International Journal of Soft Computing 9 (1): 20-26, 2014
ISSN: 1816-9503

Assessing the Medical Data using Ranking Based Weighted Fuzzy Associative Classifier

[1]N.S. Nithya and [2]K. DuraiSwamy
[1]K.S.R. College of Engineering, [2]K.S. Rangasamy College of Technology,
Tiruchengode, Tamil Nadu, India

Abstract: Fuzzy Association Rule Mining algorithm is very efficient for diagnosis, prognosis and treatment of diseases in medical field compared to other classification technique. But it suffers from exponential growth of rules produced. Identifying the most important risk factor is one of the main tasks in medical data mining. To obtain these objectives the new algorithm using information gain ranking based weight for fuzzy associative classification is proposed. The ranking of attributes eliminates irrelevant attributes and assign weight value used for assessing the risk factor of diseases. Elimination of irrelevant attributes and ranking used to extract important rules in fuzzy association rule mining which reduce the computation time and increase the classification accuracy. The results are verified using the breast cancer dataset, heart diseases dataset with different categories of attributes to demonstrate the effectiveness of the proposed approach.

Key words: Information gain, ranking based weight, heart disease prediction, breast cancer, medical data mining, fuzzy weighted support and confidence, fuzzy weighted association rule mining

INTRODUCTION

Data mining set to play an important role in tackling the data overload in medical informatics. Mining frequent patterns from large scale database is very important for data mining. Apriori algorithm is used to extract frequent patterns and association rules from transactional data base. Find frequent patterns using Apriori needs several iterations, produce large amount of inefficient item sets and can't find rarely occur events. This proposes an efficient approach based on weight factor and utility for mining high utility patterns (Dhanda, 2011).

Experts may not give the right fuzzy sets and their corresponding membership functions. This research proposes a method based on weighted pattern sequential mining dataset (Chueh, 2011). Find all significant items with W-support above the given threshold and derive rules from the item sets if an item set satisfies the min w-support then all its subsets satisfy the min W-support as well. Medical dataset is a combination of numerical and binary attributes. Crisp associative classification is not suitable for diagnosis of medical data contain either numerical (or) combination of numerical as well as binary (Somol *et al*., 2010).

Loss of information arising due to sharp partitioning like age (25, 60) is replaced by age (middle, aged) using fuzzy conversion. Conventional associative classifiers are only worked with binary attributes and expect any quantitative attributes to be converted into binary ones. Crisp association rules are not suitable for diagnosing medical diseases.

An effective way to solve this problem by using fuzzy membership function. It is done by two step process contain preprocess the data involves fuzzy partitioning the data and the second step involves generation of fuzzy association rule using iterative scaling and maximum entropy. Quantitative attributes are partitioned into several fuzzy sets by Fuzzy C-Means algorithm and membership values are generated and supervised Association Rule algorithm is used to discover interesting fuzzy association rules which are used to build classification systems (Kumar *et al*., 2011).

A wider application of mining association rules is to identify customer behavior from various business databases. For example in the medical field mining association rules can help us to predict the diseases by frequent occurrence of symptoms and causes of that patient database. Mining fuzzy association rules is finding the fuzzy item sets that frequently occur together from a given database. The fuzzy ARM works using Apriori is inefficient and slow when it comes to dealing with very large datasets (Prakash and Parvathi, 2011). It consequently converts the given dataset to fuzzy dataset heads less human interaction for even very large datasets.

The inductive learning of fuzzy rule based classification system suffers from exponential growth of

Corresponding Author: N.S. Nithya, K.S.R. College of Engineering, Tiruchengode, Tamil Nadu, India

the fuzzy rule search space when the number of patterns (or) variables becomes high (Alcala-Fdez *et al.*, 2011; Mangalampalli and Pudi, 2011; Hsu *et al.*, 2000). This growth makes the learning process more difficult and in most causes it leads to problem of scalability (in terms of time and memory consumed) and/or complexity with respect to the number of rules obtained and the number of variables included in each rule. According to the literature, many fuzzy association rule have found but still more filtration is needed to reduce the number of rules produced. In the previous survey, researchers propose novel method based on various weighted association rule that is implemented in this study (Nithya and Duraiswamy, 2011).

The proposed Fuzzy Association Rule Based Classification Method is used to obtain an accurate and compact fuzzy rule based classifier for high dimensional database with low computational cost. In this classifier, number of rules generated is reduced through improved filter based fuzzy weighted association measure.

MATERIALS AND METHODS

The proposed algorithm: The proposed classifier is developed by combining information gain and weighted fuzzy association classifier. It is done in two phase process:

- Phase 1: filter based feature ranking selection
- Phase 2: fuzzy weighted association rule generation

Filter based attribute ranking selection: Feature selection is an important process for extracting relevant attribute in medical data mining. The Feature Selection (FS) approach does not attempt to generate new features but to select the best ones from the original set of features. Depending on the outcome of a FS procedure, the result can be either a set of weighting scoring, a ranking or a subset of features (Karaolis *et al.*, 2010; Somol *et al.*, 2010; Lee *et al.*, 2011). FS leads to savings in measurements cost since some of the features are discarded and the selected features retain their original physical interpretation.

Information Gain is suitable for multi class problems and a powerful criterion for filter based approach. Filter methods asses the relevance of features as scores by looking only at the information of the data. Features can be sorted by their scores and low scoring features can be removed. Filter methods are the most commonly applied in bio-informatics studies since they are computationally simple, fast and independent of other analysis algorithms. Also, they allow features to be quantified and prioritized according to the scores which is particularly important for biological interpretation.

Info Gain is mostly used for deciding which variables to use first in a classification problem. The idea of selecting a split attribute is to select an attribute that reduces the uncertainty by the largest amount. Higher the information gain, higher are the chances of getting pure classes in a target class if split on thevariable with the highest gain. Information Gain (IG) measures the amount of information in bits about the class prediction if the only information available is the presence of a feature and the corresponding class distribution. Concretely, it measures the expected reduction in entropy.

In any prediction model, all attributes do not have same importance in predicting the class label. So, the different weights can be assigned to different attributes according to their information gain measure we will select the attribute with highest information gain as the next split attribute.

Assume there are two classes, P and N and let the set of training data (S) (with the total number of objects (s) contain p elements of class P and n elements of class N. The amount of information is defined as:

$$I = -\left(\frac{n}{s}\right)\log\left(\frac{n}{s}\right) - \left(\frac{p}{s}\right)\log\left(\frac{p}{s}\right) \qquad (1)$$

If p = n, I = 1, if p = s, I = 0. Researchers define information gain for sample S using attribute A as follows:

$$\text{Gain}(S, A) = I - \sum\nolimits_{i\in \text{values}(A)}\left(\frac{t_i}{s}\right)I_i \qquad (2)$$

Where:

I = The information before the split

$\sum_{i\in \text{values}(A)}\left(\frac{t_i}{s}\right)I_i$ = The sum of information after the split where, I_i is the information of node i and t_i is the number of objects in node i

Thus, the total information after the split is:

$$\sum\nolimits_{i\in \text{values}(A)}\left(\frac{t_i}{s}\right)I_i$$

which the sum of the information of each node is. Once researchers have computed the information gain for every remaining attribute, the attribute with the highest information gain is selected which asses the risk factor of diseases.

WEKA data mining tool (Schumaker *et al.*, 2010) provides the environment to calculate the Information Gain.

Algorithm 1:

```
Input:   C-Condition attribute of Medical datasets
         D-Decision attribute of Medical datasets
Output:  B-Selected attribute
         R-Ranking of attribute
         $W_i$-Weight value for each attribute
Information gain based attribute ranking
Step 1: Let R = Ø
Step 2: For every attribute $a \in C-B$, compute the significance of condition
        attribute a, information gain (a, B, D)
Step 3: Select the attribute which maximize the Information Gain (a, B,D),
        record it as a and $B \leftarrow B \cup \{a\}$
Step 4: If gain (a, B, D)>0 then $B \leftarrow B \cup \{a\}$, goto step 2, else goto step 5
Step 5: The set B is the selected attributes with the ranking value R based
        on information gain value
Step 6: Assign user defined weight value $W_i$ for each selected feature B
```

Fuzzy weighted association rule generation: Fuzzy weighted association rules (Fig. 1) are generated according to the information gain and fuzzy partitioning clustering method will reduce the number of rules produced incase of high dimensional datasets.

Calculation of support and confidence: The support of the rule X→Y is the percentage of transactions in T that contain X∩Y. It determines how frequent the rule is applicable to the transaction set T. The support of a rule is represented by the equation:

$$\text{supp}(X \rightarrow Y) = \frac{X \cap Y}{n} \quad (3)$$

Where:

$X \cap Y$ = The number of transactions that contain all the items of the rule

n = The total number of transactions

Equation 3 computes the relative support value but there also exists an absolute support. It works similarly but simply counts the number of transactions where the tested item set occurs without dividing it through the number of tuples (Table 1).

The confidence of a rule describes the percentage of transactions containing X which also contain Y:

$$\text{conf}(X \rightarrow Y) = \frac{X \cap Y}{X} \quad (4)$$

This is a very important measure to determine whether a rule is interesting or not. It looks at all transactions which contain a certain item or item set defined by the antecedent of the rule. Then, it computes the percentage of the transactions also including all the items contained in the consequent.

Fuzzy partition of quantitative attribute: In the medical field, datas are measured in quantitative attributes such as age of patient is 54, cholestrol level of 240 mg dL^{-1} and blood pressure 140, etc. in a fuzzy logic system these quantitative values are to be converted into fuzzy sets. Each of fuzzy sets can be viewed as a [0, 1] valued attribute, called fuzzy attribute. For example, researchers might want to partition the attribute age in to three fuzzy sets. The convenient way is to define the borders of the sets and split the overlapping part equally between the so generated fuzzy sets. We will define by age: low {0-33}, medium {27-55} and high {48-∞}.

Table 1: Fuzzy partition of heart disease data

Name	Fuzzy sets
Age (27-80)	(Low, medium, high)
Sex (value 1: male; value 0: female)	(Male, female)
Cp (1.0, 4.0)	(Normal, abnormal, severe, very severe)
Trestbps (blood pressure) (0.0-1.0)	(Low, medium, high)
Chol (serum cholesterol) (126.0-564.0)	(Low, medium, high)
Fbs (Fasting blood sugar) (94.0, 200.0) Yes if >120 mg dL^{-1}	(Low, medium, high)
Restecg (Resting electrographic results) (0.0-2.0)	(Low, medium, high)
Thalach (maximum heart rate achieved) (71.0-202.0)	(Low, medium, high)
Exang (exercise induced angina) real (0.0-1.0)	(Yes, no)
Oldpeak (ST depression induced by exercise relative to rest) (0.0-6.2)	(Low, medium, high)
Slope (the slope of the peak exercise ST segment) (1.0-3.0)	(Upslopping, flat, downsloping)
Ca (number of major vessels colored by floursopy) (0.0-3.0)	(Low, medium, high)
Thal real (3.0-7.0)	(Low, normal, high)

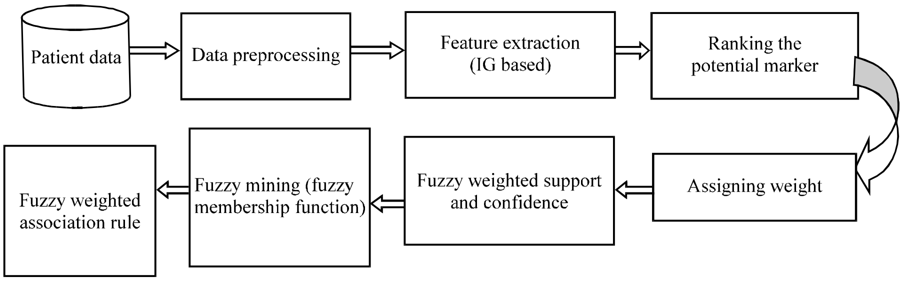

Fig. 1: Attribute ranking fuzzy weighted association rule mining

The generated fuzzy sets is shown in Fig. 2. For all areas having no overlap of the sets, the support will simply be 1 for the actual item set. If there is an overlap, the membership can be computed by using the borders of the overlapping fuzzy sets. The added support will here always sum up to 1. The formula for computing the membership varies depending on whether it is at the upper border of a set or at the lower border. For the computation of membership at the high border:

$$\mu(x) = \frac{hb(f_{ik}^{n}) - x}{hb(f_{ik}^{n}) - lb(f_{ik}^{n+1})} \quad (5)$$

For the lower border:

$$\mu(x) = \frac{x - lb(f_{ik}^{n})}{hb(f_{ik}^{n}) - lb(f_{ik}^{n+1})} \quad (6)$$

Where:

$lb(f_{ik}^{n})$ = The lower border of the set
$hb(f_{ik}^{n})$ = The high border of the set
x = The original value of the attribute

Fuzzy weighted support: For single items sets the support is the sum of the product calculation for each weighting/fuzzy member ship pair (w×f). For 2-datasets and larger the support is the sum of the products of all the weightings and fuzzy membership calculations. Thus, the sample calculation for the given data set as: <age = young, 0.25> <thal = normal, 0.5> <chest pain = very high, 1.0> <age = middle, 0.5> <chest pain = high, 0.75> <age = middle, 0.5> <thal = low, 0.25>.

The associated weighting file (based on Info Gain ranking): 0.3, 0.95 and 0.9. The support calculations will be as follows:

$$\{\text{Age}\} = \frac{(0.25\times0.3)+(0.5\times0.3)+(0.5\times0.3)}{4}$$

$$= \frac{0.75+0.15+0.15}{4} = 0.2625$$

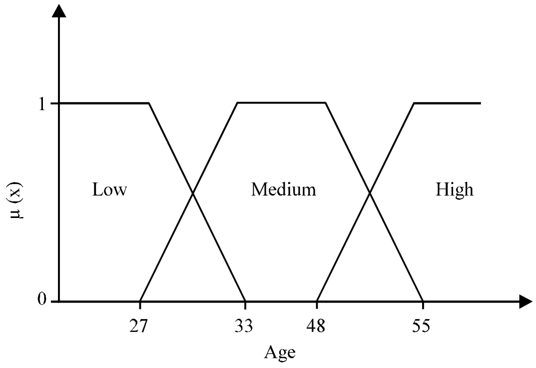

Fig. 2: Fuzzy partition of age

$$\{\text{Thal}\} = \frac{(0.5\times0.95)+(0.25\times0.95)}{4} = \frac{0.475+0.2375}{4} = 0.1781$$

$$\{\text{Chest pain}\} = \frac{(1.0\times0.9)+(0.75\times0.9)}{4} = \frac{0.9+0.675}{4} = 0.3938$$

$$\{\text{Age, Thal}\} = \frac{(0.25\times0.3\times0.5\times0.95)+(0.5+0.3\times0.25\times0.95)}{4}$$

$$= \frac{0.036+0.036}{4} = 0.018$$

$$\{\text{Age, chest pain}\} = \frac{0.5\times0.3\times0.75\times0.9}{4}$$

$$= \frac{0.1013}{4} = 0.02531$$

Fuzzy associative classifier: A fuzzy classifier (Agarwal and Singh, 2011) usually includes many fuzzy rules which state the relationship between the attributes in the antecedent part of a rule of the class label in the consequent part. The fuzzy classifier is a combination of fuzzy set and fuzzy clustering. In the antecedent part, there are many attributes which are mapped to the fuzzy sets by some continuous membership functions and in the consequent part of rules, there is a crisp set of class labels.

If $(n_1>50) \wedge \ldots \wedge (n_m, \text{yes})$ then class 1. Where, n_1, n_2..., n_m are the attributes of the colon cancer dataset, ∧ is an operator and 50, yes are quantitative attribute. So, making a fuzzy classifier needs to convert quantitative data as by Krishna and Krishna (2008) into fuzzy membership functions and rule mining. To make the fuzzy membership functions, the quantitative attributes are categorized by clustering process. Clustering can be used as preprocessing step for mining association rules. For example, when attribute R = {age, smoking}, a fuzzy item set can be like <age: middle > <smoking: high>.

After applying fuzzy sets, heart disease data set is converted into fuzzy data set. After finding the clusters, apply the Apriori in each cluster in mining association rules. Apriori algorithm as by Al-Daoud (2010) is a well known method for rule mining which can be used in Fuzzy Association Rule Mining (FARM). In this algorithm, the attributes which are characterized with a common linguistic term and have a support value more than a predefined threshold are mined and make the antecedent of the rules. Fuzzy weighted support (Muyeba *et al*., 2009) is a well-known evaluation measure associated rule mining which shows the usefulness of a fuzzy attribute set. If FWS of a fuzzy attribute set is larger than a predefined value (λ), it is called a frequent fuzzy attribute set.

Fuzzy weighted support is first used to find frequent item sets exploring its downward closure property to prune the search space. Then, fuzzy weighted confidence is used in a second step to produce rules from the frequent item sets that exceeds a min-confidence threshold. A Fuzzy association rule has a form like (X:A) (Y:B) where there is no interaction between (X:A) and (Y:B) is the consequent of a rule is a class label, the rule is considered as a classifier rule.

Algorithm 2:

```
Fuzzy weighted support:
Input:   $F_{ij}$-Fuzzy membership
         $W_{ij}$-Weight value
         Support value
         M-Size of medical datasets
         N-Total number of features
Output:  $Q_{ij}$-Fuzzy weighted support of entire size of dataset
Step 1:  For each data find the occurrence of attribute (support value).
Step 2:  Calculate the product of fuzzy membership associate with the
           attribute and weighting value based on information gain $F_{ij}\times W_{ij}$
         for the single attribute
Step3:   Calculate the sum of product of fuzzy membership with the
         corresponding weight value for two or more attributes
Step 4:  Divide this sum of product by support measurement from the
         entire data
```

Algorithm 3:

```
Fuzzy weighted association rule mining:
Input:   Medical Data Set: MDS
         A minimum support
         Threshold: $\sigma$
         An approximate factor, $\in$
         Normalized weights of the attribute
Output:  Fuzzy Weighted Frequent Pattern generation (FWAFP) Begin
Initialize FWAFP {}; // Let FWAFP
Be the result set of approximate fuzzy weighted frequent patterns. Scan MDB
once and find the global approximate weighted frequent attributes whose
maximum approximate fuzzy weighted supports ($|MWS(X)\pm\in|$) are no less
than a minimum threshold $\sigma$, remove the approximate weighted infrequent
attributes in each record by the weight ascending order. Scan the MDB again
and build a global FP-tree based on the weight ascending order. call FWAF
(FP-Tree, FWAFP)
End
```

Dataset collection: This exercise uses a dataset available from the UCI Machine Learning Repository which was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg and is known as Wisconsin Breast Cancer Original, http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+ (Original).

The Wisconsin breast cancer data is widely used to test the effectiveness of classification algorithms. The aim of the classification is to distinguish between benign and malignant cancers based on the available measurements (attributes): clump thickness, uniformity of cell size, uniformity of cell shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli and Mitoses. This database has 9 input features (attributes), 2 classes and 699 samples. The input attributes have integer value in the range [1-10]. The original database contains 699 instances, however it is advised not to use 16 of these because they are incomplete which is common in data mining. The class distribution is 65.5% benign and 34.5% malignant. To compare these data mining classification techniques Cleveland cardiovascular disease dataset from UCI repository dataset has 14 attributes and 303 records were used.

RESULTS AND DISCUSSION

Ranking based weight calculation: The initial attribute selection based on Information Gain Rank Search Method processed on 303,699 and 50 instances from the real world database of heart disease, breast cancer data, breast cancer from cancer centre using WEKA tool. WEKA is a collection of machine learning algorithms for data mining tasks. The algorithms can be directly applied to dataset from Java code. It contains tools for data visualization, data analysis and predictive modeling (Schumaker *et al*., 2010). The input files to the WEKA are datasets that is used here in CSV, format.

Table 2 gives the attribute for heart disease and Table 3 gives the attribute for breast disease. Table 4 and 5 gives the information gain based ranking of attributes and it also assigns weight value based on ranking. From

Table 2: UCI repository heart disease data

Attributes	Value range
Age	30-75
Sex	0-1
Cp	1-4
Trestbps	110-170
Chol	175-330
Restecg	0-2
Thalac	90-180
Exang	0-1
Old peak	0.2-4.4
Slope	1-3
Ca	0-3
Thal	3-7
Fbs	0-1

Table 3: UCI repository breast cancer data

Attributes	Value range
Clump thickness	1-10
Uniformity of cell size	1-10
Uniformity of cell shape	1-10
Marginal adhesion	1-10
Single epithelial cell size	1-10
Bare nuclei	1-10
Bland chromatin	1-10
Normal nuceoli	1-10
Mitoses	1-10
Class	Benign, Malignant

Table 4: Information gain based ranking using Weka tool for heart disease data

Attributes	Ranking	Weight value
Thal	0.2048	0.95
Cp	0.2017	0.90
Ca	0.1681	0.80
Old peak	0.1542	0.70
Exang	0.1391	0.60
Thalac	0.1260	0.50
Slope	0.1113	0.40
Age	0.0602	0.30
Sex	0.0573	0.20
Restecg	0.0220	0.10
Fbs	0.000	0.00
Trestbps	0.000	0.00
Chol	0.000	0.00

Table 5: Information gain based ranking using Weka tool for breast cancer data

Attributes	Rank	Weight value
Uniformity of cell size	0.674	0.9
Uniformity of cell shape	0.660	0.8
Bare nuclei	0.564	0.7
Bland chromatin	0.542	0.6
Single epithelial cell size	0.505	0.5
Normal nucleoli	0.466	0.4
Clump thickness	0.458	0.3
Marginal adhesion	0.443	0.2
Mitoses	0.198	0.1

Table 4, there are three attributes with zero gain that attributes are eliminated. Table 5 attribute 9 has the lowest gain value which is pruned during Fuzzy Weighted Association Rule generation.

Experimental results: The proposed system Ranking based Fuzzy Weighted Association Rule Mining (RFWARM) is implemented using KDD Software Tool and the result is compared with Fuzzy Association Rule Mining (FARM) for both breast cancer data and heart disease data. The entire data set is divided into testing and training data. The performance of this system is evaluated by execution time and classification accuracy.

Evaluation of execution time by varying the support value: The proposed Ranking based Fuzzy weighted Association Rule Mining algorithm is compared with Fuzzy Association Rule Mining algorithm (FARM) for both breast cancer data and heart disease data to evaluate the performance of the RWFARM algorithm. The execution time is varied according to different values of support value. Figure 3 shows the execution time for both the algorithms with different support value. Figure 3 shows the proposed RWFARM algorithm need less execution time when compared with FARM because the lowest ranking attributes get eliminated and puning is also based on this ranking.

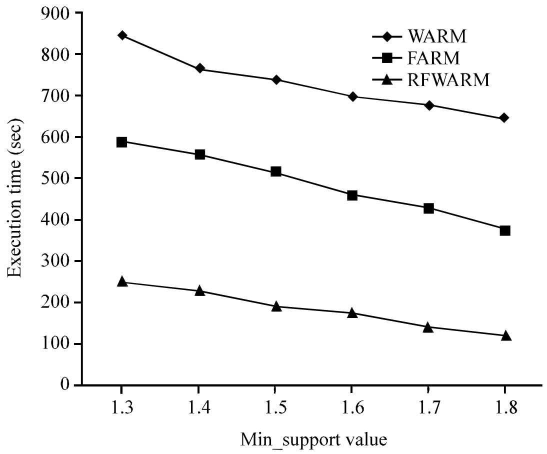

Fig. 3: Comparative analysis of execution time for min_support value

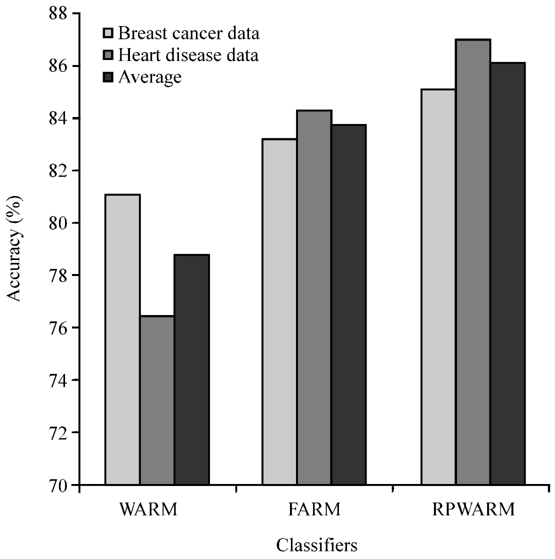

Fig. 4: Comparative analysis of accuracy

Evaluation of classification accuaracy by varying the medical diagnosis data: The classifier accuracy for both the algorithms are determined by the number of classes is correctly classified. The training and testing data set of different types of medical data sets are used to evaluate the classifier accuracy. Figure 4 shows the variation of classification accuracy of three classifiers. Figure 3 and 4, both of them shows that proposed Ranking based Fuzzy Weighted Association Rule Mining algorithm (FWARM) has better accuracy performance compared with WARM and FWARM classifiers.

CONCLUSION

The proposed Ranking Based Fuzzy Weighted Association Rule algorithm ranks the risk factors of diseases and reduces the computation time. Experimental results show that the RFWARM algorithm could improve the accuracy of diagnosis of heart disease and breast cancer data. The research is further extended by applying for prognosis and treatment of diseases in the field of medical.

REFERENCES

Agarwal, S. and L. Singh, 2011. Mining fuzzy association rule using fuzzy artmap for clustering. J. Eng. Res. Stud., 2: 76-80.

Al-Daoud, E., 2010. Cancer diagnosis using modified fuzzy network. Universal J. Comput. Sci. Eng. Technol., 1: 73-78.

Alcala-Fdez, J., R. Alcala and F. Herrera, 2011. A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning. IEEE Trans. Fuzzy Syst., 19: 857-872.

Chueh, H.E., 2011. Mining target-oriented fuzzy correlation rules to optimize telecom service management. Int. J. Comput. Sci. Inf. Technol., 3: 74-83.

Dhanda, M., 2011. An Approach to extract efficient frequent patterns from transactional database. Int. J. Eng. Sci. Technol., 3: 5652-5658.

Hsu, W., M.L. Lee, B. Liu and T.W. Ling, 2000. Exploration mining in diabetic patients databases: Findings and conclusions. Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining, August 20-23, 2000, Boston, MA., USA., pp: 430-436.

Karaolis, M.A., J.A. Moutiris, D. Hadjipanayi and C.S. Pattichis, 2010. Assessment of the risk factors of coronary heart events based on data mining with decision trees. IEEE Trans. Inf. Technol. Biomed., 14: 559-566.

Krishna, G.V. and P.R. Krishna, 2008. A novel approach for statistical and fuzzy association rule mining on quantitative data. J. Sci. Ind. Res., 67: 512-517.

Kumar, D.S., N. Ananthi and D. Yadavaram, 2011. New approach to weighted pattern sequential mining-dataset. Int. J. Sci. Eng. Res., Vol. 2.

Lee, I.H., G.H. Lushington and M. Visvanathan, 2011. A filter-based feature selection approach for identifying potential biomarkers for lung cancer. J. Clin. Bioinformat., Vol. 1.

Mangalampalli, A. and V. Pudi, 2011. Fuzzy associative rule-based approach for pattern mining and identification and pattern-based classification. Proceedings of the 20th International Conference Companion on World Wide Web, March 28-April 1, 2011, Hyderabad, India, pp: 379-384.

Muyeba, M., M.S. Khan and F. Coenen, 2009. Fuzzy weighted association rule mining with weighted support and confidence framework. Proceedings of the International Workshops on New Frontiers in Applied Data Mining, May 20-23, 2008, Osaka, Japan, pp: 49-61.

Nithya, N.S. and K. Duraiswamy, 2011. An information gain based fuzzy classifier for predictive analysis in colon cancer data. Int. J. Comput. Appl., 31: 45-48.

Prakash, S. and R. Parvathi, 2011. Qualitative approach for quantitative association rule mining using fuzzy rule set. J. Comput. Inf. Syst., 7: 1879-1885.

Schumaker, R.P., O.K. Solieman and H. Chen, 2010. Open Source Data Mining. In: Tools for Sports, Sports Data Mining, Schumaker, R.P., O.K. Solieman and H. Chen (Eds.). Springer, USA., ISBN-13: 9781441967305, pp: 89-100.

Somol, P., J. Novovicova and P. Pudil, 2010. Efficient feature subset selection and subset size optimization. In: Pattern Recognition Recent Advances, Herout, A. (Ed.). InTech, USA., ISBN-13: 9789537619909, pp: 75-97.