Asian Journal of Information Technology 15 (17): 3247-3255, 2016
ISSN: 1682-3915

Intelligent Temporal Model Using Neuro Fuzzy Decision Tree Classification Algorithm for Online Social Network Analysis

[1]Indira Priya Ponnuvel, [2]Uma Rani Saravanan and [3]Kannan Arputharaj
[1]Department of Computer Science and Engineering,
Tagore Engineering College Chennai, Tamil Nadu, India
[2]Chennai, Tamil Nadu, India
[3]Department of Information Technology, College of Engineering Guindy,
Anna University Chennai, Tamil Nadu, India

Abstract: In social data mining, classification is considered as the most effective decision making techniques among all the human activities. However, the existing algorithms for classification were based on decision making by a central decision manager. Therefore, the aim of this research, a new intelligent temporal model with an inference engine and a new algorithm called Temporal Neuro Fuzzy Decision Tree Classifier (TNFDTC) has been proposed for social network analysis. In this proposed model, inductive methods are proposed and used to classify values of attributes of unknown objects based on temporal features by providing appropriate classification using decision tree rules. Rules provided by TNFDTC are also useful for understanding the combinations of contents driving popularity over certain social networks. This proposed temporal neuro fuzzy decision tree has a fuzzy decision tree and a fuzzy decision model to handle uncertainty. It also uses temporal constraints to improve the classification accuracy by enhancing the existing neuro fuzzy decision tree classification algorithm. The parameters of the existing fuzzy decision trees have been adapted in this research which are based on stochastic gradient de-scent algorithm and hence it traverses back from leaf to root nodes. This research is useful to provide connection between consolidated features of users based on network data and also using the traditional metrics used in the analysis of social network users. From the experiments conducted in this research, it is observed that the proposed research provides better classification accuracy due to the application of neuro fuzzy classification method in decision model analysis.

Key words: Fuzzy decision tree, neuro-fuzzy decision tree, social network, classification, India

INTRODUCTION

In social data mining, classification is considered as the most effective decision making technique among all the human activities. In classification problem a person or an entity is put into a class based on the predefined properties of the person or entity. Traditional methods available in statistics for classification including discriminate analysis were used in the past for decision making under uncertainty using Bayes theorem Indira (Priya *et al.*, 2015). In this system, a new probability based model has been proposed with the addition of temporal constraints to compute the posterior probability which is used for effective classification. The effectiveness of this proposed model is con-strained by the assumptions made in this work and also based on the related work. Therefore, it is necessary to have a thorough understanding of the proposed model, application constraints and the type of data used. In the past, many classification techniques have been proposed by various researchers based on the application of neural networks for effective classification. Most of these techniques are based on neural networks. This is due to the fact that neural networks provide a number of advantages in the e effective classification of data. First, neural networks based classification techniques are flexible and based on weight assignment. Second, they use function approximations with respect to the activation functions. Third, they are useful to form rules for learning and decision making. Finally, training and testing for neural networks are easy to implement.

Temporal data mining deals with the application of conventional data mining techniques for effective mining of temporal data. Temporal data mining techniques are helpful for obtaining the temporal relationships from data sets (Kottas *et al.*, 2006). Temporal mining can be performed either by applying the classification or clustering techniques for analyzing various time oriented

Corresponding Author: Uma Rani Saravanan, Chennai, Tamil Nadu, India

data including social network data. Therefore, the ultimate goal of temporal data mining is to discover hidden temporal relations between sequences and subsequences of events (Agrawal and Srikant, 1995). The discovery of temporal relations between sequences of events involves mainly three steps: Building a temporal data representation technique effectively to model the temporal data sequence in a suitable form. Classification of these temporal data based on similarity measures between data sequences and time intervals. Use the past and present data for the prediction of future using the classification results.

Neural networks, genetic algorithms and fuzzy logic are effective soft computing techniques which can solve complex problems efficiently. Each method handles the uncertainty and ambiguity of data differently. In many applications, these technologies are combined to utilize the features of each to achieve impressive results. For example, a combination of artificial neural networks and fuzzy logic can result in a model that improves speed, fault tolerance and adaptiveness. In the past, the fusion of neural networks and fuzzy inference systems has attracted the attention of many researchers in various scientific and engineering areas due to the growing need for intelligent decision support systems which are used to solve the real world problems. However, a hybrid model which integrates decision trees, temporal features and fuzzy logic is useful for effective decision making. Madadipouya (2015) to form effective classification accuracy in the generated decision tree is improved, the depth of the tree is reduced and the weaknesses such as more time for computation and it is not still able to escape from being trapped into local optimum.

In this research, a decision tree based temporal model is proposed for effective social network analysis. A decision tree is a classifier in which each branch node is useful to make a choice between a numbers of alternatives where each leaf node represents a suitable decision on data. Moreover, decision trees are commonly used for gaining information for the purpose of effective decision making. However, they do not provide importance to time attributes and hence temporal extensions are proposed on decision trees in this study. This proposed temporal decision tree starts with a time stamped root node from which it grows up to leaf nodes where are nodes are time stamped using time intervals. From the root node, the proposed algorithm splits each node recursively using time stamps and by applying temporal constraints on the steps of the decision tree learning algorithm. The final result is a temporal decision tree in which each branch represents a possible decision which is valid for an interval. Since, temporal decision trees used in classification problems are concerned with temporal data, they are often called as temporal classification trees. In a temporal decision tree, each leaf node contains labels that indicate a predicted class of a given feature vector. Such classification can be more effective if it is enhanced with rules which are formed by training using neural networks.

In this study, a new intelligent temporal model is proposed with a modified inference mechanism based on the matching and ordering of rules with temporal aspects that enhances the ring strengths of fuzzy decision tree paths in response to user queries. This system have been trained and tested with large volume of social network data and hence it is suitable for the analysis of big data analysis on social networks. Moreover, this work enhances the back propagation algorithm with temporal constraints. The temporal decision tree proposed in this study directly uses the rules from the neural networks on the structure of fuzzy decision trees. Therefore, it improves its learning accuracy without compromising the comprehensibility. The proposed methodology has been validated using experiments conducted on real world datasets. Finally, a new temporal mining model is proposed in this study for social network analysis. The main aim of this study is to propose a new temporal neuro fuzzy decision tree algorithm provides best to make it harder for such events to occur by applying the intelligent classification methods for the temporal data in social network analysis. The agent based decision making process increases the decision time.

Literature review: Nozaki *et al*. (1996) proposed an adaptive method for fuzzy rule extraction for classification. Their method consists of two learning procedures: an error correction-based learning scheme and an additional learning procedure. The first learning procedure adjusts the grade of certainty of each rule based on its classification performance. Classification is often posed as a supervised learning problem (Qi and Davison, 2009) in which a set of labeled data are used to train a classifier which can be applied to label future examples. Many types of classification methods are present in the literature for the classification of social network data. It includes subject classification, functional classification, sentiment classification and others. Subject classification is concerned about the subject or topic of a web page. Other types of classification include genre classification (Stein and Eissen, 2008) and search engine spam classification.

Debnath *et al*. (2005) proposed a new algorithm to partition a Hypher Text Markup Language (HTML) page into constituent web page blocks. They proposed four new algorithms, namely Content Extractor, Feature Extractor, K-Feature Ex-tractor and L-Extractor. These algorithms identify the primary content blocks by looking for blocks that do not occur a large number of times across web pages by looking for blocks with desired features and by using classifiers, trained with block-features, respectively. Web page classification algorithm based on Sup-port Vector Machine (SVM) was discussed by Xue *et al*. (2006). Pach *et al*. (2008) proposed a novel classification model that is based on easily interpretable fuzzy association rules which fills the efficiency criteria also.

Luengo and Herrera (2010) discussed about fuzzy association rules which are generated based on fuzzy support and fuzzy confidence. According to these authors, fuzzy support indicates the compatibility grade between all the data and the rule while fuzzy confidence indicates the accuracy of the fuzzy rule. The goal of association rule based fuzzy classification is to construct a fuzzy classifier with strong classification ability. Classification accuracy and interpretability are two major criteria used in the literature to evaluate fuzzy classifiers. Accuracy is the ability to correctly classify unseen data. On the other hand, interpretability is the level of understanding and insight that is provided by the model. Fernandez *et al*. (2010) discussed about multi-class classification for linguistic fuzzy rule based classification systems. They decomposed the original data-set into binary classification problems using the pair wise learning approach and obtained an independent fuzzy system for each one of them.

Fuzzy decision trees are powerful, top-down, hierarchical search methodology to extract human interpretable classification rules. However, they are often criticized to result in poor learning accuracy. Bhatt and Gopal (2006) proposed a Neuro-Fuzzy Decision Trees (N-FDTs) which is a fuzzy decision tree structure with neural like parameter adaptation strategy. In the forward cycle, the researchers constructed fuzzy decision trees using the standard induction algorithm namely fuzzy ID3. By applying back propagation algorithm directly on the structure of fuzzy decision trees, they found that this method improves its learning accuracy without compromising the comprehensibility. Their methodology was validated by them using suitable experiments on real-world datasets.

Temporal classification techniques were paid much attention in the past two decades. Moreover, representation and analysis of temporal data pertaining to web mining is gaining importance recently. This helps to perform temporal pattern analysis, where temporal databases are used for storing web log data. These databases provide features for temporal rules extraction Meamarzadeh *et al*. (2009) which may be further used for decision support system.

Rodrigo Coelho Barros developed a new decision-tree algorithm that provides an up-to-date overview that fully focused on evolutionary algorithms. This is due to the fact that decision trees do not concentrate on any specific evolutionary approach. They also provided a taxonomy which addresses only the works that evolve decision trees. A number of references are provided by the researchers that describe applications of evolutionary algorithms for decision-tree induction in different domains. According to them, the important advantage of the evolution of decision trees is the possibility of explicitly biasing the search space through multi objective optimization. The main limitation of evolutionary induction of decision trees is related to the handling of time constraints.

Bhatt and Gopal (2006) proposed a framework of multi-genre movie recommender system based on Neuro-Fuzzy Decision Tree (NFDT) methodology. The system is capable of recommending a list of movies in descending order of preference in response to user queries and pro les. This system also takes care of attempt to vote stung using novel application of fuzzy c-means clustering algorithm. The rules represented by NFDT also act as a tool for understanding the combinations of contents driving popularity (and unpopularity) over certain social network. Moreover, they proposed a modified inference mechanism based on matching and ordering of ring strength of each fuzzy decision tree path in response to user queries. Their proposed approach is also useful in linguistic genre analysis to understand what drives popularity of certain movies with respect to combinations of genres.

Many researches are present in the literature on classification techniques based on decision trees for social network analysis. However, the temporal nature of social network data is not handled efficiently due to the absence of temporal constraints. In this study, we propose a new fuzzy decision tree which temporal constraints to handle the analysis of past and present data and to predict the future used behaviour and needs.

MATERIALS AND METHODS

Figure 1 shows the architecture of the proposed temporal rule based classification model. It consists of seven major components namely, social network data, data collection agent, classification module, rule

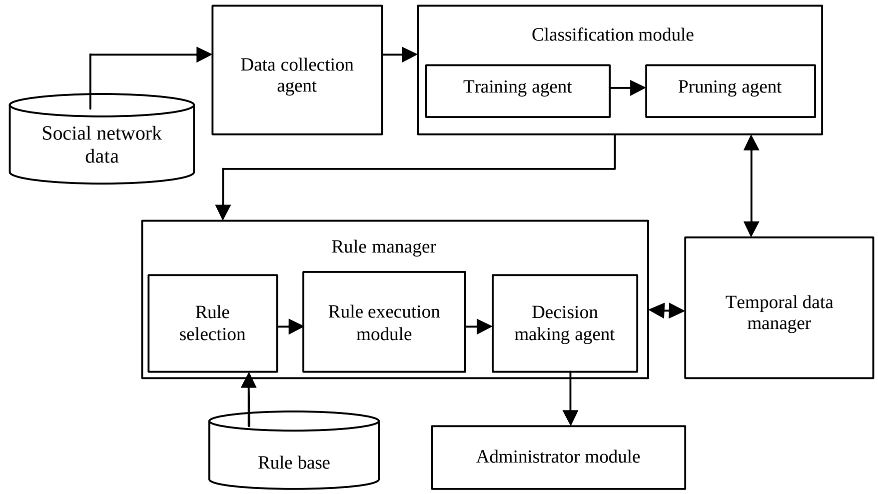

Fig. 1: Architecture of the temporal rule based classification model

manager, rule base, temporal data manager and administrator. The input to the Temporal Neuro Fuzzy Decision Tree Classifier (TNFDTC) system is referred from the social network trace data which is used in this research for carrying out the experiments. The data collection agent collects the necessary data from the network trace data set. These data are sent to the classification module for classification of the data where it classifies the data by using the training and pruning agents. The training agent trains the data which are received from the data collection agent. The pruning agent prunes the data which are received from the training agent. The rule manager is responsible for rule extraction, rule selection and decision making. First, rules are extracted based on the information provided by the classification module. They are stored in the rule base. During testing, the rule selection subsystem selects the suitable rules and executes them to get the results which are sent to the decision making agent. The decision making agent classifies the test data using these rules and makes suitable decisions on the social network data.

The rule base contains fuzzy temporal rules for classifying the data in IF-THEN format. The data set features are classified based on the features of the data set. Moreover, fuzzy IF-THEN rules are formed based on the classification module and are stored in the rule base. The temporal data manager collects the classified data cooperates in decision making by executing the temporal constraint satisfaction rules. Moreover, it performs validation using tenfold cross validation whenever the user initiates it.

Proposed work: In this research, a new algorithm called TNFDTC has been proposed. It consists of three major steps as given:

- Step 1: Call the rule creation procedure
- Step 2: Tune the rule parameters
- Step 3: Optimize rules by pruning

Each rule is fired by using a rule matching procedure and a deductive inference method.

Construction of rules: The steps required for rule extraction are as follows:

- Step 1: Whenever a data point is classified at a particular node, it marks that node as a classifier node
- Step 2: In every node it maintains the mean (μ) and standard deviation (σ) of the data points that pass through it

The rules are then constructed as follows:

- Each path starting from the root up to a classifier node is converted as a rule
- Since, the number of clauses in the antecedent part of the rule is the same as the number of levels encountered in the corresponding path, a validation is performed to check this
- Each node encountered in the path (except the root node) is converted to a clause and is associated with an initial membership function of Gaussian-type with position = μ and width = σ, i.e., membership function is given by $m(x) = c^{-(x-\mu)^2/2\sigma^2}$ where is the corresponding feature value of the data point
- The consequent of the rule, i.e., the class assigned by the rule is the same as the class label assigned by the corresponding path

The antecedent parts of different rules generated by the above scheme have different number of clauses. A path involving two levels is converted to a rule having two clauses in the antecedent part whereas a path involving five levels are converted to a rule having five antecedent clauses. To calculate the final output of a rule, the product is used in this research as the intersection operator. For example, the ring strength of the rule on the data point is calculated based on the product where the membership value of the data point in the fuzzy set associated with the clause of the rule and the total number of clauses in the rule. The initial values of the Gaussian membership function associated with a node are taken as the mean and standard deviation of the feature values of all training data points passing through that node. When a data point is obtained with unknown class, the ring strength of all rules is calculated using rules present the rule-base by selecting the rule with the maximum ring strength. This process is repeated using rule chaining to perform deductive inference.

Tuning of the extracted rules: Classification by this proposed classifier is threshold-based. Therefore, a node as well as its child belonging to the same class may both classify some data points and thus both may get marked as classifier nodes. As a result, two rules may be generated but the will have two different time stamps. Hence, the antecedent clauses of these two rules will be almost similar except the time-stamp values. In fact, the antecedent part of the rule corresponding to the parent node is contained in the antecedent of the rule corresponding to the child and both of them classify for the same class. The chance that one of these two rules will become redundant is very high. The μ and σ of the fuzzy sets associated with the rules are very important in this research and they increase the performance of the rule-base. In this case, these parameters are initialized by the mean and standard deviation of the data points passing through that node. This system does not analyze initially about how many of them are properly classified and how many are misclassified. Moreover, a lot of data points belonging to other classes may also pass through this node. Hence, with the μ and σ initialized by the mean and standard deviation of those data points may (usually will) not be the best possible choice. Thus, they should be tuned for a better performance. This system has used gradient descent technique to tune the parameters of the rules.

Pruning the rule-base: The proposed rule extraction method, however, does not rule out the possibility of presence of redundant rules as well as conflicting rules, which should be removed from the rule-base for providing better performance as well as efficiency. Therefore, a rule optimization procedure is proposed in this research which is very simple but effective scheme for rule minimization. After tuning is over the training sample is classified once more with the rule-base by performing the following.

- With each rule, namely two variables, Correct_Count and Incorrect_Count are maintained
- Whenever any data point is classified correctly by a rule, the Correct_Count is incremented by 1
- Whenever any data point is classified incorrectly by a rule, the incorrect_Count is incremented by 1
- After all training samples are exhausted; the rules that satisfy any one of the following two criteria are delete
- A rule whose Correct_Count is less than its Incorrect_Count
- A rule whose Correct_Count is less than a certain percentage of the cardinality of the training set. The optimized rules are stored in the rule base

Temporal neuro fuzzy decision tree classifier algorithm: The temporal neuro fuzzy decision tree classifier proposed in this study is capable of handling uncertainty and temporal aspects. The steps of the proposed algorithm are as follows.

Step 1: Generate the root node that has a set of all data, i.e., a fuzzy set of all data with the membership value 1. Use its time stamp for ordering.

Step 2: If a node t with a fuzzy set of data D satisfies the following conditions:

a. Check whether the proportion of a data set of a class C_k is greater than or equal to a threshold _ : during the interval[t1,t2].

b. If the number of a data sets is less than a threshold _:, there are no attributes for more classification's, then it is a leaf node and assigned by the class name.

Step 3: If it does not satisfy the above condition, it is not a leaf node and the test node is generated as follows:

a. For A_i, S(i = 1, 2, n) calculate the information gains G and select the test attribute A_m that maximizes them in the presence of temporal constraints

b. Divide D into fuzzy subset D_1, D_2, D_m according to A_{max} where the member-ship value of the data in D_j is the product of the membership value in D and the value of A max in D order them using time stamps

c. Generate new nodes n_1,n_2,n_m for fuzzy subsets D_1, D_2, .., .D_m and label the fuzzy sets.

d. Replace D by D_j (j = 1, 2, ..m) and repeat steps 2 and 3 recursively

Repeat this procedure until all nodes are generated and data are exhausted.

Proposed inference mechanism using rule firing:

- Sort the elements of the user query in ascending order of time stamps
- Optimize the query using rules send the query to inference engine
- Find fuzzy membership values and apply temporal constraint
- Select the rules and test their antecedent using scheduler and interpreter
- Perform forward chaining inference using retrieved data and rules
- Output the result to the user

The rules extracted from the training phase are stored in the rule base. During the testing phase, rules are selected and red using forward chaining inference for making effective decisions. The rule manager developed into this work uses fuzzy temporal constraints in rule matching and hence the decisions on medical datasets are provided with necessary accuracy. Rule execution is carried out during the testing phase for checking the input data effectively. Moreover, decision is performed through classification heads in which the decision based classification technique is effective in reducing the classification data on user details and the important data are classified.

RESULTS AND DISCUSSION

The performance of the rule-base extracted from T-NFDTC is shown in Table 1. From this Table 1, it can be observed that tuning of the rule-base improves the performance of the rule base a lot, which implies that the initialization of the parameters of the rules was not the optimum. For all data sets, there is a significant improvement in the recognition score achieved by the rule-base for social network dataset the performance remains the same.

Table 1 shows the performance analysis for the different datasets. From the Table 1, it can be observed that the classification accuracy for face book data is better than considered other datasets. Table 2 provides the comparative analysis of recall and precision values of the proposed algorithm with the existing algorithms. Moreover, the decision for the temporal data is performed only through the classification heads.

From Table 2, it can be seen that the proposed classification algorithm shows better performance in terms of precision, recall and F-measures values when it is compared with the existing systems Indira (Priya *et al*., 2014). From Table 2, it can be observed that precision, recall and F-measure are improved when the data are classified using Neuro Fuzzy Decision tree. This is due to the fact that the intelligent decision tree uses intelligent agents for decision making. However, the agent based decision making process increases the decision time. However, the proposed approach consumes less time for training and testing when it is compared with the NFDTC based approach. From the Fig. 2, it can be observed that the classification accuracy for Facebook data is better than considered other data sets. The overall precision and recall for classifying the data related to one application is reduced by >10% when the proposed neuro fuzzy decision tree efficient classification method is used.

Table 1: Performance of the rule base on different data sets

	Correct classification (%)			Number of rules	
Data set used	Not tuned	Tuned	Pruned	Not pruned	Pruned
Facebook	96.0	98.0	98.0	4	4
Youtube	85.7	91.1	91.1	11	7
Weblog	80.6	95.8	97.6	8	5

Table 2: Performance analysis of classification accuracy

Data set	Approach	Precision (%)	Recall (%)	F-measure (%)	Train (sec)	Test (sec)
Facebook	T-NFDT	88.67	97.97	93.34	6.88	2.02
	NFDT	88.19	97.82	92.73	6.91	2.04
Youtube	T-NFDT	99.99	97.23	98.72	25.22	14.83
	NFDT	99.98	97.05	98.50	26.59	15.17
Weblog	T-NFDT	94.93	28.32	42.97	5.78	3.32
	NFDT	94.70	27.08	42.08	5.30	2.96
Blog	T-NFDT	55.23	63.56	59.03	0.99	0.86
Catalog	NFDT	55.07	62.35	58.19	0.85	0.67

Figure 2-6 show the accuracy of number of visit analysis for Facebook, Youtube, Weblog, Blog catalog Dataset. From these figures, it can be observed that the proposed TNFDTC provides higher accuracy all these types of Dataset. This is due to the fact that intelligent agents and rules applied in the proposed model enhanced the detection accuracy. The results obtained by applying the FDT is used to get the precision graphs.

Figure 7 proposed the TNFDTC algorithm is compared with existing NFDT algorithm. From this figure, it has been observed that the performance of the proposed algorithm is improved with respect to precision. This improvement helps to retrieve customized web users. Moreover, the use of intelligent agents helps to identify optimal decision in which the agent based decision making process increases the decision time hence the overall classification method performance is increased.

Figure 8 shows the performance analysis of the online purchase for different set of age groups. From Fig. 8, it can be observed that the classification accuracy is more for No. of visits when it is compared with purchased due to the behavior of the dataset.

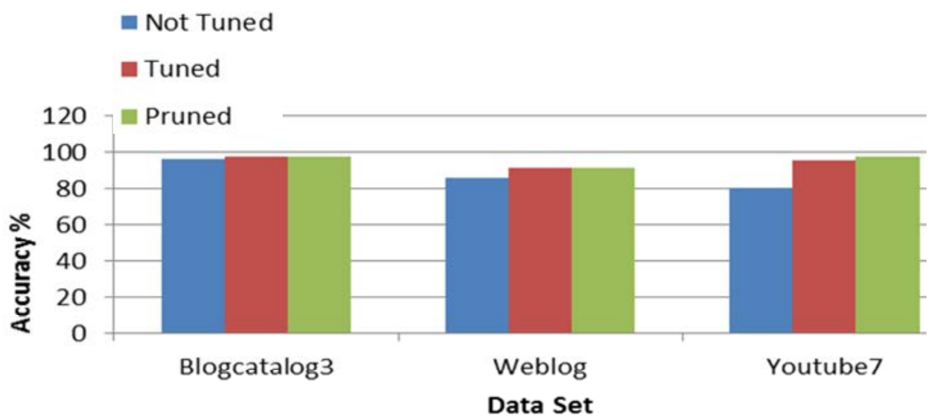

Fig. 2: Performance analysis for rules

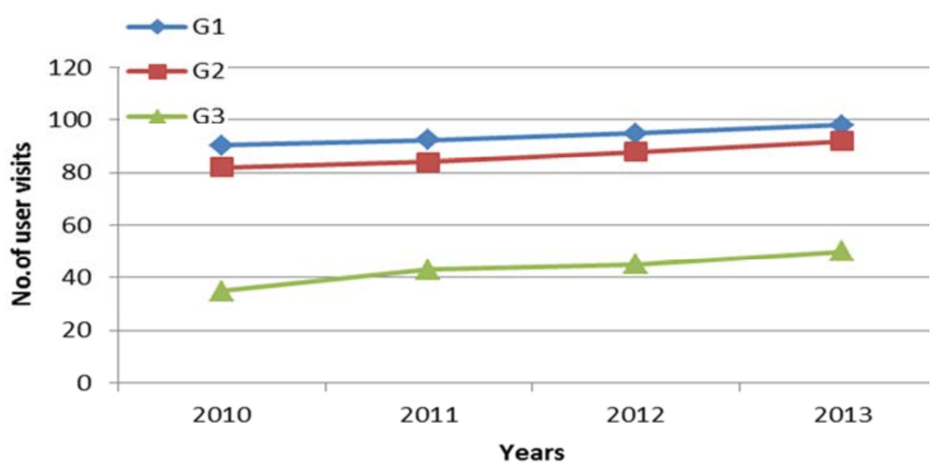

Fig. 3: Facebook analysis

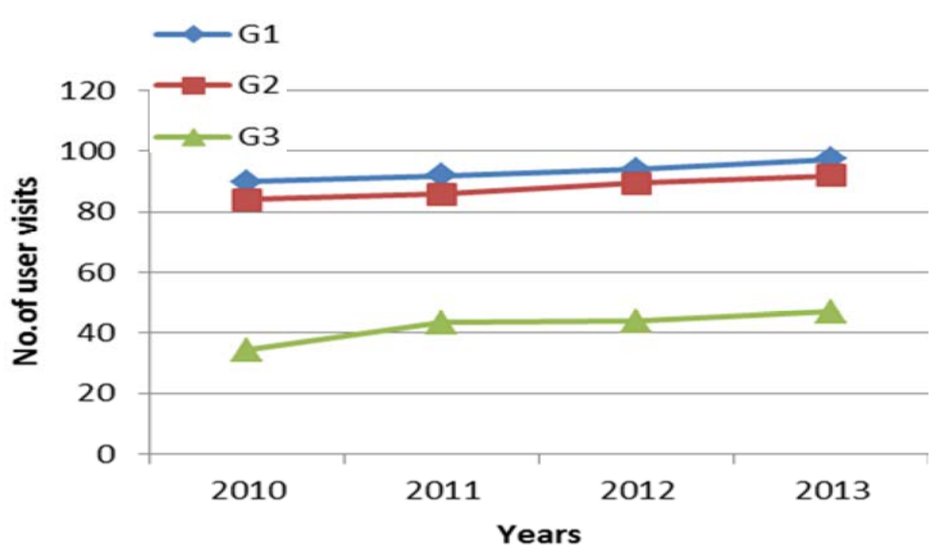

Fig. 4: Youtube analysis

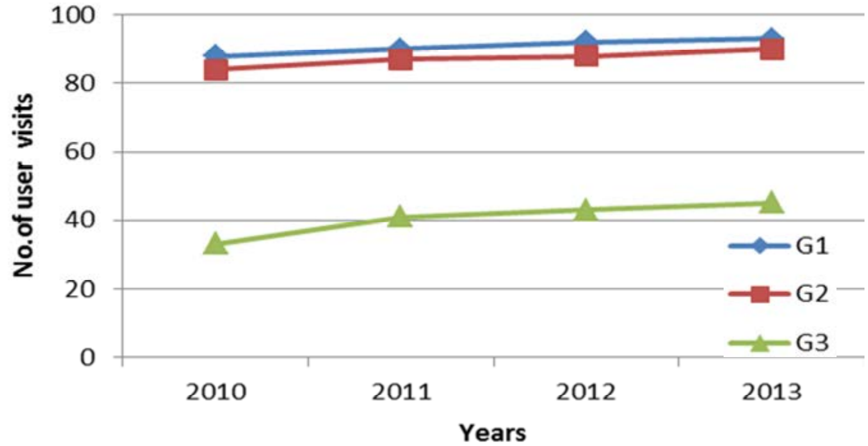

Fig. 5: Weblog analysis

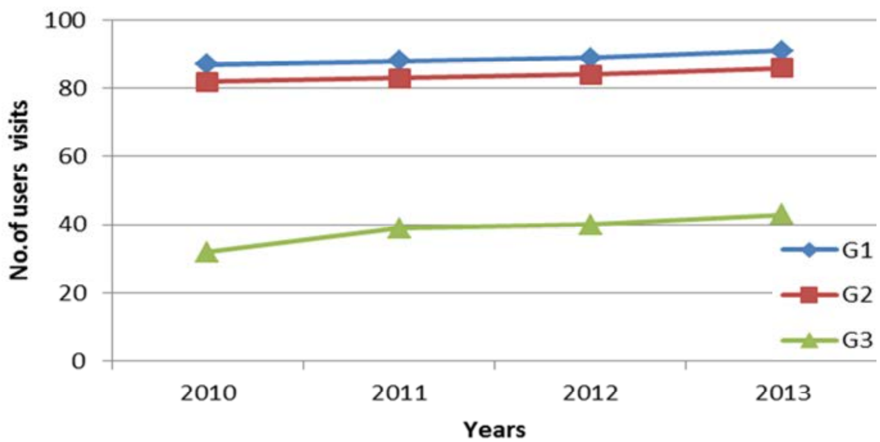

Fig. 6: Blog catalog analysis

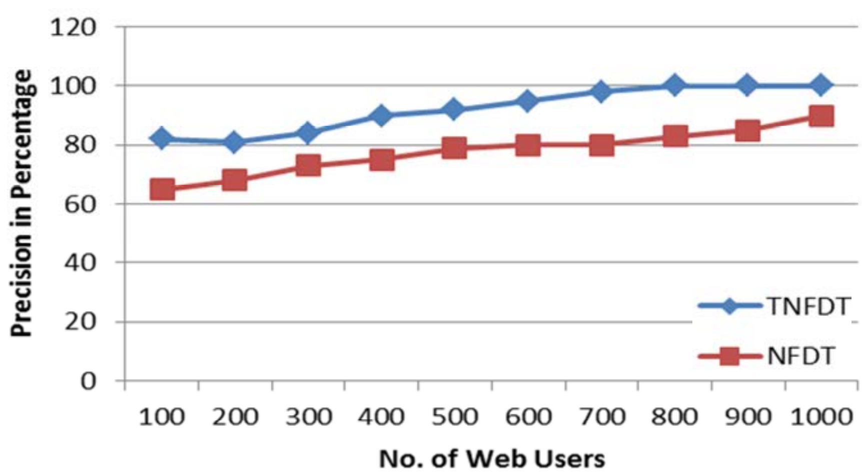

Fig. 7: Performance analysis of proposed recommendation system

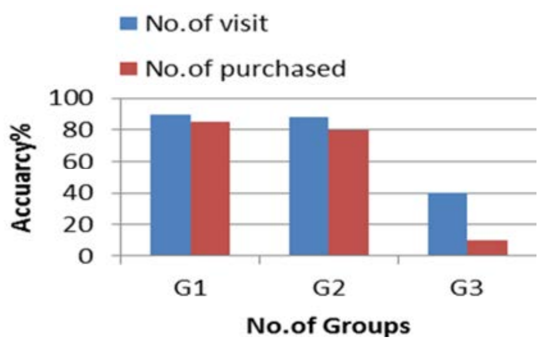

Fig. 8: Performance analysis for online purchase

CONCLUSION

In this study, a new classification algorithm called TNFDTC is proposed for the effective classification of temporal data. Moreover, new rule optimization methods are proposed for rule pruning temporal constraint satisfaction and to perform deductive inference. The data points which are classified with low confidence in this system are declared as unclassified. Conflicts are resolved by classifying reasonable number of data points and temporal constraints are used in this classifier to analyze the past and to predict the future using temporal rules. From the experiments conducted in this research work, it is observed that the proposed temporal classification algorithm provides better classification accuracy than the existing decision tree classifiers. Therefore, the decisions made by applying these algorithms are more suitable for effective social networks analysis. Further works in this direction can be the proposal of a new temporal logic for effective temporal decision making.

REFERENCES

Agrawal, R. and R. Srikant, 1995. Mining sequential patterns. Proceedings of the 11th International Conference on Data Engineering, March 6-10, 1995, Taipei, Taiwan, pp: 3-14.

Bhatt, R.B. and M. Gopal, 2006. Neuro-fuzzy decision trees. Int. J. Neural Syst., 16: 63-78.

Debnath, S., P. Mitra, N. Pal and C.L. Giles, 2005. Automatic identification of informative sections of Web pages. IEEE Trans. Knowledge Data Eng., 17: 1233-1246.

Fernandez, A., M. Calderon, E. Barrenechea, H. Bustince and F. Herrera, 2010. Solving multi-class problems with linguistic fuzzy rule based classification systems based on pairwise learning and preference relations. Fuzzy Sets Sys., 161: 3064-3080.

Kottas, T.L., Y.S. Boutalis and A.D. Karlis, 2006. New maximum power tracker for PV arrays using fuzzy controller in close cooperation with fuzzy cognitive networks. IEEE Trans. Energy Convers., 21: 793-804.

Luengo, J. and F. Herrera, 2010. Domains of competence of fuzzy rule based classification systems with data complexity measures: A case of study using a fuzzy hybrid genetic based machine learning method. Fuzzy Sets Syst., 161: 3-19.

Madadipouya, K., 2015. A new decision tree method for data mining in medicine. Adv. Comput. Intell.: Int. J., 2: 31-37.

Meamarzadeh, H., M.R. Khayyambashi and M.H. Saraee, 2009. Extracting temporal rules from medical data. Proceedings of the International Conference on Computer Technology and Development, Volume 1, November 13-15, 2009, Kota Kinabalu, Malaysia, pp: 327-331.

Nozaki, K., H. Ishibuchi and H. Tanaka, 1996. Adaptive fuzzy rule-based classification systems. IEEE Trans. Fuzzy Syst., 4: 238-250.

Pach, F.P., A. Gyenesei and J. Abonyi, 2008. Compact fuzzy association rule-based classifier. Expert Syst. Applic., 34: 2406-2416.

Priya, P.I., A. Annapurni, S.I. Gandhi, S. Muthurajkumar, M. Vijayalakshmi and A. Kannan, 2015. Intelligent classification techniques for effective prevention of inference attacks in online social networks. Aust. J. Basic Applied Sci., 9: 42-46.

Priya, P.I., D.K. Ghosh, A. Kannan and S. Ganapathy, 2014. Behaviour analysis model for social networks using genetic weighted fuzzy C-means clustering and neuro-fuzzy classifier. Int. J. Soft Comput., 9: 138-142.

Qi, X. and B.D. Davison, 2009. Web page classification: Features and algorithms. ACM Comput. Surveys, Vol. 41, 10.1145/1459352.1459357.

Stein, B. and S.M.Z. Eissen, 2008. Retrieval models for genre classification. Scand. J. Inform. Syst., 20: 93-119.

Xue, W., H. Bao, W. Xue, W. Huang and Y. Lu, 2006. Web page classification based on SVM. Proceedings of the 6th World Congress on Intelligent Control and Automation, Volume 2, June 21-23, 2006, Dalian, China, pp: 6111-6114.