A Novel Modified Adaptive Fuzzy Inference Engine and Its Application to Pattern Classification

J. Hossen, A. Rahman, K. Samsudin, F. Rokhani, S. Sayeed, R. Hasan

Abstract—The Neuro-Fuzzy hybridization scheme has become of research interest in pattern classification over the past decade. The present paper proposes a novel Modified Adaptive Fuzzy Inference Engine (MAFIE) for pattern classification. A modified Apriori algorithm technique is utilized to reduce a minimal set of decision rules based on input output data sets. A TSK type fuzzy inference system is constructed by the automatic generation of membership functions and rules by the fuzzy c-means clustering and Apriori algorithm technique, respectively. The generated adaptive fuzzy inference engine is adjusted by the least-squares fit and a conjugate gradient descent algorithm towards better performance with a minimal set of rules. The proposed MAFIE is able to reduce the number of rules which increases exponentially when more input variables are involved. The performance of the proposed MAFIE is compared with other existing applications of pattern classification schemes using Fisher's Iris and Wisconsin breast cancer data sets and shown to be very competitive.

Keywords—Apriori algorithm, Fuzzy C-means, MAFIE, TSK

I. INTRODUCTION

IN past decades, fuzzy systems have been combined with neural networks mainly for performing pattern classifications [1]. Many approaches have been proposed to address the issue of automatic generation of fuzzy membership functions and a fuzzy rule base from an input-output data set and also subsequent adjustment of them towards more satisfactory performance [2], [3]. Most of these schemes that incorporate the learning property of neural networks within a fuzzy system framework provide encouraging results. However, most of these techniques also have difficulties associated with the number of resulting fuzzy rules, which increase exponentially when high numbers of input attributes are employed. The computational load required to search for a corresponding rule becomes very heavy as the number of fuzzy rules in a complicated situation is increased.

Institute of Advanced Technology (ITMA), Universiti Putra Malaysia (UPM), Malaysia (*jakirhossen@yahoo.com*, 006-062523382)
Faculty of Engineering, Universiti Putra Malaysia (UPM), Malaysia (arr@eng.upm.edu.my)
Faculty of Engineering, Universiti Putra Malaysia (UPM), Malaysia
Faculty of Engineering, Universiti Putra Malaysia (UPM), Malaysia
Faculty of Information Science and Technology (FIST), Multimedia University (MMU), Malaysia
Institute of Advanced Technology (ITMA), Universiti Putra Malaysia (UPM), Malaysia

Apriori algorithm (a shortened form of *a priori* algorithm) from data mining field has been used with fuzzy inference system to obtain more compact information from a data set [4]. This is a popular algorithm used in data mining using associative rules [5]. Apriori algorithm techniques provides a methodology to do this in data analysis based on empirical data and it has been applied to a variety of areas including web text mining, data mining, medical data analysis, and so on [6].

It is known that the Apriori algorithm approach [5] is able to find a minimal set of decision rules that map input-output (I/O) variables. The TSK type of fuzzy model has an ability to exactly approximate non-linear systems with a combination of linear systems [6]. Consequently, if a minimal set of rules obtained by the modified apriori approach is able to be used to carry out the TSK type fuzzy inference, not only the number of fuzzy inference rules but also the number of fuzzy antecedent variables involved can be effectively reduced. The advantages of both the modified apriori approach and the TSK fuzzy model are combined in order to introduce a novel modified adaptive fuzzy pattern classifier. After this initial construction of the adaptive fuzzy inference model, the membership functions (MFs) are adjusted to achieve better performance.

This paper is organized as follows: Section 2 provides a brief review of Apriori algoritm approach, and the TSK type fuzzy inference model in section 3. Section 4 presents the design of the proposed modified adaptive fuzzy inference engine along with other subsections. Experimental results based on Fisher's Iris and Wisconsin breast cancer data sets for the MAFIE are compared with results for other existing pattern classification algorithms in Section 5, and final conclusions are drawn in Section 6.

II. APRIORI ALGORITHM

The Apriori algorithm was originally proposed in [5] to study the "shopping basket" problem. The "shopping basket" problem may be stated briefly as follows: if a customer is purchasing a certain group of items, what is the likelihood of the person buying another group of items in the same shopping session. The set of items that a customer purchases is called an itemset. Assume for convenience that each customer's transaction has items. The algorithm finds all itemsets L_k greater than a threshold that each transaction has. The L_k is then used to generate the candidate set C_k. The candidate set is the union of $L_k \cup L_{k-1}$. The candidate set is used to form another new larger itemset by removing all those itemsets which are below the threshold. The algorithm repeats

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:5, No:8, 2011

until L_k is empty [41]. We will illustrate this algorithm using a simple example. The procedure is shown graphically in Fig. 1.

The input data to the Apriori algorithm consists of a set of transaction records. The TID (Transaction Identifier) column is the transaction ID and the 'items' is the item number involved in each transaction. For example, in the transaction ID T100, the items purchased are 1, 2, and 4. In this example, there are six transactions. We assume that a threshold of 1 has been set. Thus, we look at the TID column, and find out if the occurrence of items 1, 2, 3, or 4 is less than or equal to one. In this example, all items occurred at least twice. Hence we cannot eliminate any item. In this case, we will form an itemset denoted as Table L_1 in Fig. 1. Since all items are present (as none of them is below the threshold), and hence we have four itemsets in this set L_1. Since there is only one candidate item in each itemset, hence the count column in this table essentially counts the occurrence of each item. From the TID table we find that there are four occurrences of the item 1 (in T100, T103, T104, and T105), and hence the entry in Table for itemset set $\{1\}$ is 4. The Table C_2 is obtained by joining the itemsets in Table L_1
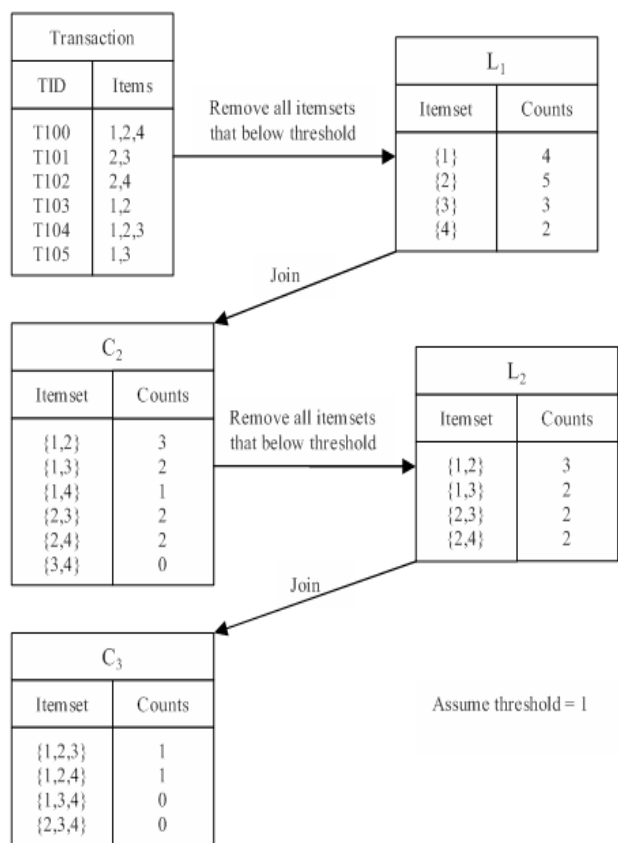

Fig. 1 An example illustrating the determination of the maximum itemset in the apriori algorithm

together. Here the joining is performed in a lexicographical order, with non-repeats. Thus, for example, from itemset $\{1\}$ in Table L_1, we can form the following itemsets $\{1, 2\}$, $\{1, 3\}$, $\{1, 4\}$. Now we can compare the pattern of the itemset $\{1, 2\}$ with the TID column and find out the number of occurrence of this pattern. In this case we find that there are three occurrences (T100, T103, and T104). Hence the entry in the column Count in Table C_2 is 3. We can remove all itemsets in Table C_2 which are below the threshold. In this case, we have itemsets $\{1, 4\}$ and $\{3, 4\}$ which are below the threshold, and hence they will not be considered further. Table L_2 is formed by removing all these itemsets which are below the threshold with the corresponding count of occurrences. Table 3 can be formed by joining (concatenating) the itemsets in Table L_2 together with those in Table L_1. Thus the first candidate itemset would be $\{1, 2\}$ from Table L_2 with itemset $\{1\}$. However, this cannot happen as item 1 already exists in the itemset $\{1, 2\}$. Hence the only possibility would be $\{1, 2\}$ in Table L_2 concatenate with itemset 3. This results in candidate itemset $\{1, 2, 3\}$. Then we compare this pattern with those in Table TID, and find that there is only one such occurrence (T104). It is found that from Table C_3 that all occurrences are less than or equal to the threshold. Hence the process stops.

In this "shopping basket" example, we find that not every item combination exists in the transaction record. The Apriori algorithm removes such a combination if it does not exist or if it is below a prescribed threshold. The procedure may be extended to provide information on the support and confidence of a particular rule found.

III. TSK FUZZY INFERENCE SYSTEM

The TSK type fuzzy model suggested by Takagi Sugeno Kang [7] is able to represent a general class of nonlinear systems. It can be modeled as a linear combination of input variables plus a constant term as defined by (1),

R_i : IF x_{k1} is F_{i1} AND x_{k2} is F_{i2} … AND x_{km} is F_{im} THEN $y_i = c_{i0} + c_{i1}x_{k1} + … + c_{im}x_{km}$

$$y = \frac{\sum_{i=1}^{N} w_i y_i}{\sum_{i=1}^{N} w_i}, \quad w_i = \prod_{j=1}^{m} F_{ij}(x_{kj}) \quad (1)$$

where R_i (i=1, 2, …, N) is the i-th T-S type fuzzy rule, x_{kj} (j=1, 2, …, m) is the j-th input feature of the k-th pattern vector, and F_{ij} is a fuzzy variable of the j-th input feature in the i-th rule. Also Π is a fuzzy T-norm operator and w_i is a rule firing strength of the i-th rule, and y_i is the i-th rule output and y is the total output.

Without a loss of generality, a fuzzy inference system with Multi-Input-Single-Output (MISO) is assumed since it is known that Multi-Input-Multi-Output (MIMO) system can be decomposed into a several number of MISO systems [8]. The Takagi and Sugeno fuzzy model approximates a nonlinear system with a combination of several linear systems by decomposing the entire input domain into several partial spaces and representing each input/output (I/O) space with a linear function. In order to find the coefficients of the linear systems, the least-square fit method has been widely used. It is crucial to fully examine the minimal set of rules in the process of a fuzzy rule generation. If the minimal set of decision rules obtained from Apriori algorithm is appropriate to be used as a set of fuzzy inference rules in the TSK model, the numbers of fuzzy antecedent variables and fuzzy rules in a knowledge-base are able to be reduced effectively.

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:5, No:8, 2011

IV. DESIGN OF A MODIFIED ADAPTIVE FUZZY INFERENCE ENGINE (MAFIE)

A. Automatic Generation of MFs

In order to build a TSK type adaptive fuzzy inference system, firstly an automatic generation of fuzzy membership functions is required. The Fuzzy C-Means (FCM) clustering algorithm [9] is used to find each cluster adaptively. The FCM clustering algorithm is an unsupervised clustering method whose aim is to establish a fuzzy partition of a set of pattern vectors in C number of clusters and the corresponding set of cluster prototypes towards the local minimum of their objective function [10]. An objective function J_m defined by (2) measures the fitting between the clusters and their cluster prototypes.

$$J_m(M,v) = \sum_{k=1}^{n} \sum_{i=1}^{C} (u_{ik})^m (d_{ik})^2 \qquad (2)$$

where $u_{ik} \, \varepsilon \, [0 \ 1]$ is a membership degree of the k-th pattern vector to the i-th cluster represented by its cluster prototype v_i, and v is a cluster prototype of u_i. The distance measure d_{ik} used in the FCM clustering is the *Euclidean* norm $d_{ik} = \|x_k - v_i\|$ on R^p if a pattern vector is in a p-dimensional space, and m is a weighting exponent so-called fuzzifier, $m \, \varepsilon \, [1,\infty]$, which then makes the resulting partitions more or less fuzzy.

After the FCM clustering, each membership function of the j-th feature, x_j, is obtained by plotting the elements of each row of the membership matrix M versus x_j values. Two procedures are applied for each membership function to form their shapes and to fit their membership values.

1) Finding outer shapes: Amongst all data points of each membership function after plotting the entries as above, select only the maximum membership degree for each value of the jth feature, x_j. These maximum membership degrees will be used in the fitting process to generate prototypes of their corresponding fuzzy membership functions as follows.

2) Fitting without false representation: Since the FCM clustering algorithm applies normalization as in (3),

$$\sum_{i=1}^{C} u_{ik} = 1.0 \qquad (3)$$

This condition causes a pattern vector to have a very small amount of representation within a membership function where it should have no membership values in the ideal case. In other words, the FCM algorithm assigns a small noise as a same membership value $1/C$ to each cluster. To overcome this handicap due to the false representation, a modified α–cut method [18] is utilized to remove the noise. Then, to fit those processed membership values for each fuzzy set, a modified asymmetric Gaussian membership function (*gauss2mf*) as defined by (4) is chosen for the adaptive membership function scheme that provides more flexibility.

$$\mu_{ij} = \mu_{ij1} \times \mu_{ij2}$$

$$\mu_{ij1} = e^{-\left(\frac{x_{kj} - v_{ij1}}{\sigma_{ij1}}\right)^2} \times Index_{v_{ij1}} + (1 - Index_{v_{ij1}}) \qquad (4)$$

$$\mu_{ij2} = e^{-\left(\frac{x_{kj} - v_{ij2}}{\sigma_{ij2}}\right)^2} \times Index_{v_{ij2}} + (1 - Index_{v_{ij2}})$$

$$if \ x_{jk} \leq v_{ij1}, Index_{vij1} = 1, \text{otherwise } 0$$

$$if \ x_{jk} \leq v_{ij2}, Index_{vij2} = 1, \text{ otherwise } 0$$

The membership value μ_{ij} is determined by the j-th feature value of the k-th pattern vector, x_{kj}, the cluster prototype value for the j-th feature of the i-th cluster, v_{ij}, and two different standard deviations, σ_1 and σ_2. The Levenberg-Marquardt type non-linear least square fit is utilized to estimate the parameters, $\{v_{ij1} \ v_{ij2} \ \sigma_{ij1} \ \sigma_{ij2}\}$ for each membership function for each fuzzy cluster. The initial values of cluster prototypes v_{ij} are obtained from the final cluster prototypes using the FCM, and deviations σ_{ij1}, σ_{ij2} are initialized to the average deviation of pattern vectors in each cluster. The height of this modified asymmetric Gaussian membership function initialized as 1.0, but it is able to be controlled to be less than 1.0 during the fitting process when $v_{ij1} > v_{ij2}$. This characteristic regarding the height of the membership function provides the proposed adaptive fuzzy inference system with more flexibility to model the best shapes of the training data using Gaussian basis functions.

B. A Modified Apriori Algorithm for Rule Formation

In this section, we will give bit details of a proposed algorithm for rule formation. This algorithm is inspired by the ways of finding the maximum itemset in the Apriori algorithm. However, our proposed algorithm is different from the one used in the Apriori algorithm. In a way the proposed algorithm is like running the maximum itemset determination algorithm backwards. Instead of considering each item by itself, we will start with the clusters identified in the fuzzy c-means clustering method

We will consider a simple example to illustrate the proposed rule formation method. We will first consider the d=1 axis. There are two clusters: {1, 2, 3} and {4, 5, 6} respectively. This is shown in the table called Clustered Data in Fig. 2. For convenience we will label {1, 2, 3} as cluster 1, and {4, 5, 6} as cluster 2. Each cluster consists of three data labels. This information is displayed in the table called L_1 in Fig. 2. The threshold is 1. The column itemset has two elements each denotes the cluster {1, 2, 3} and cluster {4, 5, 6} respectively. The column "Combination of clusters" denotes the label we provided these two clusters, i.e., cluster 1, and cluster 2 respectively. The column "Count of elements" denotes the number of elements in the cluster. In both cases, there are three elements in the cluster. Then we concatenate the clusters in dimension d= 1 with those in dimension d = 2 using a join operation. As there are two clusters in the d=2 dimension: {1, 2} and {3, 4, 5, 6}, we can concatenate the clusters of the dimension d= 1 with those in dimension d = 2 and find the common elements. Thus, in the table called C_2 in Fig. 2, the first element in the column "itemset" shows the concatenation of the first cluster {1, 2, 3} in dimension d= 1 with the first

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:5, No:8, 2011

cluster {1, 2} in dimension d=2. This is denoted by {1, 2, 3}∩{1,2}. This is denoted by (1), 1 in column

Clustered Data

{1,2,3}, {4,5,6}
{1,2,3}, {3,4,5,6}
{1,2,3,4}, {5,6}

L₁

Itemset	Combination of clusters	Counts of common element
{1,2,3}	1	3
{4,5,6}	2	3

C₂

Itemset	Combination of clusters	Counts of common element
{{1,2,3}∩{1,2}}	(1),1	2
{{1,2,3}∩{3,4,5,6}}	(1),2	1
{{4,5,6}∩{1,2}}	(2),1	0
{{4,5,6}∩{3,4,5,6}}	(2),1	3

L₂

Itemset	Combination of clusters	Counts of common element
{1,2}	(1),1	2
{4,5,6}	(2),2	3

C3

Itemset	Combination of clusters	Counts of common element
{{1,2}∩{1,2,3,4}}	(1,1),1	2
{{1,2}∩{5,6}}	(1,1),2	0
{{4,5,6}∩{1,2,3,4}}	(2,2),1	1
{{4,5,6}∩{5,6}}	(2,2),2	2

L₃

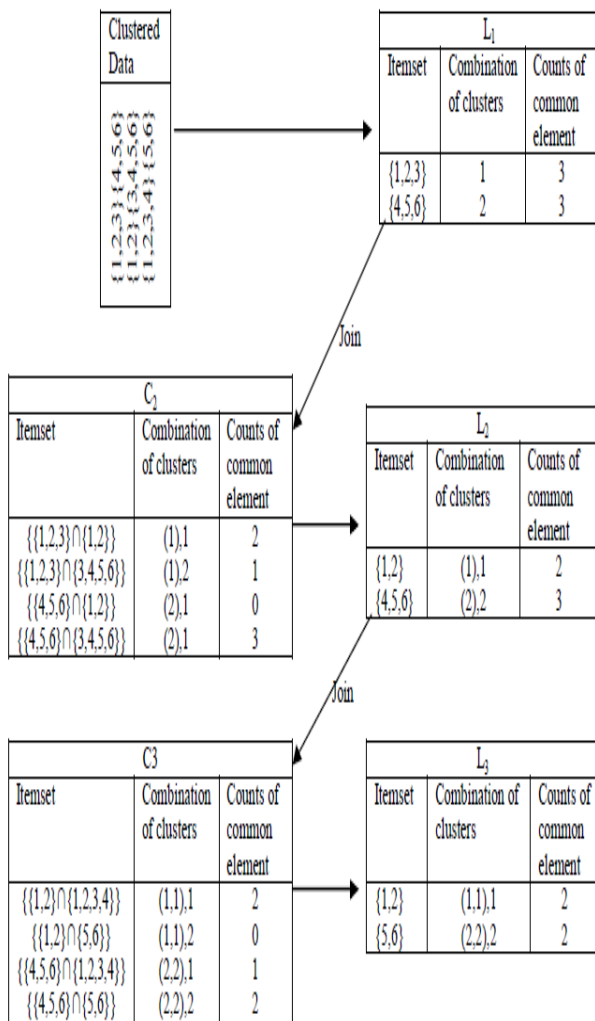| Itemset	Combination of clusters	Counts of common element
{1,2}	(1,1),1	2
{5,6}	(2,2),2	2

Fig. 2 An example to illustrate the proposed rule formation algorithm

"Combination of clusters" in Fig. 2. In this case, we find that there are two common elements 1, and 2 and hence the entry of 2 in the column "Counts of common elements". In a similar manner, we can find the values on all the columns of the table C_2. Since the threshold is 1, we can remove the entry corresponding to {4, 5, 6,}∩{1, 2}, as there are no common elements in this concatenation. Then we transfer this information into the table L_2. The entries in the column "Itemset" are the concatenation of the clusters and the common elements. Thus for example, the first entry of column "itemset" is obtained by the concatenation of cluster 1 {1, 2, 3} in d= 1 dimension and cluster 1 {1, 2} in dimension = 2. The common elements in these two clusters are {1, 2}. This will be the itemset. There are only two common elements, and hence the entry in the column "Counts of common elements" is 2. The entry in the column "Combination of clusters" denoted by (1), 1 signifies the result is obtained by the concatenation of cluster 1 in d= 1 dimension with that of cluster 1 in the d= 2 dimension. The entries in the table called

C_3 denote the join operation of the results of Table L_2 with those clusters on the d= 3 dimension. Thus the first element is formed by concatenation of the common elements found by concatenating cluster 1 in dimension d = 1 and cluster 1 in dimension d = 2 with cluster 1 in the d = 3 dimension. This is denoted by {1, 2} ∩ {1, 2, 3, 4}. Thus the meaning of the first entry in the column "Combination of clusters" (1,1),1. Here we find that there are two common elements, viz. {1 2}. Hence the first entry in the column "Counts of common elements" is 2. This process is repeated for the other clusters, and Table C_3 is fully populated. As the threshold is 1, and hence we can eliminate entries {1,2}∩{5, 6}, and {4, 5, 6}∩{1,2,3,4}. The remaining information is transferred to the table called L_3. Here there are only two values which are above the threshold {1,2}∩{1, 2, 3, 4}, and {4, 5, 6}∩{5, 6}. Hence the entries in the final column of Table L_3 are both 2 denoting that there are only two common elements. The entries in the column "Combination of clusters" denote the way in which the clusters are formed. For example, the first element is formed by the concatenation of cluster 1 in d = 1 dimension, cluster 1 in d= 2 dimension and cluster 1 in d= 3 dimension. The "Itemset" column denotes the common elements as a result of the concatenation process. Since there are only three input dimensions, and hence the process stops.

In this example, we finally conclude that there are two fuzzy rules (as in Table in Fig. 2 there are only two remaining entries):

Rule1: IF cluster1 in d= 1 dimension ^ cluster1 in d= 2 dimension ^ cluster1 in
d = 3 dimension THEN Consequence1.
Rule2: IF cluster2 in d= 1 dimension ^ cluster2 in d= 2 dimension ^ cluster2 in
d = 3 dimension THEN Consequence2.

From this description it can be observed that the proposed procedure is quite different from the maximum itemset determination in the Apriori algorithm. It seeks to find the combination of clusters such that there are common elements in the clusters. Note that these common elements are represented by the data labels store in the clusters. Nevertheless the proposed algorithm is inspired by the maximum itemset determination algorithm in the Apriori algorithm. It is possible similar to the Apriori algorithm [5] to compute the support and the confidence of the rules formed.

C. Construction of MAFIE

Once the parameters of antecedent MFs are found via the FCM clustering and the minimal set of decision rules is obtained through the Apriori algorithm approach, the proposed modified adaptive fuzzy inference engine (MAFIE) can be constructed. The proposed system is built as a MISO TSK type fuzzy model as mentioned in Section 3. All attributes are set as antecedent variables with the corresponding adaptive cluster information after the FCM clustering. A type of *Generalized Modus Ponen* (GMP) compositional rules is used to form fuzzy rules in the knowledge base and the algebraic minimum operator is utilized to calculate fuzzy *T-norm* operation ('*AND*') between the antecedent variables. The coefficients of the consequent variable are fitted into constant

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:5, No:8, 2011

terms after the least squares fitting, since the values of the output classes are usually assigned as discrete integers in a pattern classification scheme.

D. Tuning Process of MFs

The performance of the system needs to be evaluated and enhanced towards a higher accuracy after the construction stage. If the RMSE error measure in (5) is not satisfactory when compared to an arbitrary error criterion, the parameters of antecedent membership functions are adjusted using the Polak-Ribiere conjugate gradient algorithm based on the difference between the desired and the actual output.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(err_i)^2}{n}}, err_i = y^d - y^o \qquad (5)$$

where, err_i is the error between the desired output, y^d, and the actual output, y^o, from the fuzzy inference system at one epoch. Once the coefficients of the TSK type consequent variable are fitted with the training data, the performance evaluation is done first with the training data to compare its RMSE with a user-defined error criterion. If the RMSE is not satisfactory, the adjustment of antecedent membership functions is carried out with the training data set.

V. EXPERIMENTAL RESULTS

In the past, many different approaches have been suggested to achieve a higher accuracy on a variety of data sets in the pattern classification scheme. For example, as reported in [18], conventional methods [12], [13] and fuzzy-based classifiers; Adaptive Fuzzy Leader Clustering (AFLC) [14], Wu and Chen's algorithm [15], Fuzzy Entropy-Based Fuzzy Classifier (FEBFC) [18], Influential Rule Search Scheme (IRSS) [19] and Adaptive Rough-Fuzzy Inference System (ARFIS) [20,21} have been applied on Iris data set and Wisconsin breast cancer data set to achieve better performance. However, some of these approaches still have difficulties with the number of fuzzy rules when a higher dimensional data set is applied, because in fuzzy inference systems the size of their knowledge base is directly associated with the computational complexity and the system performance.The proposed MAFIE has been developed to resolve this problem by reducing the number of fuzzy rules and antecedent variables effectively through the knowledge-reduction process and by adjusting the antecedent MFs after the performance evaluation.The Fisher's Iris data set [11] and the Wisconsin breast cancer data set [11] were retrieved from the UCI machine learning repository to use for the experiments. The MAFIE was applied using these two data sets to compare its results with other existing pattern classifiers. For each data set, the FCM clustering was done with C=5 and was chosen as 0.02 for applying the modified α-cut method [18] in the process of an automatic generation of membership functions. The error criterion was assigned to 0.2 in the adjustment of antecedent MFs and the experiments were carried out on 10 independent runs for both data sets.

A. Fisher's Iris Data

The Fisher's Iris data set [11] contains 150 pattern vectors with four features (sepal length, sepal width, petal length, and petal width) and one output of three classes (Setosa,

Versicolor, and Virginica). Without a loss of generality, MAFIE selected the first 70 percent of the data for the training data set and the last 30 percent data for the testing data set for each output class. The training and the testing data were swapped once the system was implemented and tested with those data sets, respectively. The results shown in Table 1 represent the average percent accuracy with this N-fold cross validation technique when N=2. The final adjusted antecedent membership functions for sepal length, for instance, are shown in Fig. 3.
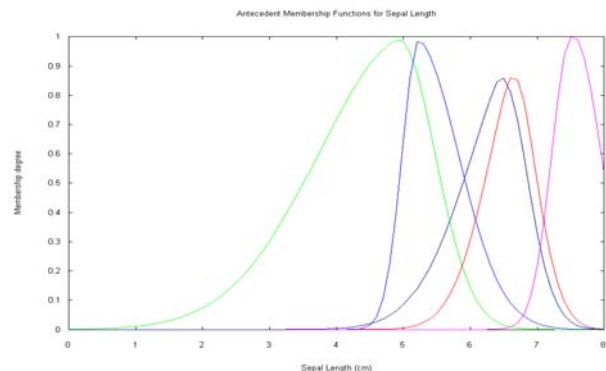

Fig. 3 The final membership functions adjusted for 'Sepal Length'

As a result of the reduction algorithm, the input attributes were reduced to {sepal length, petal length, petal width}. Also in contrast to the size of the rule base of IRSS [19], which increased exponentially as 5^4, the number of fuzzy rules generated by MAFIE was on average 16 after 10 independent runs.In Table I, the accuracy of the proposed MAFIE system was compared with other existing pattern classification schemes on the Iris data set. From the results shown in Table 1, it can be said that the proposed MAFIE is comparable with other algorithms and can be considered to be one of the most efficient fuzzy pattern classifiers for this data set.

TABLE I
COMPARISON WITH OTHER CLASSIFICATION SCHEMES ON IRIS DATA

Algorithms	Setosa (%)	Versicolor (%)	Verginica (%)	Average classification ratio (%
GVS [13]	100	94.00	94.00	96.00
Fisher ratios [12]				96.00
AFLC [14]	100	86.00	100	95.33
Wu and Chen [15]	100	93.38	95.24	96.21
FEBFC [18]				97.12
IRSS [19]	100	92.00	96.00	96.00
ARFIS [20]	100	93.60	95.24	96.28
MAFIE	100	95.41	97.32.	97.57

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:5, No:8, 2011

B. Wisconsin Breast Cancer Data

The proposed MAFIE was also applied using the Wisconsin breast cancer data set [11] to determine whether any classification approach is efficient enough to handle such a high dimensional data. This data set has 699 samples with nine input attributes and one output to classify the sample as a "benign" or a "malignant" sample. In order to create the training and the testing data sets, the steps described below were performed.
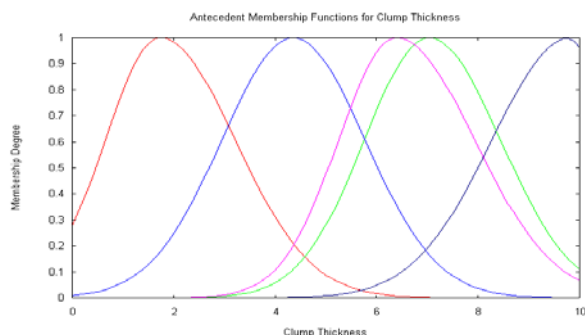

Fig. 4 The final membership functions adjusted for "Clump Thickness"

Amongst all 699 pattern vectors from the original data set, samples that include missing attributes were firstly removed. Then, in a random manner, 70 percent of instances for each class were assigned as the training data set and another 30 percent was selected for testing. In Fig. 4 displays the final membership functions adjusted for the feature "Clump Thickness".The number of input attributes was reduced to 5 and the number of fuzzy rules obtained as a minimal set of rules was on average 12 after 10 runs. As shown in Table 2, our proposed algorithm MAFIE provides encouraging results for classification on the Wisconsin breast cancer data set when compared with other classification methods. Note that our proposed system had a much better performance despite the higher dimensionality of input attributes. This was achieved by the effective reduction process of the apriori algorithm methodology and by the adjustment procedure.

TABLE II
COMPARISON WITH OTHER CLASSIFICATION SCHEMES ON CANCER DATA

Algorithms	Testing Accuracy (%)
Setiono's neuro classifier [16]	93.99
MSC [17]	94.90
FEBFC [18]	95.14
IRSS [19]	95.89
ARFIS [20]	96.63
MAFIE	97.24

VI. CONCLUSION

A novel modified adaptive fuzzy inference system (MAFIE) has been proposed which automatically generates fuzzy membership functions via the FCM clustering and fuzzy rules from the modified Apriori algorithm based on input-output data sets. The performance evaluation was done to achieve better performance through the adjustment of antecedent membership functions. It is significant that the number of rules generated by MAFIE were reduced effectively by the Apriori approach towards better performance. The comparisons with other pattern classifiers indicated that the performances of MAFIE were found to be encouraging and satisfactory. Research is continuing on the refinement process of the fuzzy rules to achieve better accuracy in a pattern classification scheme.

REFERENCES

[1] P.K. Simpson, "Fuzzy Min-Max Neural Networks–Part 1:Classification," IEEE Transaction on Neural Networks, vol. 3, no.5, pp.776-786, Sept. 1992.

[2] S. Abe and M.S. Lan, "Fuzzy Rules Extraction Directly from Numerical Data for Function Approximation," IEEE Transaction on System, Man, and Cybernetics, vol. 25, no.1, pp.119-129, Jan. 1995.

[3] G.O.A. Zapata, R.K.H. Galvao, and T. Yoneyama, "Extracting Fuzzy Control Rules from Experimental Human Operator Data," IEEE Transaction on System, Man and Cybernetics – Part B: Cybernetics, vo. 29, no. 3, pp 25-40, Feb. 1999.

[4] Han, J. and Kamber, M., Data Mining: Concepts and Techniques, Second Edition, Morgan Kaufmann publishers, San Francisco, 2006

[5] R. Agrawal, R. Srikant, "Fast Algorithms for Mining Association Rules", Proceedings of the 20th VLDB Conference, Santiago, Chile, 1994.

[6] Bilal I. S., Keshav P. D., Alamgir M. H., Mohammad S A., " Diversification of Fuzzy Association Rules to Improve Prediction Accuracy", Fuzzy Systems (Fuzz) in IEEE Explorer, 2010.

[7] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," IEEE Transaction on Systems, Man, and Cybernetics, vol. SMC-15, pp. 116-132, Jan.-Feb. 1985

[8] X. Zeng and M. G. Singh,, "Approximation Theory of Fuzzy Systems-MIMO Case IEEE Transactions on Fuzzy Systems, vol. 3, no. 2, May 1995.

[9] James C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, pp.65-86, 1981.

[10] Mohanad A., Mohammad M., Abdullah R., " Optimizing of Fuzzy C-Means Clustering Algorithm Using GA", World Academy of Science, Engineering and Technology, Vol. 39, 2008.

[11] C.L. Blake and C.J. Merz, "UCI Repository of Machine Learning Databases," University of California, Irvine, Department of Information and Computer Science, http://www.ics.uci.edu/~mlearn /MLRepository.html, 1998

[12] S. Fahlman and C. Lebiere, "The Cascade-Correlation Learning Architecture," Carnegie Melloin Univ., School of Computer Science, Technical Report CMU-CS- 90-100, Feb. 1990.

[13] T-P. Hong and S.-S. Tseng, "A Generalised Version Space Learning Algorithm for Noisy and Uncertain Data," IEEE Transaction on Knowledge and Data Eng., vol. 9, no. 2, pp. 336-340, Mar.-Apr. 1997.

[14] S.C. Newton, S. Pemmaraju, and S. Mitra, "Adaptive Fuzzy Leader Clustering of Complex Data Sets in Pattern Recognition," IEEE Transaction on Neural Networks, vol. 3, no.5, pp.794-800, Sept. 1992.

[15] T.P. Wu and S.M. Chen, "A New Method for Constructing Membership Functions and Fuzzy Rules from Training Examples," IEEE Transaction on System, Man, and Cybernetics – Part B: Cybernetics, vol. 29, no.1, pp.25-40, Feb. 1999.

[16] R. Setiono, "Extracting M-of-N Rules from Trained Neural Networks," IEEE Transaction On Neural Networks, vol. 11, no. 2, pp.512- 519, Mar. 2000.

[17] B.C. Lovel and A.P. Bradley, "The Multiscale Classifier," IEEE Transaction On Pattern Analysis and Machine Intelligence, vol. 18, no. 2, pp. 124-137, Feb. 1996.

[18] H.-M. Lee, C.-M. Chen, J.-M. Chen, and Y.-L. Jou, "An Efficient Fuzzy Classifier with Feature Selection Based on Fuzzy Entropy," IEEE Transaction on Systems, Man, and Cybernetics – Part B: Cybernetics, vol. 31, no.3, pp.426-432, June 2001.

[19] A. Chatterjee and A. Rakshit, "Influential Rule Search Scheme (IRSS) – A New Fuzzy Pattern Classifier," IEEE Transaction on Knowledge and Data Engineering, vol. 16, no. 8, pp. 881-893Aug. 2004.

[20] ChangSu L., Anthony. Z., Tomas B., "An Adaptive T-Stype Rough-Fuzzy Inference System (ARFIS) for Pattern Classification", Fuzzy Information Society, IEEE Explorer, pp. 117-122, 2007.

[21] Sandeep C., and Rene V. M., "RANFIS: Rough Adaptive Neuro-Fuzzy Inference System", International Journal of Computational Intelligence, vol. 3, No. 4, 2006.