faults. This problem is resolved by rearranging the computational part of FP-growth in a way of I/O conscious FP-tree through the process of page blocking. Therefore, if the system needs to load any block from the secondary storage devices to the main memory, it needs to load the block only once and this reduces the page faults dramatically.
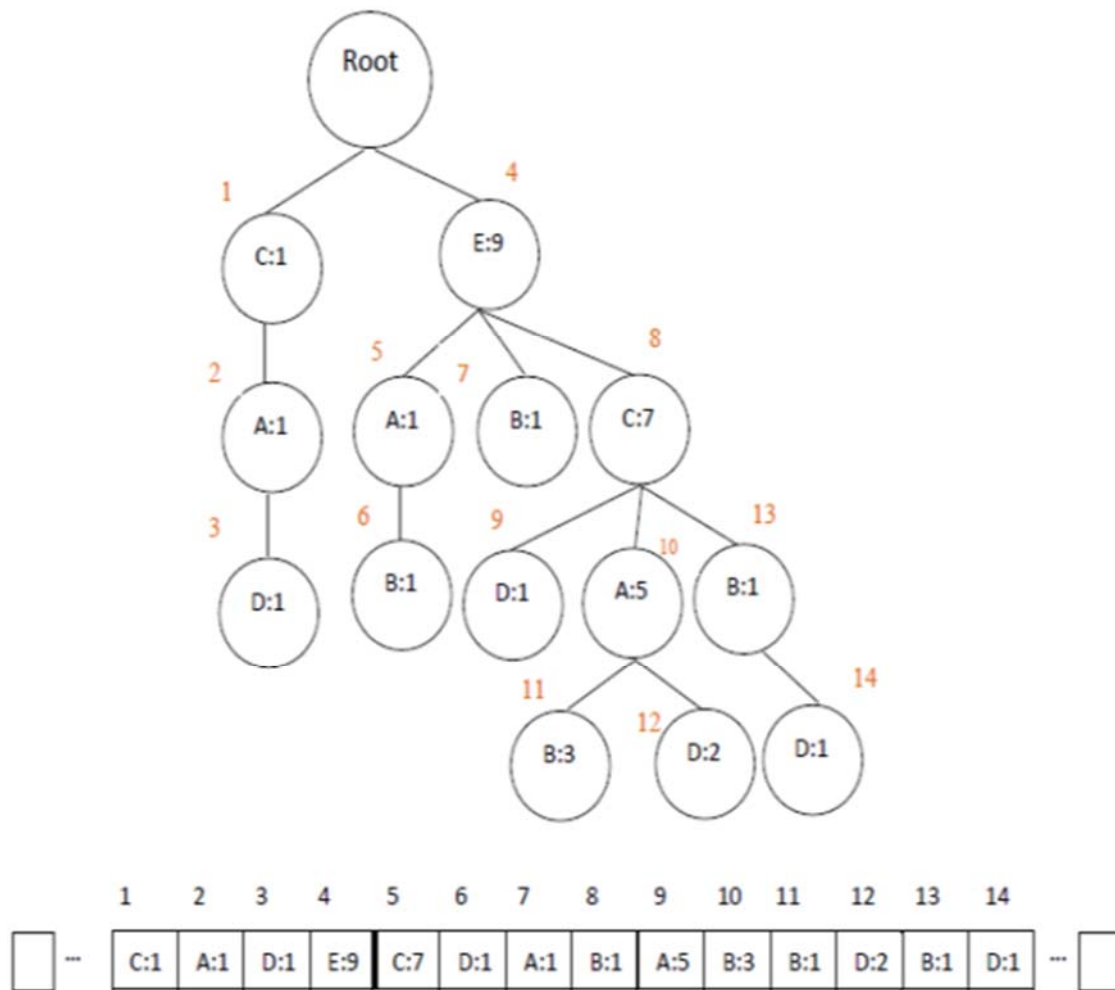
Fig. 3.   FP-tree in depth-first order

After the FP-tree construction and allocation of memory, the FP-growth is applied to generate association rules which are used to categorize the accident locations as high frequency, moderate frequency and low frequency accident locations. The attributes which have high support values are processed in DNNJ48 for classifying the accident type as critical and non- critical accident.

3.3. *Solving Spatio-Temporal locality problem in FP-growth using BFP-growth to categorize the accident locations Deep Neural Network with J48 for classification of accident type*

The DNNJ48 gets accident data attributes which have high support value as input to get accident type as output. DNNJ48 is a neural network that is trained with gradient descent. DNNJ48 finds a potentially better solution than J48 by simultaneously searching the structure and parameters of J48 with stochastic gradient descent. The DNNJ48 consists of the binning layer and Kronecker product.

The soft binning function is used in the binning layer which is used to make split decisions in J48. The soft binning function takes the accident data attributes x which has high support value as input and generates an index of the accident type to which x belongs. The accident attribute value x is continuous which is bin into $n + 1$ interval which leads to the need of n cut points, which are trained using neural network. The cut points of J48 are denoted as $[\beta1, \beta2, \dots \beta n]$ and one-layer neural network is constructed with softmax activation function which is given in Eq. (1)

$$fw, (x) \ = \ softmax \ (wx \ + \ b) \qquad (1)$$

In Eq. (1), w is a constant and its value ranges from 1 to n+1 and b is construct

$$b \ = \ [0, -\beta1, -\beta1 \ - \ \beta2, \dots, -\beta1 \ - \ \beta2 \ - \dots - \ \beta n] \qquad (2)$$

Given the binning function, the key idea is to build J48 decision tree through the Kronecker product. Each accident data attributes x binning by its neural network (x), the accident types can comprehensively find by,

$$z \ = \ f1(x1) \ \otimes \ f2(x2) \ \otimes \ \dots \ \otimes \ fD(xD) \qquad (3)$$

In Eq. (3), z indicates the index of the leaf node where the accident data attribute x arrives in the J48 decision tree. Finally, by traversing the J48 decision tree, the accident types are classified as critical and non-critical types.

4. Result and Discussion

The data about the accident is collected from the traffic police in the Coimbatore region. The data is collected from the period of 1st January 2013 to 30 November 2014 which consists of 2644 transactions. As part of this research each road traffic accident incident is logged with all potential attributes. This information is manually typed from the physically maintained record of Traffic Investigation Wing (TIW) of Police in their First Information Report (FIR) records which is maintained manually. The day of the incident, Incident Date ,Incident Time stamp ,Affected Vehicle Type , Causing Vehicle Type , Region of occurrence, Place of occurrence ,Place category, Number of Minors ages and Major ages affected, senior citizens , Vehicle Speed during the incident, Reason for accident , number of accidents by means of severity(fatal, grievous, minor injury ,only damage to the property).

Dataset is divided into two parts, one for training and other for testing .Initially the higher confidential attributes are selected from Bounded FP-growth. Then, the instances of selected attributes are given into DNN with class labels for assigning weight values to attributes .Once the weight values for attributes obtained from DNN; the J48 decision tree is constructed. The rules are generated from constructed decision tree. The generated rules are used to classify the test data. The performance of BFP-growth-DNNJ48 is measured in terms of accuracy, precision, recall, memory and time utilization. The proposed work is compared with the existing works AdaBoost-SO [15] and TASP-CNN [14].

4.1. *Accuracy*

Accuracy is defined as the proportion of the true outcomes for road accident prediction and non accidents prediction among the sum of the cases observed. It is calculated as,

$$Accuracy = \frac{True\ Positive\ (TP) + True\ Negative\ (TN)}{TP + TN + False\ Positive\ (FP) + False\ Negative\ (FN)} \quad (4)$$

In Eq. (4), if the class label is positive and the road accident prediction outcome is positive then it is TP. If the class label is negative and the road accident prediction outcome is negative then it is TN. If the class label is negative and the road accident prediction outcome is positive then it is FP.If the class label is positive and the road accident prediction outcome is negative then it is FN. Table 3 shows the accuracy of AdaBoost-SO, TASP-CNN and BFP-growth-DNNJ48 for road accident prediction.

Table 3. Comparison of Accuracy(%)

AdaBoost-SO	TASP-CNN	BFP-growth-DNNJ48
84.16	91.4	93.93

The comparison of AdaBoost-SO, TASP-CNN and BFP-growth-DNNJ48 methods in terms of accuracy is shown in Fig. 4. The accuracy of BFP-growth-DNNJ48 is 11.51% greater than AdaBoost-SO and 2.76% greater than TASP-CNN From this analysis; it is proved that the proposed BFP-growth-DNNJ48 method has high accuracy than other road accident prediction methods.
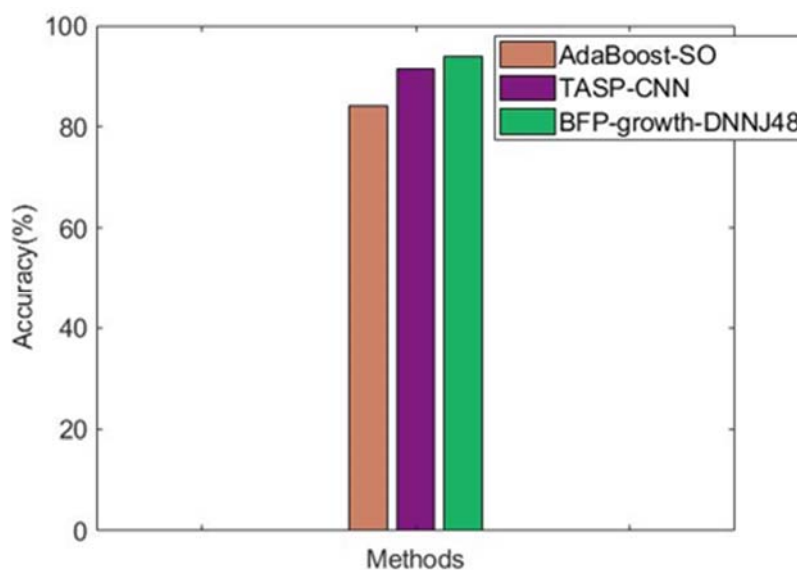

Fig. 4. FP-tree in depth-first order

4.2. *Precision*

Precision value is calculated as ratio between correctly predicted road accidents cases and correctly predicted all cases. Precision value is computed based on road accident prediction at true positive prediction, false positive. It is calculated as

$$Precision = \frac{TP}{TP+FP} \qquad (5)$$

In Eq. (5), if the class label is positive and the road accident prediction outcome is positive then it is TP.If the class label is negative and the road accident prediction outcome is positive then it is FP. Table 4 shows the precision of AdaBoost-SO, TASP-CNN and BFP-growth-DNNJ48 for road accident prediction.

Table 4. Comparison of Precision (%)

AdaBoost-SO	TASP-CNN	BFP-growth-DNNJ48
85.16	87.4	92.1

Fig. 5 shows the precision of AdaBoost-SO, TASP-CNN and BFP-growth-DNNJ48 for road accident prediction of a real-time accident dataset. The precision of BFP-growth-DNNJ48 is 8.23% greater than AdaBoost-SO and 5.74% greater than TASP-CNN. From this analysis, it is proved that the proposed BFP-growth-DNNJ48 method has high precision than other road accident prediction methods.
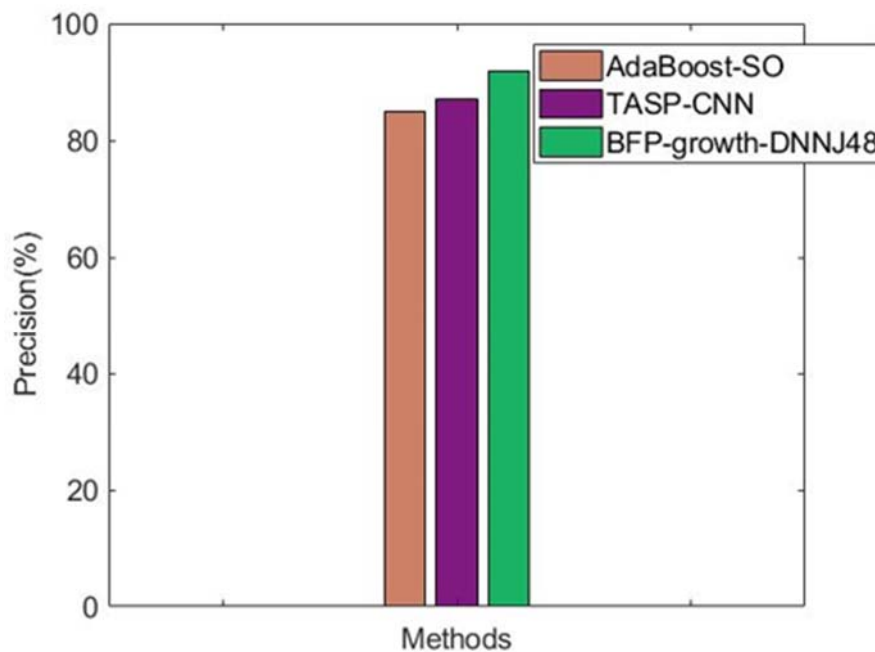

Fig. 5. Comparison of Precision

4.3. *Recall*

Recall is the ratio of the correctly predicted road accident cases from all available road accident cases in test dataset. Recall value is calculated based on road accident prediction at True Positive and False Negative prediction.

$$Recall = \frac{TP}{TP+FN} \qquad (6)$$

In Eq. (6), if the class label is positive and the road accident prediction outcome is positive then it is TP.If the class label is positive and the road accident prediction outcome is negative then it is FN.Table 5 shows the accuracy of AdaBoost-SO, TASP-CNN and BFP-growth-DNNJ48 for road accident prediction.

Table 5. Comparison of Recall (%)

AdaBoost-SO	TASP-CNN	BFP-growth-DNNJ48
87	90	95

Fig. 6 shows the comparison of recall between AdaBoost-SO, TASP-CNN and BFP-growth-DNNJ48 for road accident prediction. The recall of BFP-growth-DNNJ48 is 8% greater than AdaBoost-SO and 5% greater than TASP-CNN. From this analysis, it is proved that the proposed BFP-growth-DNNJ48 method has high recall than other road accident prediction methods.
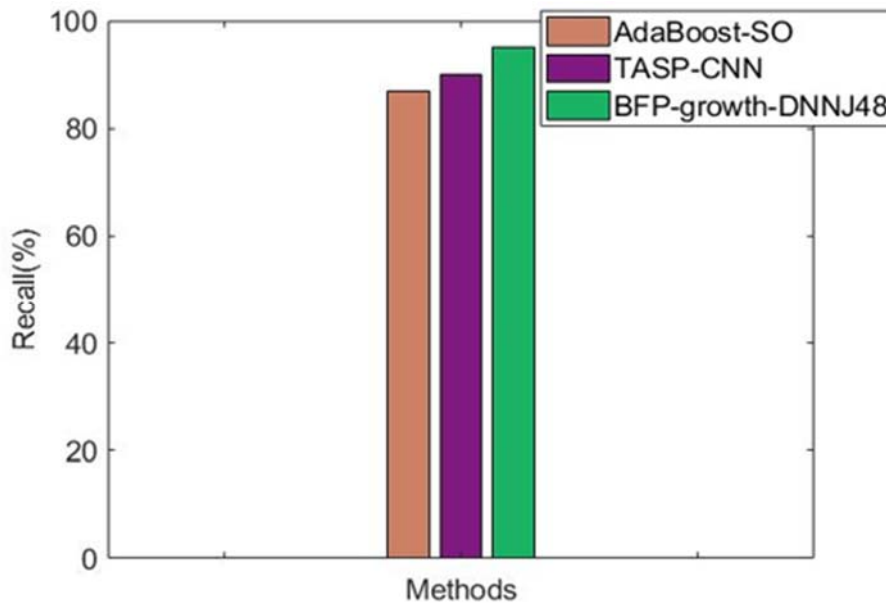

Fig. 6. Comparison of Recall

4.4. *Mining Time*

The mining time represents the time required to mine the accident data items using AdaBoost-SO,TASP-CNN and BFP-growth-DNNJ48.Table 6 shows the mining time of AdaBoost-SO, TASP-CNN and BFP-growth-DNNJ48 based road accident prediction.

Table 6. Mining Time (s)

Minimum Suppor(%)	AdaBoost-SO	TASP-CNN	BFP-growth-DNNJ48
2	200	190	150
4	175	166	127
6	152	141	103
8	128	117	79
10	103	91	54

The mining time of AdaBoost-SO, TASP-CNN and BFP-growth-DNNJ48 for different minimum support values is shown in Fig. 7. When the support is 10%, the mining time of BFP-growth-DNNJ48 method is 47.57% less than AdaBoost-SO and 40.66% less than TASP-CNN. From this analysis, it is proved that the proposed BFP-growth-DNNJ48 method has less mining time than other methods for road accident prediction.

4.5. *Memory usage*

The memory usage represents the memory required by AdaBoost-SO, TASP-CNN and BFP-growth-DNNJ48 for road accident prediction. Table 7 shows the mining time of AdaBoost-SO, TASP-CNN and BFP-growth-DNNJ48 for road accident prediction. The memory usage of AdaBoost-SO, TASP-CNN and BFP-growth-DNNJ48 methods for different support values is analyzed from Table 7. When the support is 10%, the memory usage of BFP-growth-DNNJ48 method is 37.93% less than AdaBoost-SO and 21.74% less than TASP-CNN. From this analysis, it is proved that the proposed BFP-growth-DNNJ48 method has less memory usage than other methods for road accident prediction.

Table 7. Comparison of Memory Usage(KB)

Minimum Suppor(%)	AdaBoost-SO	TASP-CNN	BFP-growth-DNNJ48
2	400	350	180
4	380	330	160
6	350	290	140
8	310	270	130
10	290	230	110

The memory usage of AdaBoost-SO, TASP-CNN and BFP-growth-DNNJ48 methods for different support values is analyzed from Table 7. When the support is 10%, the memory usage of BFP-growth-DNNJ48 method is 45% less than AdaBoost-SO and 30% less than TASP-CNN. From this analysis, it is proved that the proposed BFP-growth-DNNJ48 method has less memory usage than other methods for road accident prediction.

4.6. *Comparison with parts of proposed methods*

The performance of FP-growth-J48, FP-growth-DNNJ48 and BFP-growth-DNNJ48 based road accident predictions are analyzed. The analyzed results are shown in Table8.This results are analyzed for the support value 10%.From Table 8, it is concluded that the proposed work performs better than its predecessor methods in terms of all the performance metrics.

Table 8. Comparison of Bounded FP-growth-DNNJ48 with its antecedent methods

Methods	Minimum Suppor(%)	AdaBoost-SO	TASP-CNN	BFP-growth-DNNJ48
Accuracy(%)	2	78.57	85.69	90.34
Precision(%)	4	62.5	78.9	85.9
Recall(%)	6	42.5	60.8	72.9
Time(s)	8	103	91	54
Memory usage (KB)	10	240	210	180

The existing methods AdaBoost-SO, TASP-CNN directly learns knowledge from training dataset whereas BFP-growth-DNNJ48 learns from the training instances for highly confidential attributes only. So quality attributes used for training the classifier improve the performance of BFP-growth-DNNJ48 than AdaBoost-SO, TASP-CNN.

The memory management techniques used in BFP-growth reduces the considerable memory usage and mining time. In BFP-growth-DNNJ48, the weight value of the attribute for forming J48 decision tree is computed in the DNN classifier. So the DNN used in BFP-growth-DNNJ48 have very less number of iteration compared to AdaBoost-SO, TASP-CNN. It attributes for acquiring best results with less memory usage and mining time.

5. Conclusion

In this paper, BFP-growth-DNN-J48 is proposed for efficient road accident prediction. The experimental results proves that the proposed BFP-growth-DNNJ48 has 94% accuracy, 92% precision, 95% recalls for road accident prediction. The average mining time of proposed method is 40% less than AdaBoost-SO and TASP-CNN. Average Memory usage reduction of proposed method is 29%.

Initially, the accident data is collected and it is given as input to the BFP-growth where the association rules are generated by constructing BFP-tree. Specialized memory management is used in BFP-growth to optimize the spatial and temporal locality issues of FP-tree nodes. The attributes which has high confidence values are processed in DNNJ48 to classify the accident types with the help of NN and J48. Therefore, by resolving spatio-temporal locality issues in FP-growth and by using DNNJ48 the road accident prediction is improved which is shown in the experimental analysis. The Deep learning models has to be rebuilt from the very beginning once the feature-space distribution changes. In order to solve this issue, the transfer learning will be taken into account for the future work.

References

[1] An, S.;. Zhang, T ; Zhang, X; Wang,J.(2014).Unrecorded accidents detection on highways based on temporal data mining, Mathematical Problems in Engineering, 2014, pp.1-7.

[2] Arun Prasath, N; Punithavalli, M. (2019).Prediction of road accident locations in road accident database by mining spatio-temporal association rules, ARPN Journal of Engineering and Applied Sciences, 14(12) pp.2310-2314.

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

[4] Jain, A;Ahuja, G; Mehrotra,D.(2016).Data mining approach to analyse the road accidents in India, In IEEE 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO), pp.175-179.

[5] Krishnaveni, S; Hemalatha,M . (2011) .A perspective analysis of traffic accident using data mining techniques, International Journal of Computer Applications, 23(7), pp.40-48.

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

[7] Labib,M. F; Rifat, A. S;. Hossain, M. M; Das, A. K ; Nawrine,F.(2019): Road Accident Analysis and Prediction of Accident Severity by Using Machine Learning in Bangladesh, In: Proceedings of IEEE 7th International Conference on Smart Computing & Communications (ICSCC), 2019,pp.1-5.

[8] Mohamed, E. A. (2014).Predicting causes of traffic road accidents using multi-class support vector machines, Journal of Communication and Computer, 11(5) pp.441-447.

[9] Ozbayoglu, M ; Kucukayan, G; Dogdu,E.(2016).A real-time autonomous highway accident detection model based on big data processing and computational intelligence, In: IEEE International Conference on Big Data (Big Data), pp.1807-1813.

[10] Shanthi, S; Ramani, R. G.(2011)Classification of vehicle collision patterns in road accidents using data mining algorithms", international Journal of Computer Applications, 35(12), pp.30-37.

[11] Theofilatos, A;Yannis,G;   Kopelias, P; Papadimitriou, F .(2016).Predicting road accidents: a rare-events modeling approach",Transportation research procedia, 14, pp.3399-3405.

[12] Zhao, H;Yu, H;Li, D;.Mao,T; Zhu, H. (2019).Vehicle accident risk prediction based on AdaBoost-so in vanets",IEEE Access, 7, pp.14549-14557.

[13] Zheng, J;Wu,X.Prediction of Road Traffic Accidents Using a Combined Model Based on IOWGA Operator,Periodica Polytechnica Transportation Engineering, 43(3), pp.146-153.

[14] Zheng, L; Meng, X; (2011). An approach to predict road accident frequencies: application of fuzzy neural network, In: 3rd International Conference on Road Safety and Simulation.

[15] Zheng, M; Li, T;.Zhu, R; Chen, J; Ma, Z.;.Tang, M ;.Wang, Z.(2019).Traffic accident's severity prediction: A deep-learning approach-based CNN network,IEEE Access, 7, pp.39897-39910.