A Fuzzy Associative Classification with Evolutionary Algorithm based Rule Selection

D. Sindhu Bairavi [1] , Prof. R. Sumathi [2] M.E., (Ph.D.) and P. Dhana Lakshmi [3]

[1] Dept of Computer Science & Engg , J.J. College of Engg & Technology, Tiruchirapalli-620 009, sindhu.bairavi3@gmail.com.

[2] Prof & Head Of the Department, Dept of Computer Science & Engg , J.J. College of Engg &Technology,Tiruchirapalli-620 009, sumathi_rajmohan@yahoo.com.

[3] Assist Prof, Dept of Information Technology, J.J. College of Engg & Technology, Tiruchirapalli-620 009, Gana-lxm@yahoo.co.in.

Abstract

FRBCS suffers from exponential growth of the fuzzy rule search space when the number of patterns and/or variables be-comes high. This growth makes the learning process is more difficult and, in most of the cases, it leads to problems of scalability (in terms of the time and memory consumed) and/or complexity (with respect to the number of rules obtained and the number of variables included in each rule).FARC method for high-dimensional problems, this is based on three stages to obtain an accurate and compact fuzzy rule-based classifier. This FARC method limits the order of the associations in the association rule extraction and considers the use of subgroup discovery, this is based on an improved weighted relative accuracy measure to preselect the most interesting rules before a genetic post processing process for rule selection and parameter tuning.

1. Introduction

Classification can be classified into traditional techniques and recent techniques. Traditional techniques such as Neural Network, Bayesian Network, Bayesian Belief Network and Decision Tree Induction. A recent technique consists of associative classification that can be further classified into fuzzy based and non fuzzy based method.

A Neural Network (NN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the process information and brain. The key element of this system is the novel structure of the information processing system. A large number of highly interconnected processing elements (neurones) working in unison to solve specific problems. Bayesian Network is a probabilistic graphical model (a type of statistical model) that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG). For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. The network can be used to compute the probabilities of the presence of various diseases for given symptoms.

The decision tree includes root node, leaf and branch node. Each internal node represents a test on attribute, each branch represents the outcome of test and each leaf node holds the class label. The topmost node of the tree is the root node. Fuzzy associative classifier this classifier is used for actual classification. Before the classifier training can ensue, need to extract the fuzzy frequent item sets from fuzzy dataset E.

1.1. Association Rule Mining

Association rule mining is a simple probabilistic statement about the co-occurrence of certain events in a database, and is mainly applicable to sparse transaction datasets. The interestingness of an association rule is quantified by two measures known as support and confidence. The support of an association rule is a measure of coverage denoted by the number of instances for which the association rule predicts correctly. On other hand, the confidence of an association rule is a measure of accuracy denoted by the ratio of the number of instances that it predicts correctly to the number of instances to which it applies.

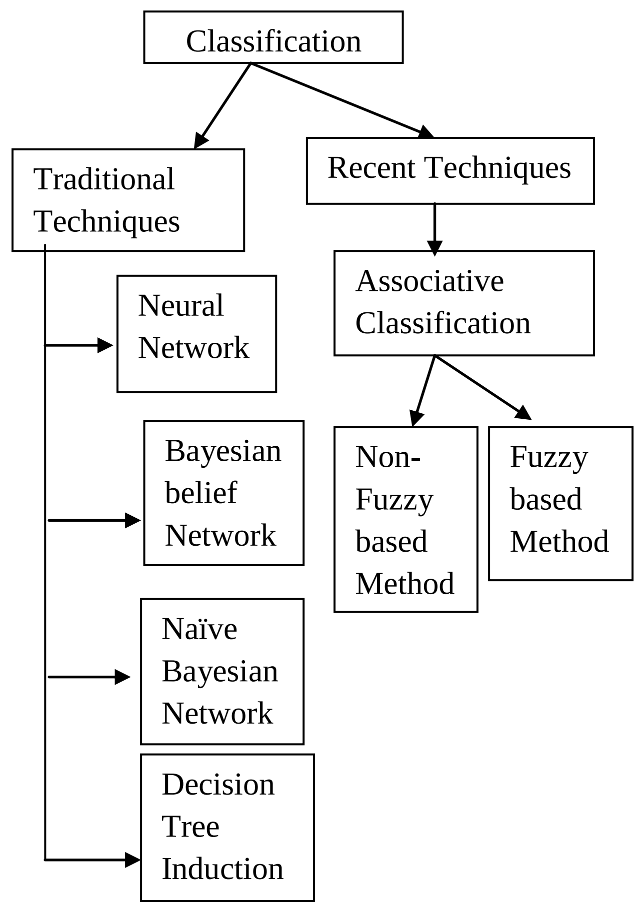

Figure 1.1 Taxonomy of Classification

Association rule mining is one of the most widely used functionalities in data mining. A common example of association rule mining is the market basket analysis. Market basket analysis is used in determining the buying habits of the customers by looking at the various association and combinations of the items they have purchased together.

2. Related Work

2.1. Elicitation of classification rules by fuzzy data mining

Data mining is the exploration and analysis of the data in order to discover meaningful patterns. To propose a fuzzy data mining method that can automatically find a set of fuzzy if–then rules for classification problems. Actually, data mining problems involving classification can be viewed within a common framework of rule discovery. The advantage for mining fuzzy if–then rules for classification problems is that knowledge acquisition can be achieved for users by carefully checking these rules discovered from training patterns. Additionally, data mining can also ease the knowledge acquisition bottleneck in building prototype expert systems or rule-based systems. The discovery of association rule is an important topic in data mining techniques. In addition, association rules elicited from transaction databases have been applied to help decision makers determine which items are frequently purchased to get her by customers. In Apriori algorithm consisting two phases:

In the first phase, frequent item sets are generated, whereas a candidate k – item set (kX_1) containing k items, is frequent (i.e., frequent k-item set) if its support is larger than or equal to a user-specified minimum support. In the second phase, association rules are generated by frequent item sets discovered in the first phase. Additionally, the comprehensibility of fuzzy representation by human users is also a criterion in designing a fuzzy system.

2.2.Compact Fuzzy Rule Base

1. Feature selection: A genetic feature selection process is used to determine a set of feature subsets. They proposed an irrelevant feature elimination algorithm based on the analysis of class regions generated by a fuzzy classifier.
2. Rule selection: GA-based method to minimize the number of linguistic

fuzzy rules for high-dimensional fuzzy classifiers.

3. Selecting the best one rule at a time iteratively.
4. Partitioning of feature space
5. Fuzzy clustering with model reduction.

Intelligent Genetic Algorithm (IGA)

Initiation: Randomly to generate an initial population with individuals. Evaluation: To evaluate fitness values of all individuals. Selection: To use the simple ranking selection that replaces the worst. Individuals with the best. That Individuals to form a new population where is a selection probability. Let be the best individual in the population.Crossover: To Randomly select Individuals including, where is a crossover probability. Perform intelligent crossover operations for all selected pairs of parents.

Mutation: Apply a conventional bit-inverse mutation operator to the population using a mutation probability. To ensure the best fitness value from deteriorating, mutation technique is not applied to the best individual.

2.3.Adaptive Fuzzy Rule Based classifier

They have proposed an adaptive method to construct a fuzzy rule-based classification system with high performance for pattern classification problems. The proposed method consists of two steps: an error correction based learning procedure and an additional learning procedure. On the contrary, when a pattern is correctly classified then the grade of certainty is increased. Because the error correction-based learning method is not meaningful after all the given patterns are correctly classified, it cannot adjust a classification boundary in some case. To acquire a more intuitively acceptable boundary, They have been proposed an additional learning procedure. They also proposed a method for selecting the most significant fuzzy rules by pruning unnecessary or unwanted fuzzy rules, which consists of the error correction-based learning procedure and the concept of forgetting. But in this construct a compact fuzzy rule based classification system with high performance.

3. Problem Definition

The inductive learning of FRBCSs suffers from exponential growth of the fuzzy rule search space. This growth makes the learning process more difficult and, in most of the cases, it leads to the problems of scalability (in terms of the time and memory consumed) and/or complexity (with respect to the number of rules obtained and the number of variables included in each rule).

Association discovery is one of the most common Data Mining techniques used to extract interesting knowledge from large datasets. Many efforts have been made to use its advantages for classification under the name of associative classification. Association discovery aims to find interesting relationships between the different items in a database, while classification aims to discover a model from training data that can be used to predict the class of test patterns. Both association discovery and classification rules mining are essential in practical Data Mining applications and their integration could result in greater savings and convenience for the user.

A associative classification system is constructed in two stages:

1) To discovering the association rules inherent in a database;

2) To selecting a small set of relevant association rules to construct a classifier.

In order to enhance the interpretability of the obtained classification rules and to avoid unnatural boundaries in the partitioning of the attributes, different studies have been presented to obtain classification systems based on fuzzy association rules. So this problem can be achieved by presenting a Fuzzy Association Rule-based Classification method for High-Dimensional problems (FARCHD) to obtain an accurate and compact fuzzy rule-based classifier with a low computational cost.

4. Fuzzy Association Rule-based Classification method for High-Dimensional problems (FARCHD)

To obtain an accurate and compact fuzzy rule-based classifier with a low computational cost. This method is based on three stages:

1) Generate Fuzzy association rule for classification
2) Candidate rule filtering
3) Genetic rule selection using GA

4.1 Generate Fuzzy association rule for classification

A search tree is employed to list all possible frequent fuzzy item sets and to generate fuzzy association rules for classification, it limiting the depth of the branches in order to find a small number of short (i.e., simple) fuzzy rules. The root (initial) or level 0 of a search tree is an empty set. Then all attributes are assumed to have an order (the order of appearance in the training data), and the one-item sets corresponding to the attributes are listed in the first level of the search tree according to their order. If an attribute has j possible outcomes (q_j linguistic terms for each quantitative attribute), it will have j one-item sets listed in the first level. If an attribute has j > 2 possible outcomes, that can be replaced by j binary variables to ensure that no more than one of these j binary attributes can appear in the same node in a search tree. The depth of the trees is limited to a fixed value ($Depth_{max}$). To generate the RB and employ a search tree to list all the possible fuzzy item sets of a class. The root or level 0 of a search tree is an empty set. Then all attributes are assumed to have an order (the order of appearance in the training data), and the one-item sets corresponding to the attributes are listed in the first level of the search tree according to their order.

If an attribute has j possible outcomes (q_j linguistic terms for each quantitative attribute), it will have j one-item sets listed in the first level of the search tree. The children of a one-item node in tree for an attribute A are the two-item sets that include the one item set of attribute A and a one-item set for another attribute behind attribute A in the order, and so on. If an attribute has j > 2 possible outcomes, that can be replaced by j binary variables to ensure that no more than one of these j binary attributes can appear in the same node in a search tree. An example with two attributes (V1 and V2) with two linguistic terms (L and H) . An item set with a support higher than the minimum support is a frequent item set. If the support of an n th item set in a node J is less than the minimum support, it does not need to be extended more because the support of any item set in a node in the sub tree led by node J will also be less than the min support. The same as if a candidate item set generates a classification rule with confidence higher than the maximum confidence, this rule has reached the quality level demanded by the user and it is again unnecessary to extend it further. These properties can greatly reduce the number of nodes needed for searching.

4.2 Candidate rule Filtering

Even though the order of the associations is limited in the association rule extraction, the number of rules generated can be very large. In order to decrease the computational cost of the genetic post-processing stage consider the use of subgroup discovery based on an improved Weighted Relative Accuracy measure (wWRAcc’) to preselect the most interesting rules by means of a pattern weighting scheme. This scheme treats the patterns in such a way that covered positive patterns are not deleted when the current best rule is selected. In the first iteration all target class

patterns are assigned the same weight w (e_j ; 0) = 1 and then it will be updated using the wWRAcc' was used to evaluate the quality of interval rules in APRIORI-SD. This measure was defined as follows:

$$wWRAcc'(A \rightarrow Cj)\left(\frac{n'(A)}{N'}\right).\left(\left(\frac{n'(A.cj)}{n'(A)}\right) - \left(\frac{n(Cj)}{N}\right)\right)$$

Where N_0 is the sum of the weights of all patterns, N_0 (A) is the sum of the weights of all covered patterns, N_0 (A$\rightarrow$ C_j) is the sum of the weights of all correctly covered patterns, n(C_j) the number of patterns of class C_j, and N is the number of all patterns.

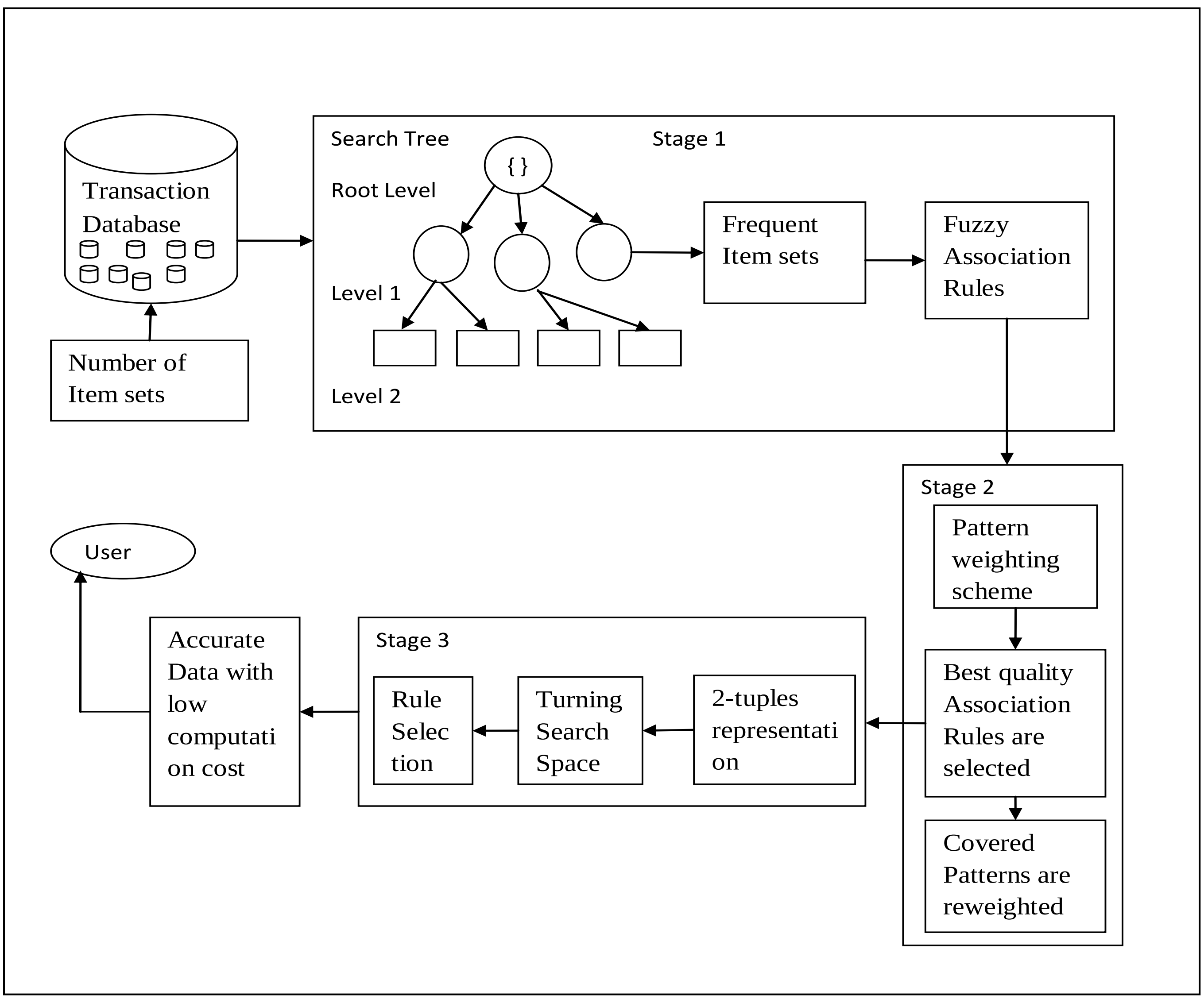

Figure 4.1 System Architecture

4.3 Genetic rule selection using GA

Finally, to make use of GAs to select and tune a compact set of fuzzy association rules with high classification accuracy in order to consider the known positive synergy that both techniques present. Many works have successfully combined the selection of rules with the tuning of membership functions (MFs) within the same process. Considering rules are based on the linguistic 2-tuples representation. This representation allows the lateral displacement of the labels considering only one parameter (symbolic translation parameter), which involves a simplification of the tuning search space that eases the derivation of optimal models, particularly when it is combined with a rule selection within the same process enabling it to take advantage of the positive synergy that both techniques present. The symbolic translation parameter of a linguistic term is a number within the interval [-0.5, 0.5) that expresses the domain of a label when it is moving between its two lateral labels. Let us consider a set of labels S representing a fuzzy partition. Formally, the pair, (Si, _i), Si 2 S, _i 2 [-0.5, 0.5).

5. Conclusion

To obtain an accurate and compact fuzzy rule-based classifier with a low computational cost using fuzzy based association classifier. Using pattern weighting scheme in order to reduce the number of candidate rules with the best quality and to reduce the search space. To mine fuzzy association rules limiting the order of the association in order to obtain a reduced set of candidate rules with less attributes in the antecedent. Transaction Database set up with Fuzzy Items, Fuzzy Association Rule Extraction for Classification methods are solved in this project phase 1. The remaining methods (Candidate rule Prescreening and Genetic rule Selection and Lateral tuning) will be solved in phase 2.A genetic rule selection and lateral tuning have been applied to select a small set of fuzzy association rules with a high classification accuracy.

References

[1]. B. Liu, W. Hsu, and Y. Ma,"Integrating classification and association rule mining", in Proceedings of the International

Conference on Knowledge Discovery and Data Mining (SIGKDD), New York, USA, 1998, pp. 80–86.

[2] J. Li, G. Dong, K. Ramamohanarao, and L. Wong, Elicitation of classification rules by fuzzy data mining", Machine Learning, vol. 54, no. 2, pp. 99– 124, 2004.

[3] Y. Yi and E. Hullermeier, "Learning complexity-bounded rule-based classifiers by combining association analysis and genetic algorithms", in Proc.4th Conf. Eur. Soc. Fuzzy Logic Technol., Barcelona, Spain, 2005, pp. 47–52.

[4].S. Ho, H. Chen, S. Ho, and T. Chen,"Design of accurate classifiers with a compact fuzzy-rule base using an evolutionary scatter partition of feature space", IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 34, no.2,pp. 1031–1044, Apr. 2004.

[5] K. Nozaki, H. Ishibuchi, and H. Tanaka, "Adaptive fuzzy rule-based classification systems", IEEE Trans. Fuzzy Syst., vol. 4, no. 3, pp. 238–250,Aug. 1996.