

Gain ratio based fuzzy weighted association rule mining classifier for medical diagnostic interface

N S NITHYA^{1,*} and K DURAISWAMY²

¹K. S. R. College of Engineering, Tiruchengode 637 215, Tamil Nadu, India

²K. S. Rangasamy College of Technology, Tiruchengode 637 215, Tamil Nadu, India
e-mail: sachinnithya@yahoo.com; drkduraiswamy@yahoo.co.in

MS received 3 June 2013; revised 17 August 2013; accepted 21 August 2013

Abstract. The health care environment still needs knowledge based discovery for handling wealth of data. Extraction of the potential causes of the diseases is the most important factor for medical data mining. Fuzzy association rule mining is well-performed better than traditional classifiers but it suffers from the exponential growth of the rules produced. In the past, we have proposed an information gain based fuzzy association rule mining algorithm for extracting both association rules and membership functions of medical data to reduce the rules. It used a ranking based weight value to identify the potential attribute. When we take a large number of distinct values, the computation of information gain value is not feasible. In this paper, an enhanced approach, called gain ratio based fuzzy weighted association rule mining, is thus proposed for distinct diseases and also increase the learning time of the previous one. Experimental results show that there is a marginal improvement in the attribute selection process and also improvement in the classifier accuracy. The system has been implemented in Java platform and verified by using benchmark data from the UCI machine learning repository.

Keywords. Feature selection; gain ratio; ranking based weight; medical data mining; fuzzy weighted support and confidence; UCI repository.

1. Introduction

Data mining is the non-trivial and proactive process of extracting valid, comprehensible and interesting knowledge from data. Human medical data are the most rewarding and difficult of all the biological data to mine and analyse. Data mining techniques are used to diagnose the disease accurately. Data mining technique in medicine is distinct from that of other fields due to the special nature of data: heterogeneous with ethical, legal and social constraints. The most commonly used technique is classification and prediction with different methods applied for different cases. Associated rules describe the data in the database.

*For correspondence

In medical applications (Zubair Rahman & Balasubramanie 2009), an association rule may be more interesting if it reveals the relationship among some useful concepts, such as 'high pressure', 'low cholesterol' and 'normal sugar'. Thus interesting concepts are defined using fuzzy items and are interpreted based on fuzzy set. We combine these fuzzy terms with the association rule mining as fuzzy association rule mining, for example from the terms 'Age = middle' and 'chest pain = high', 'a risk level of heart disease = medium' is a fuzzy quantitative association rule, where 'middle', 'high' and 'medium' are fuzzy terms. This is a process of fuzzifying numerical numbers in linguistic terms which is often used to reduce the information overload in the medical decision making process. One way of determining this membership is either from the expert's opinion or by people's perception. In developing fuzzy association rule mining algorithm there are two important tasks, one is to find the appropriate membership function and second to generate the fuzzy association rule automatically.

2. Related work

The calculation of fuzzy membership function of fuzzy expert system is based on domain knowledge. The disadvantage of this method is that the system is not clearly giving overlapping membership function (Oladipupo *et al* 2012). del Jesus *et al* (2007) proposed a subgroup discovery based on subjective quality measures like usefulness, actionability, operationability, unexpectedness, novelty and redundancy. The rule contains 'cond->class' and the antecedent part is the conjunction of features selected from the features describing the training instances. Dimension of search space has an exponential relation to the number of features and values considered. In Shaik & Yeasin (2009) proposed ASI [Adaptive Subspace Iteration] adapts one to many mappings between genes and cellular functions which relates associated sub functions. Chen *et al* (2008) proposed a cluster-based evaluation in fuzzy genetic data mining. This proposed technique speed up the evaluation process of large item sets for chromosome division using K-means clustering approach.

The clustered chromosome represents the membership function which is to calculate fitness value. It is not suitable for complex problems and only fit for large item sets and mathematical proof is not confirmed. Mabu *et al* (2011) describes a Genetic Network programming of directed graph structure based fuzzy class association rule mining can deal with both continuous and discrete attributes to detect intrusion effectively. Feature extraction selects feature which classifies intrusion is not clearly defined. Then Kaya & Alhaji (2005) proposed a method based on utilizing the mining process effectively by fuzziness and online analytical processing [OLAP] based mining. They introduced the fuzzy data cube OLAP architecture which facilitates effective storage and processing of the state information reported by mobile agents. They select the attribute by using multilevel association rule mining which discovers not many experienced states. But they were not considered about the number of association rules produced. Generalization of all the state information and selection based on association rule mining leads to increase the number of rule generation. Marin *et al* (2008) proposed online analytical mining based association rule for imprecision data. They used a fuzzy multidimensional model based association rule extraction which is capable of representing and managing imprecise data. They select the attribute by taking inter-dimensional, intra-dimensional and hybrid dimension. To reduce the number of rules produced, they consider lossless generalization. Let us consider the sample of two rules:

- 1) If [Patient is 13 years old], then [Severity is low].
- 2) If [Patient is 20 years old], then [Severity is low].

In this case both the rules were translated into the same one as in the following generalization. If [Patient is Young], then [Severity is low].

The disadvantage of this work is in considering the generalization at each step and evolves a local minimum which is suitable only for analytic purpose and not for future prediction. If the new attribute is entered, the generalization process does not consider the new one. To avoid the costly generation of candidate sets of quantitative association rules (Lotfi & Sadreddini Kong 2009) proposed mutual information based fuzzy association rule mining algorithm. The mutual information evaluates the attribute with the strong informative relationship between the attributes in which potential frequent item set will be discovered. This method assigns a different support value at each level and the model works well with problems involving uncertainty in data relationship of fuzzy set concepts. The disadvantage of this method is to assign different support value at each level of abstraction to produce a large number of rules generated as a specific manner and not as a generalized one (Gautam *et al* 2010). In this paper (Zubair Rahman & Balasubramanie 2009) weighted support is calculated by weight and frequency of occurrence of an item in the translation. The method uses N processors for partitioning the database and generates closed frequent item sets in parallel. The disadvantage of this method is to support weight calculation using N processors which require efficient and parallel processing capacity. Weight calculation mainly depends upon frequency of occurrence which may lead to missing of some attributes which has less frequency of occurrence but more important attribute.

The proposed fuzzy weighted association rule based classification method is used to obtain an accurate and compact fuzzy rule based classifier for high dimensional database with a low computational cost. In this classification, the number of rules generated is reduced through improved gain ratio based fuzzy weighted association measure. The potential causes of the diseases are assessed by gain ratio and weightage is given to that attribute. The proposed classifier is developed by combining gain ratio and weighted fuzzy association classifier. The remaining parts of this paper are organized as follows. The gain ratio ranking based feature selection is explained in section 3. The fuzzy weighted association rule generation is described in section 4. The experimental results are stated in section 5. The conclusion is given in section 6.

3. Gain ratio based attribute ranking selection

Exploring the hidden patterns in the datasets of the medical field is the tedious task in medical data mining. These patterns can be utilized for clinical diagnosis. Data preprocessing includes data cleaning, data integration, data transformation and data reduction. These data preprocessing techniques can substantially improve the overall quality of the interesting patterns mined and/or the time required for the actual mining. Data preprocessing is important for knowledge discovery process as quality decisions is based on quality data. The aim of data reduction is to find a minimum set of attributes such that the resulting probability distribution of the data classes is as close as possible to the original distribution obtained using all attributes. Mining on the reduced set of attributes has additional benefits. It reduces the number of attributes appearing in the discovered patterns, helping to make the patterns easier to understand. Further, it improves the accuracy of classification and learning runtime (Han & Kamber 2001).

Split method is the most important component of decision tree learner. To attain high predictive accuracy in different situations, the split method (information gain ratio) is the best one. The

information gain measure is biased towards tests with many outcomes. The major drawback of using information gain is that it tends to choose attributes with large numbers of distinct values over attributes with fewer values even though the latter is more informative (Asha *et al* 2012). For example, consider an attribute IE, name of the disease in the patient database. A split on a disease name would result in a large number of partitions; as each record in the database has a different name for different patients. So the information required to classify database with this partitioning would be of little importance and such a partition is useless for classification.

C4.5, a successor of ID3 (<http://www.cs.waikato.ac.nz/~ml/weka/>), uses an extension to gain information known as a gain ratio (GR), which attempts to overcome the bias. The WEKA (Quinlan 1986) classifier package has its own version of C4.5 known as J4.8. We have used J4.8 to identify the significant attributes. Let D be a set consisting of ' d ' data samples with n distinct classes. Then the expected information needed to classify the sample is given by

$$I(D) = - \sum_{i=1}^n p_i \log_2 p_i, \quad (1)$$

where p_i is the probability of an arbitrary sample which belongs to class C_i . Let attribute A have ' V ' distinct values. Let ' d_{ij} ' be number of samples of class C_i in a subset D_j . D_j contains those samples in D that have value ' a_j ' of A . The entropy based on partitioning into subsets by A , is given by

$$E(A) = - \sum_{i=1}^n I(D) \frac{d_{1i} + d_{2i} + \dots + d_{mi}}{d}. \quad (2)$$

The encoding information that would be gained by branching A is

$$Gain(A) = I(D) - E(A). \quad (3)$$

C4.5 applies a kind of normalization to information gain using a 'split information' value which is defined analogously with $Info(D)$ as

$$Split\ info_A(D) = - \sum_{j=1}^v \left(\frac{|D_j|}{|D|} \right) \log_2 \left(\frac{|D_j|}{|D|} \right). \quad (4)$$

This value represents the information computed by splitting the dataset D , into v partitions, corresponding to the v outcomes of a test on attribute A (Han & Kamber 2001). For each possible outcome, it considers the outcome of the number of tuples with respect to the total number of tuples in D . The gain ratio is defined as

$$Gain\ ratio(A) = \frac{Gain(A)}{Split\ info(A)}. \quad (5)$$

Attribute with maximum gain ratio is selected as the splitting attribute. WEKA data mining tool (Schumaker *et al* 2010) provides the environment to calculate the information gain ratio. The gain ratio based ranking is used as a user defined weight value for each potential attribute as shown in algorithm 1.

4. Fuzzy weighted association rule generation

Fuzzy Weighted Association Rules are generated by combining gain ratio based weight and fuzzy association rule mining. The first step of rule generation is to partition the data into fuzzy

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**Gain ratio based attribute ranking****Step 1:** Let $R = \emptyset$;**Step 2:** For every attribute $a \in C - B$, compute the significance of condition attributes a, information gain ratio (a, B, D);**Step 3:** Select the attribute which maximize the information gain ratio (a, B, D), record it as a; and $B \leftarrow B \cup \{a\}$;**Step 4:** If gain ratio (a, B, D) > 0, then $B \leftarrow B \cup \{a\}$, go to Step 2, else go to Step 5;**Step 5:** The set B is the selected attribute with the ranking value R based on information gain ratio value;**Step 6:** Assign user defined weight W_i value based on ranging value R for each selected feature B.

membership values and calculate the fuzzy weighted support. The next step is to generate the fuzzy weighted association rule.

4.1 Fuzzy partition of quantitative attribute

In the medical field, data are measured in quantitative attributes, such as the patient age as 54, cholesterol level as 240 mg/dl and blood pressure as 140. 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, if we want to partition the attribute age into three fuzzy sets, the most convenient way is to define the borders of the sets and split the overlapping part equally between the so generated fuzzy sets. Thus we define age by

$$\text{Age/Low} = \{0-33\}$$

$$\text{Age/Medium} = \{27-55\}$$

$$\text{Age/High} = \{48-\infty\}.$$

The generated fuzzy sets are shown in figure 1. For the areas where there is 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 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

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

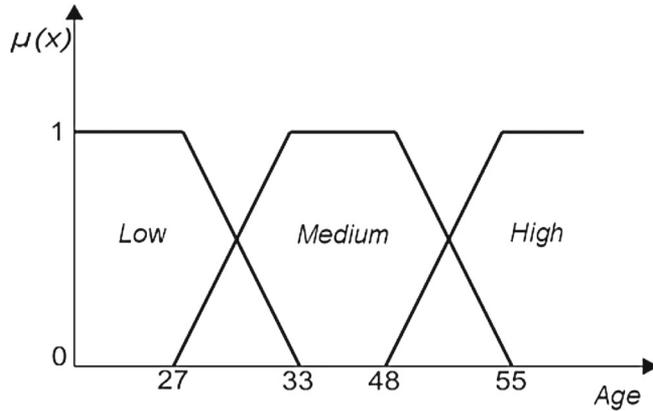


Figure 1. Fuzzy partition of age.

For the computation of membership at the high bidder,

$$\mu(x) = \frac{x - lb(f_{ik}^n)}{hb(f_{ik}^n) - lb(f_{ik}^{n+1})}, \quad (7)$$

for the lower border.

Where $lb(f_{ik}^n)$ and $hb(f_{ik}^n)$ are the lower and high borders of the set and x is the original value of the attribute.

4.2 Calculation of support and confidence

The support of the rule $X \rightarrow Y$ is the percentage of transactions in T that contain $X \cap Y$. It determines how frequently the rule is applicable to the transaction set T . The support of a rule is represented by the formula

$$supp(X \rightarrow Y) = \frac{X \cap Y}{n}, \quad (8)$$

where $X \cap Y$ is the number of transactions that contain all the items of the rule and n is the total number of transactions. The above formula computes the relative support value, but there exists an absolute support also. It works similarly but simply counts the number of transactions where the tested item set occurs without dividing it through the number of tuples.

The confidence of a rule describes the percentage of transactions containing X and also Y .

$$conf(X \rightarrow Y) = \frac{X \cap Y}{X}. \quad (9)$$

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 which includes all the items contained in the consequent.

4.3 Fuzzy weighted support

For single item sets, the support is the sum of the product calculation of each weighting/fuzzy membership pair ($w \times f$). 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 dataset is shown below

$$\langle \text{age} = \text{young}, 0.25 \rangle \langle \text{thal} = \text{normal}, 0.5 \rangle \langle \text{chest pain} = \text{veryhigh}, 1.0 \rangle$$

$$\langle \text{age} = \text{middle}, 0.5 \rangle \langle \text{chest pain} = \text{high}, 0.75 \rangle$$

$$\langle \text{age} = \text{middle}, 0.5 \rangle \langle \text{thal} = \text{low}, 0.25 \rangle$$

The associated weighting file (based on gain ratio ranking) such as 0.3, 0.95 and 0.9.

The support calculations will be as follows:

$$\begin{aligned} \{\text{age}\} &= ((0.25 * 0.3) + (0.5 * 0.3) + (0.5 * 0.3))/4 \\ &= (0.75 + 0.15 + 0.15)/4 = 0.2625 \end{aligned}$$

$$\{\text{thal}\} = ((0.5 * 0.95) + (0.25 * 0.95))/4 = (0.475 + 0.2375)/4 = 0.1781$$

$$\{\text{chest pain}\} = ((1.0 * 0.9) + (0.75 * 0.9))/4 = (0.9 + 0.675)/4 = 0.3938$$

$$\begin{aligned} \{\text{age, thal}\} &= ((0.25 * 0.3 * 0.5 * 0.95) + (0.5 * 0.3 * 0.25 * 0.95))/4 \\ &= (0.036 + 0.036)/4 = 0.018 \end{aligned}$$

$$\{\text{age, chestpain}\} = ((0.5 * 0.3 * 0.75 * 0.9))/4 = 0.1013/4 = 0.02531.$$

4.4 Fuzzy association rule generation

The steps presented in the above section calculate the fuzzy weighted support for fuzzy weighted association rule mining and fuzzy partition. A fuzzy classifier 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 (Agarwal & Singh 2011). 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 $(n_1, 50) \wedge (n_2, 200) \dots \dots (n_{m-1}, 150) \wedge (n_m, \text{yes})$, then class 1. Where, n_1, n_2, \dots, n_m is the attributes of the colon cancer dataset, \wedge is an operator and 50, 200, 150, yes are quantitative attribute. So fuzzy classifier needs the conversion of quantitative data as in Vijay Krishna & Radha 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 $R = \{\text{age, smoking}\}$, a fuzzy item set can be like $\langle \text{age: middle} \rangle \langle \text{smoking: high} \rangle$.

After applying fuzzy sets, heart disease dataset is converted into fuzzy dataset. After finding the clusters, apply the Apriori in each cluster in mining association rules. Apriori algorithm

as in AI-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 with a common linguistic term and a support value more than a predefined threshold are mined and make the antecedent of the rules. Fuzzy weighted support is a well-known evaluation measure associated rule mining which shows the usefulness of a fuzzy attribute set (Muyeba *et al* 2010). If fuzzy weighted support of a fuzzy attribute set is larger than a predefined value (λ), it is called a frequent fuzzy attribute set.

Fuzzy weighted support is used initially to find frequent item sets exploring its downward closure property to prune the search space. Then fuzzy weighted confidence is used to produce rules from the frequent item sets that exceeds a minimum confidence threshold. A fuzzy association rule has a form like $(X : A) (Y : B)$ where there is no interaction between antecedent $(X : A)$ and $(Y : B)$ is the consequent part of a rule with a class label, then 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 associated with the attribute and weighting value based on the information gain ratio $F_{ij} \times W_{ij}$ for the single attribute;

Step 3: 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 dataset: MDS

A minimum support threshold: σ

An approximate factor, ϵ

Normalized weights of the attribute

Output:

Fuzzy weighted frequent pattern generation (FWAFP)

Step 1: Initialize FWAFP $\{\}$; Scan MDS once and find the global approximate weighted frequent attributes whose maximum approximate fuzzy weighted supports ($|MWS(X) \pm \epsilon|$) are no less than a minimum threshold σ ;

Step 2: Remove the approximate weighted infrequent attributes in each record by the weight ascending order;

Step 3: Scan the MDB again and build a global FP-tree based on the weight ascending order; Call FWAFP (FP-tree, FWAFP);

Table 1. Medical datasets used for the experiment.

Medical problem	No. of instances	Numeric	Nominal	No. of classes
Breast cancer	699	0	2	2
Heart disease	303	6	7	2
Iris	150	4	0	3
Liver	345	7	0	2
Pima Indian diabetes	768	8	0	2

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, (<http://archive.ics.uci.edu/ml/datasets.html>).

The Wisconsin breast cancer data, heart disease, iris, liver and pima Indian diabetes datasets are used to test the effectiveness of classification algorithms. The aim of the classification is to distinguish between diseases based on the available measurements (attributes). This database has 9 input features (attributes), 2 classes and 699 samples of breast cancer data, 13 input features (attributes), 2 classes and 303 samples of heart disease data, 4 input features (attributes), 2 classes and 150 samples of iris data, 10 input features (attributes), 2 classes and 345 samples of liver cancer data, 8 input features (attributes), 2 classes and 768 samples of breast cancer data as shown in table 1. The input attributes have both numerical and nominal attribute. The incomplete and noisy data are removed during data preprocessing.

5. Experimental set-up*5.1 Gain ratio based weight calculation*

Our initial attribute selection based on gain ratio rank search method processed on 303, 699, 150, 345 and 768 instances from the real world database of heart disease, breast cancer data, iris data, liver cancer data, Pima Indian diabetes data from UCI repository using WEKA tool.

WEKA is a collection of machine learning algorithms for data mining tasks (Schumaker *et al* 2010). The algorithms can directly be applied to the dataset from Java code. It contains tools for data visualization, data analysis and predictive modelling. The input files of the WEKA are datasets that are used here in CSV format.

Table 2 gives the attribute ranking based on the information gain method from all datasets and table 3 gives the attribute ranking based on a gain ratio from all datasets. From this table, we obtain that some of the attributes in the heart disease dataset and a liver dataset has information gain and gain ratio as zero value and but for some of the attributes the values are different. According to feature selection process, there is a little variation between the gain and the gain ratio method and the classification accuracy is also slightly improved when compared with the information gain that is shown in further results.

The individual ranking of the attributes is used to assess the risk factors of the diseases by assigning weight value based on the rank value. For example, from the table 3 direct bilirubin attribute of liver data has highest rank value 0.4 that is assigned as the weight value for that attribute. The weight value is combined with a fuzzy support value which is used to prune the number of rules generated by fuzzy association rule mining.

5.2 Experimental results

The proposed system gain ratio ranking based fuzzy weighted association rule mining (GRRFWARM) is implemented using KDD software Tool and the result is compared with fuzzy association rule mining based on information gain (FARM) for 5 datasets obtained from the UCI repository. The entire dataset is divided into testing and training data. The performance of this system is evaluated by execution time and classification accuracy.

5.3 Evaluation of execution time by varying the support value

The proposed gain ratio ranking based fuzzy weighted association rule mining algorithm is compared with fuzzy association rule mining algorithm based on information gain (IGFWARM), weighted association rule mining (WARM), fuzzy association rule mining (FARM) for 5 different medical datasets to evaluate the performance of the GRRFWARM algorithm. The execution time is varied according to the different values of support value. Figure 2 shows the execution time for the four algorithms with different support value. Figure 2 shows that the proposed GRRFWARM algorithm needs less execution time when compared to other fuzzy associative classifiers because the lowest ranking attributes get eliminated and pruning is also based on this ranking.

5.4 Evaluation of classification accuracy by varying the medical diagnosis data

The classifier accuracy of both the algorithms are determined by the number of classes correctly classified. The training and testing dataset of 5 different types of medical datasets are used to evaluate the classifier accuracy. Figure 3 shows the variation of classification accuracy of four classifiers. Figure 3 shows that proposed gain ratio ranking based fuzzy weighted association rule mining algorithm which has better accuracy performance when compared to other fuzzy associative classifiers.

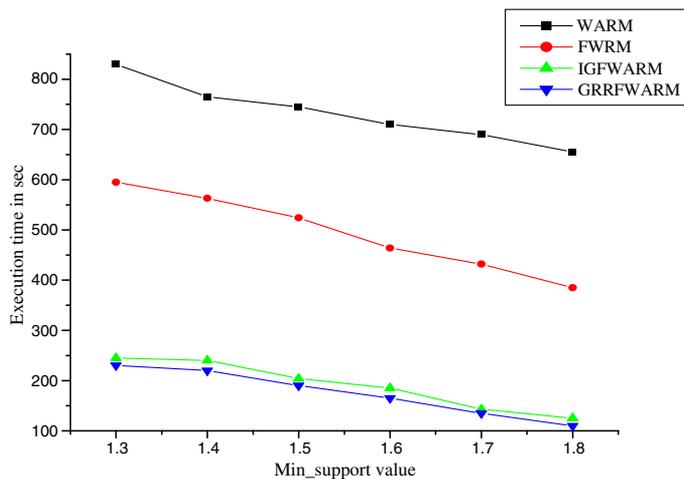


Figure 2. Comparative analysis of execution time for min_support value for heart disease data.

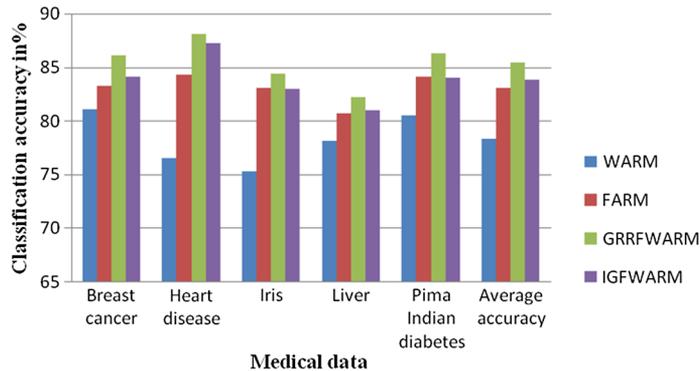


Figure 3. Comparative analysis of accuracy.

6. Conclusion

In this paper, gain ratio ranking based fuzzy weighted association rule mining algorithm was proposed. The proposed system can be applied to identify the potential causes of any real time detecting systems. The framework was outlined as follows. First, the potential risk factors of the diseases are ranked by gain ratio. According to the ranking, the weight value is assigned to potential attributes and lowest gain ratio attributes get eliminated in the second step. In the third step, the weight value is combined with the fuzzy membership function and support value generates the fuzzy association rules for classification of diseases. Pruning is done by supporting threshold value based on weight value.

This approach uses the gain ratio based ranker feature selection for assessing the risk factors of the diseases. The proposed gain ratio based ranker fuzzy weighted association rule mining reduce the computation time, reduces the exponential growth of rules produced by fuzzy association rule mining and increase the accuracy of the classification. This work is further extended to calculate the weight value using the automatic expert system. It is also extended with prognosis, treatment of diseases and also for ranking the intrusion threats in wireless networks.

References

- Agarwal S and Singh L 2011 Mining fuzzy association rule using fuzzy artmap for clustering. *J. Eng. Res. Stud.* 2: 76–80
- AI-Daoud E 2010 Cancer diagnosis using modified fuzzy network. *Universal, J. Comp. Sci. Engg. Tech.* 1(2): 73–78
- Asha G K, Manjunath A S and Jayaram M A 2012 A comparative study of attribute selection using gain ratio and correlation based feature selection. *Int. J. Inf. Technol. Knowl. Manag.* 2: 271–277
- Chen C-H, Tseng V S and Hong T-P 2008 Cluster-based evaluation in fuzzy-genetic data mining. *IEEE Trans. Fuzzy Syst.* 16(1): 249
- del Jesus M J, González P, Herrera F and Mesonero M 2007 Evolutionary fuzzy rule induction process for subgroup discovery: A case study in marketing. *IEEE Trans. Fuzzy Syst.* 15(4): 578–590
- Gautam P, Neelu Khare K R and Pardasani A 2010 Model for mining multilevel fuzzy association rule in database. *J. Comput.* 2(1): 58–64
- Han J and Kamber M 2001 *Data mining: concepts and techniques*. San Francisco: Morgan Kaufmann Publishers, pp 336–341

- Kaya M and Alhajj R 2005 Fuzzy OLAP Association rules mining-based modular reinforcement learning approach for multiagent systems. *IEEE Trans. Syst. Man Cybern.—Part B: Cybern.* 35(2): 326–338
- Lotfi S and Sadreddini Kong M H 2009 Mining fuzzy association rules using mutual information. In: *Proc. Internat. Multi Conf. Eng. Comp. Scientists 2009*, vol I, IMECS 2009, Hong Kong, pp 978–988
- Mabu S, Chen C, Lu N, Shimada K and Hirasawa K 2011 An intrusion-detection model based on fuzzy class-association-rule mining using genetic network programming. *IEEE Trans. Syst. Man Cybern.—Part C: Appl. Rev.* 41(1): 495–505
- Marin N, Molina C, Serrano J M and Amparo Vila M A 2008 A complexity guided algorithm for association rule extraction on fuzzy data cubes. *IEEE Trans. Fuzzy Syst.* 16(3): 779–794
- Muyeba M, Khan M S and Coenen F 2010 *Fuzzy weighted association rule mining with weighted support and confidence framework*. Scientific Literature Digital Library and Search Engine, pp 1–13
- Oladipupo O O, Uwadia C O and Ayo C K 2012 Improving medical rule-based expert systems comprehensibility: Fuzzy association rule mining approach. *Int. J. Artif. Intell. Soft Comput.* 3(1): 29–38
- Quinlan J R 1986 *Induction of decision tress*. Machine Learning, Kluwer Academic Publishers, pp 81–106
- Schumaker R P, Solieman O K and Chen H 2010 Sports Data Mining, chap. 9. *Open source data mining tools for sports*. Springer, pp 89–90
- Shaik J and Yeasin M 2009 Fuzzy-adaptive-subspace-iteration-based two-way clustering of microarray data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 6(2): 244–259
- Vijay Krishna V and Radha Krishna P 2008 A novel approach for statistical and fuzzy association rule mining on quantitative data. *J. Sci. Indu. Res.* 67: 512–517
- Zubair Rahman A M J Md and Balasubramanie P 2009 Weighted support association rule mining using closed itemset lattices in parallel. *IJCSNS Int. J. Comput. Sci. Netw. Secur.* 9(3): 247–253