Iranian Journal of Fuzzy Systems Vol. 3, No. 1, (2006) pp. 77-89

NEW CRITERIA FOR RULE SELECTION IN FUZZY LEARNING CLASSIFIER SYSTEMS

M. EFTEKHARI, M. J. ZOLGHADRI AND S. D. KATEBI

ABSTRACT. Designing an effective criterion for selecting the best rule is a major problem in the process of implementing Fuzzy Learning Classifier (FLC) systems. Conventionally confidence and support or combined measures of these are used as criteria for fuzzy rule evaluation. In this paper new entities namely precision and recall from the field of Information Retrieval (IR) systems is adapted as alternative criteria for fuzzy rule evaluation. Several different combinations of precision and recall are redesigned to produce a metric measure. These newly introduced criteria are utilized as a rule selection mechanism in the method of Iterative Rule Learning (IRL) of FLC. In several experiments, three standard datasets are used to compare and contrast the novel IR based criteria with other previously developed measures. Experimental results illustrate the effectiveness of the proposed techniques in terms of classification performance and computational efficiency.

1. Introduction

Recently learning classifier systems have been widely used for automatic learning, especially in the case of learning classification rules. Three different methods of Genetic Based Machine Learning (GBML) have been developed, namely, Pittsburgh, Michigan and Iterative Rule Learning (IRL). These three methods were originally developed for constructing a rule base for crisp expert systems [4-6, 9, 12-15, 18, 24]. Due to the emphasis on the transparency of extracted rules in the field of data mining and also the inherent comprehensibility of fuzzy rules, GA based methods have recently been developed to learn fuzzy rules.

In the Pittsburgh method [9, 18, 24], an entire rule base is coded by a string as a chromosome. Therefore, it is computationally demanding, especially for a high dimensional pattern space [12]. In the Michigan approach [4, 15] a single fuzzy rule is coded by a chromosome. So, as opposed to the Pittsburgh approach, this method is less CPU-time and memory demanding, but may not optimize the rule set directly. Rules in Michigan approach must cooperate to receive payoff while they also compete in the process of running GA [12] while rules sets in Pittsburgh approach only cooperate to receive payoff.

In order to overcome some of difficulties associated with the Michigan and Pittsburgh approaches, an Iterative Rule Learning (IRL) approach was proposed in [5, 6, 8, 12-14]. This method is based on searching for fuzzy rules in a sequence of iterations

Received: September 2005; Accepted: December 2005

Key words and phrases : Fuzzy classification, Rule evaluation criteria, Information retrieval, Iterative rule learning.

and a heuristic criterion is used for evaluating each rule. When a fuzzy rule is found, training patterns covered by this rule are removed or their weights are set to zero. Then another fuzzy rule is found and added to rule set using the modified training set. In the IRL method, GA is used to find a partial solution to the learning problem at every iteration so that the size of the search space at later stages becomes smaller.

One important step in implementing the IRL approach is the selection of the best rule. Most of the proposed criteria [1, 2] for evaluation of the rules are based on confidence and support concepts in association rule mining. The confidence measures the accuracy and support evaluates the coverage of a rule. Ishibuchi has extended fuzzy versions of these concepts [16, 17]. Some different heuristic criteria for fuzzy rules evaluation have been compared in [16]. The authors have reported, "The better results in terms of classification accuracy have been obtained from composite criteria of confidence and support than their individual use" [16].

A search engine or Information Retrieval (IR) system is said to be efficient, to the degree that it evaluates each element in the information base similar to an expert and retrieves the most relevant element. In other words, an IR system can be conceived as a set of rules which must be evaluated in order to achieve the ability to retrieve the most relevant objects. So the problem of classifying patterns with some fuzzy rules is similar to retrieving relevant objects by an IR system. On the other hand there is a conceptual resemblance between the pair (confidence, support) in association rule mining and the pair (precision, recall) in IR systems. The former is used to evaluate association rules and the latter is used for evaluating IR systems. Precision and recall are two primary criteria used to evaluate the performance and efficiency of IR systems [19]. Furthermore, there exist other composite criteria of precision and recall that produce a scalar metric [3, 25].

In this work, inspired by the similarity between an IR system and a fuzzy rule set for classification, composite criteria of precision and recall are considered as evaluation criteria for fuzzy rules in IRL approach. The IR based composite criteria are more comprehensible than confidence and support based methods.

Although the IRL approach is used to handle high dimensional classification problems, the main contribution of this paper is primarily to introduce new measures for rule selection and compare the relative effectiveness of various composite rule evaluation techniques. Three well-known classification problems are considered to illustrate the performance of each of the introduced measure.

2. Fuzzy Rule Based Classification Systems (FRBCS)

In FRBCS, the classifier is represented by a set of fuzzy if-then rules. The following structure is used to represent a fuzzy rule.

$$Rule\ \ R_i\ \ :\ If\ \ x_1\ \ is\ \ A_{i1}\ \ and\ \ ....\ and\ \ x_n\ \ is\ \ A_{in}\ \ then\ \ Class\ \ C_i \tag{1}$$

where n is the number of features used for the classification problem, $X = [x_1, x_2, \ldots x_n]^T$ is the pattern vector, $C_i \in [C_1, C_2, \ldots C_M]$ is the consequence class of the rule and A_{ik} is the fuzzy set associated with x_k.

Assuming that a set of m labeled training patterns $\left(X^{(1)}, X^{(2)}, \ldots X^{(m)}\right)$ is available, the task of classifier design is to find a compact set of fuzzy rules in the form of Equation (1).

In order to classify an input pattern $X = [x_1, x_2, \ldots x_n]^T$, the degree of compatibility of the pattern with each rule is calculated using the product operator to achieve and connectives in the rule antecedent:

$$\mu R_i(X) = \prod_{k=1}^{n} \mu A_{ik}(x_i), \qquad i = 1, \ldots N \tag{2}$$

where N is the number of fuzzy rules. The pattern is classified using the rule having maximum compatibility (i.e. the winner rule R_w) with the input pattern:

$$\mu R_w(X) = \max\left\{\mu R_i(X),\ i = 1, \ldots, N\right\} \tag{3}$$

where w is the $\arg\max_i(\mu R_i(X))$.

3. **Feature Selection**

Fuzzy rule based systems are linguistically interpretable. Hence fuzzy systems are more comprehensible than other systems used in data mining and classification [11]. This inherent transparency can be diminished when the number of features increases. Therefore the comprehensibility of the rule base can be improved by a feature selection method. Recent studies show that, even for a low dimensional data set such as IRIS, discarding some less effective input features can improve the average classification accuracy rate [7]. Some fuzzy similarity-driven methods have been used to simplify the rule set of a fuzzy system and to select a subset of features before the rule set generation process [7,23]. Also less-informative data and noise may cause a classifier to behave erroneously [21, 22, 23].

In this paper, to reduce dimensions by ranking the features and then selecting the best ones, the sum of estimated Mahalanobis distances [10] is considered.

4. **IRL Approach for Fuzzy Rule Learning**

There are some genetic learning processes that use the IRL approach for learning rules. SLAVE [12,13], SIA [26] and Genetic Generation Process [14] are algorithms which are based on IRL. The SLAVE algorithm uses GA and begins to search for a new rule after removing the patterns covered by the previously found rule; it was designed to work with and without linguistic information. SIA is also based on GA and works in a manner similar to SLAVE but it was designed to work with crisp data. The Genetic Generation Process uses GA to search for the best rule similarly to the aforementioned methods but it assigns a weight to each pattern covered by the last rule, eliminating all patterns with weight greater than a preset threshold value.

The common properties in all GA based IRL methods are listed below:

1) Running GA to find a rule.
2) A criterion for selecting the best rule at each iteration.
3) A penalization criterion (weighing or removing)
4) An appropriate criterion to terminate all process.

In the IRL approach the best individual found by GA at each iteration is considered as a solution of the rule set learning problem. Discarding the remaining individuals of the population in GA makes the whole problem of rule set learning more efficient than the Michigan and Pittsburgh approaches. Therefore the IRL approach is used to reduce the search space by giving a partial solution to rule set. As suggested in SLAVE [12,13], the IRL approach is considered as learning the rules of a particular class each time. Therefore, for a problem with *M* classes, the IRL algorithm must be repeated *M* times, each time, extracting a set of rules for a particular class. An iterative process with an embedded GA is used to learn the whole rule set as described below:

For each class *i* in class list $\left[C_1, C_2, \ldots C_M\right]$:

Use GA to obtain a rule for C_i
Incorporate the rule into the final set of rules.
Penalize this rule
If the set of rules is adequate to represent the examples of ith class, terminate the Process of finding rules for that class.

End For.

5. Rule Evaluation Criteria

5.1 Existing Rule Evaluation Techniques.

Confidence and support of association rules have often been used as rule selection criteria in data mining [1]. The fuzzy rule in (1) can be viewed as an association rule of the form $A_i \Rightarrow C_i$ where A_i is a multidimensional fuzzy set representing the antecedent combination of the rule. In [16,17] an extension of these measures for the case of a fuzzy association rule is given which can be expressed as:

$$C = Confidence(R_i) = \sum_{X^{(k)} \in C_h} \mu_i(X^{(k)}) \Big/ \sum_{k=1}^{m} \mu_i(X^{(k)}) \tag{4}$$

$$S = Support(R_i) = \sum_{X^{(k)} \in C_h} \mu_i(X^{(k)}) \; / m \tag{5}$$

Where C_h is the consequence class of rule R_i and *m* is the number of training patterns. Employing composite criteria of confidence and support for rule selection has proven to be more effective than their individual use [16]. For the purpose of comparison, three composite measures appearing in the literature will be considered here. The product of confidence and support is used as a composite measure for rule selection. Two other measures proposed in past research are Castro [6] and SLAVE [12,13]. These measures can be formulated using confidence and support [17] as follows:

$$f_{Product} = C.S \tag{6}$$

$$f_{SLAVE}(R_i) = f_{SLAVE}(A_i \Rightarrow C_h) = S(A_i \Rightarrow C_h) - S(A_i \Rightarrow \overline{C_h}) \tag{7}$$

$$f_{Castro}(R_i) = (\mathrm{S}(A_i \Rightarrow \mathrm{C_h}) \;/\; (n_h.\overline{n_h})\;)\;((\overline{n_h}/m) - \mathrm{S}(A_i \Rightarrow \overline{C_h})) \tag{8}$$

Where n_h and $\overline{n_h}$ are the number of training patterns in class C_h and other classes respectively.

5.2 **Proposed Rule Evaluation Technique.**

When a query is submitted to an IR system, a subset of objects (documents) present in corpus is returned to the user. This is usually presented to the user in the form of a ranked list. The object in the top rank is assumed to be the most relevant to the user's query. This can be viewed as a classification problem with two classes (relevant and non-relevant). The ultimate goal of an IR system is to present relevant documents to the user.

Precision, recall and fall-out are three common measures for evaluating an IR system. Precision is the proportion of relevant objects in the retrieved set, recall is the proportion of relevant objects retrieved and fall-out is the proportion of non-relevant patterns retrieved. These measures may be adapted for evaluating fuzzy classification rules:

$$P = precision(R_i) = \sum_{X^{(k)} \in C_h} \mu_i(X^{(k)}) \Big/ \sum_{k=1}^{m} \mu_i(X^{(k)}) \tag{9}$$

$$RE = recall(R_i) = \sum_{X^{(k)} \in C_h} \mu_i(X^{(k)}) \Big/ n_h \tag{10}$$

$$FO = Fallout(R_i) = \sum_{X^{(k)} \notin C_h} \mu_i(X^{(k)}) \Big/ \overline{n_h} \tag{11}$$

P, RE and *FO* denote precision, recall and fallout respectively and n_h is the total number of patterns which belong to the class C_h in the training set. Notice that precision and confidence, respectively Eq. (4) and (9), are equivalent.

As in our canonical IRL scheme we select the best rule for a particular class, using product of confidence and support as a criterion. This is equivalent to using the product of precision and recall.

$$\text{Product} = P.RE \tag{12}$$

Another well-known composite measure in the field of IR is the F-measure. This can be expressed as:

$$F_\beta = ((1+\beta^2)P.RE)\big/(\beta^2 P + RE) \tag{13}$$

where the parameter β can be used as a trade-off between the relative importance of recall and precision. $\beta = 1$ weights *P* and *RE* equally ($F_0 = P$, $F_\infty = RE$).

A composite measure which combines *P, RE* and *FO* is the *φ*-measure which can be expressed as:

$$\varphi = (RE - FO)\sqrt{(1-P)\Big/\Big\{RE.\big(FO + \big((1/P)-1\big)(1-FO)\big)\Big\}} \tag{14}$$

The average of the interpolated precision versus recall is another well-known composite measure in IR. Let $RE_0, RE_1, ..., RE_{10}$ be defined as the standard recall levels such that $RE_j = \frac{j}{10}$ $j = 0,1,...,10$. At first the retrieved patterns are ranked in descending order of their compatibility degree with respect to the rule" $R_i : A_i \Rightarrow C_h$ ". Then the precision and recall values are calculated only for relevant patterns (patterns that belong to class *h*) in the descending sorted list. These calculated precision and recalls are now treated as actual and denoted by $P(re_k)$ and re_k . Then the interpolated precision at standard recall level RE_j , denoted by $P(RE_j)$, is defined by:

$$P(RE_j) = \begin{cases} \max\limits_k P(re_k) & RE_j < re_k < RE_{j+1} \\ 0 & otherwise \end{cases} \tag{15}$$

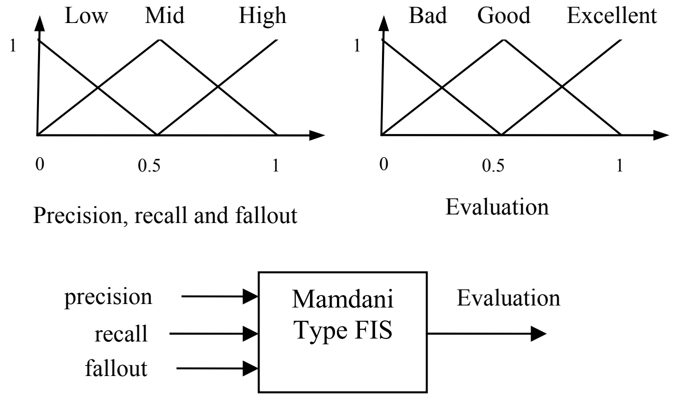
where we go back from RE_{10} to RE_0 for calculating the related interpolated precision.

Finally, a possible evaluation measure is the average of the calculated and interpolated precisions for each standard recall.

$$\text{AIP} = \sum_{j=0}^{Nrs-1} P(RE_j)\Big/Nrs \tag{16}$$

where *Nrs* is the number of considered standard recalls (e.g. the number of standard recalls is eleven in this paper (RE_0 to RE_{10})).

The φ-measure is a composite metric which uses all the three parameters (precision, recall and fallout). In order to prepare an interpretable composite criterion according to the φ-measure, a Fuzzy Inference System (FIS) is obtained by sampling from the φ-measure using the Wang-Mendel method [27]. The fuzzy sets and general schema of obtained Mamdani type FIS are given below:



The schema of the φ-FIS learned by Wang-Mandel method for evaluating a rule

If precision is Low and recall is Low and fallout is Low then Evaluation is Bad
If precision is Low and recall is High and fallout is High then Evaluation is Bad
If precision is Mid and recall is High and fallout is High then Evaluation is Bad
If precision is High and recall is High and fallout is High then Evaluation is Bad
If precision is Low and recall is Mid and fallout is Low then Evaluation is Good
If precision is Low and recall is Mid and fallout is Mid then Evaluation is Good
If precision is Low and recall is High and fallout is Low then Evaluation is Good
If precision is Low and recall is High and fallout is Mid then Evaluation is Good
If precision is Mid and recall is Low and fallout is Low then Evaluation is Good
If precision is Mid and recall is Mid and fallout is Low then Evaluation is Good
If precision is Mid and recall is Mid and fallout is Mid then Evaluation is Good
If precision is Mid and recall is High and fallout is Mid then Evaluation is Good
If precision is Mid and recall is High and fallout is Mid then Evaluation is Good
If precision is High and recall is Low and fallout is Low then Evaluation is Good
If precision is High and recall is Mid and fallout is Mid then Evaluation is Good
If precision is High and recall is High and fallout is Mid then Evaluation is Good
If precision is High and recall is High and fallout is Low then Evaluation is Excellent
If precision is Mid and recall is High and fallout is Low then Evaluation is Excellent
If precision is High and recall is Mid and fallout is Low then Evaluation is Excellent

6. **Experimental Results**

Experiments were carried out with three well-known datasets, namely IRIS, WINE and Breast-cancer-Wisconsin [20]. In each case the sum of the estimated Mahalanobis distance feature selection method was applied to rank features in terms of their importance and effects on classification performance. The higher the rank of a feature the more effective it is for classification. So for each dataset the first five features were selected to perform test and train. In order to evaluate the classification performance two types of experiments were performed.

1. Full-Train vs. Full-Test (FT-FT). The whole dataset was used to perform training and testing.

2. Ten folded Cross Validation (10-CV) technique was used to estimate the classification accuracy as the second experiment. The whole 10-CV process was repeated five times using different data partitions and the average accuracy on test data are given below for each case separately.

The main objective of this paper is to investigate the performance of the proposed IR based criteria for rule selection in terms of the accuracy. In both the experiments mentioned for each case (i.e. the number of selected features) only one rule per class is obtained per iteration of IRL process. Consequently, in the elimination phase of IRL method, only patterns covered by the rule discovered for a particular class are eliminated. In the next iteration, the IRL process goes on to discover the appropriate rule for another class.

The Real Coded GA is used and its parameters for all cases are given in Table 1:

No. of Generations	**Pop. Size**	**Probability of X-over**	**Probability of Mutation**
500	1000	0.6	0.1

TABLE 1. The GA parameters and their values.

The triangular form of fuzzy sets is used with three real numbers for coding each fuzzy set in GA. Thus, for a classification problem with 2 selected features, we have a chromosome length of 6 (2×3), where a triangular fuzzy set is assigned to each selected feature.

6.1 **WINE Data.**

Ranking of features= $\{7, 12, 13, 1, 10, 11, 6, 2, 4, 9, 8, 3, 5\}$

In the following tables the first column indicates the number of selected features from the sorted list.

Number of selected Features	F-measure $\beta=1$	Product	AIP	SLAVE	Castro	φ-measure	φ-FIS
2	0.8370	0.8377	0.9101	0.8089	0.8483	0.8371	0.8371
3	0.9325	0.9325	0.9325	0.9438	0.9382	0.9325	0.8876
4	0.9382	0.9494	0.9157	0.9494	0.9494	0.9438	0.9382
5	0.9662	0.9775	0.9550	0.9719	0.9719	0.9550	0.9494
Average Accuracy	0.918475	0.924275	0.928325	0.9185	0.92695	0.9171	0.903075

TABLE 2. Classification accuracy when train data is used as test data (1 rule /class)

Number of selected Features	F-measure $\beta=1$	Product	AIP	SLAVE	Castro	φ-measure	φ-FIS
2	0.7150	0.7371	0.7150	0.7595	0.8032	0.7550	0.7234
3	0.8741	0.8842	0.8733	0.8630	0.8771	0.8542	0.8514
4	0.8821	0.8763	0.8713	0.8625	0.8763	0.8600	0.8941
5	0.9024	0.8988	0.8900	0.9012	0.8965	0.8956	0.8444
Average Accuracy	0.8434	0.8491	0.8374	0.84655	0.863275	0.8412	0.828325

TABLE 3. Average accuracy of classifier when 10-CV is used five times (1 rule/class)

6.2 **IRIS Data.**

Ranking of features= $\{3, 4, 1, 2\}$

Number of selected Features	F-measure $\beta=1$	Product	AIP	SLAVE	Castro	φ-measure	φ-FIS
2	0.9600	0.9600	0.9667	0.9533	0.9667	0.9667	0.9600
3	0.9667	0.9667	0.9600	0.9667	0.9667	0.9533	0.9600
4	0.9600	0.9600	0.9533	0.9533	0.9600	0.9533	0.9400
Average Accuracy	0.962233	0.962233	0.96	0.957767	0.964467	0.957767	0.953333

TABLE 4. Classification accuracy when train data is used as test data (1 rule /class)

Number of selected Features	F-measure $\beta=1$	Product	AIP	SLAVE	Castro	φ-measure	φ-FIS
2	0.9520	0.94667	0.94	0.94	0.9533	0.9547	0.9387
3	0.9133	0.9227	0.9333	0.9333	0.9267	0.9347	0.9147
4	0.8893	0.8720	0.8420	0.8387	0.9093	0.8413	0.8733
Average Accuracy	0.9182	0.91379	0.9051	0.904	0.929767	0.910233	0.9089

TABLE 5. Average accuracy of classifier when 10-CV is used five times (1 rule /class)

6.3 **Breast-cancer-Wisconsin.**

Ranking of features= $\{6, 3, 2, 7, 8, 1, 4, 5, 9\}$

Number of selected Features	F-measure $\beta=1$	Product	AIP	SLAVE	Castro	φ-measure	φ-FIS
2	0.9077	0.9311	0.9033	0.9238	0.9033	0.9077	0.9021
3	0.9267	0.9355	0.9238	0.9238	0.9428	0.8521	0.8901
4	0.8901	0.9019	0.9150	0.9048	0.9165	0.9150	0.9150
5	0.9077	0.9267	0.9077	0.9136	0.9297	0.9428	0.9236
Average Accuracy	0.90805	0.9238	0.91245	0.9165	0.923075	0.9044	0.9077

TABLE 6. Classification accuracy when train data is used as test data (1 rule /class)

Number of selected Features	F-measure $\beta=1$	Product	AIP	SLAVE	Castro	φ-measure	φ-FIS
2	0.8823	0.8864	0.8525	0.8891	0.8864	0.8631	0.8486
3	0.8662	0.8532	0.8300	0.8531	0.8622	0.8271	0.8706
4	0.8830	0.8655	0.8600	0.8711	0.8800	0.8733	0.8601
5	0.8912	0.8824	0.8804	0.8876	0.8902	0.8830	0.8711
Average Accuracy	0.880675	0.871875	0.855725	0.875225	0.8797	0.861625	0.8626

TABLE 7. Average accuracy of classifier when 10-CV is used five times (1 rule /class)

The results in the Tables 2-7 show that the performances of all the proposed evaluation measures are acceptable and near each other. Moreover, a comparison of the classification accuracy of the seven introduced criteria illustrates that none of the measures is consistently better than the others over the three datasets. Using AIP takes more execution time as a rule evaluation measure to produce the final rule set for classification. However the results justify the use of AIP as a good performance evaluation measure.

7. **Conclusions**

New evaluation criteria adapted from the field of IR have been introduced for fuzzy rule selection for classification purposes. Within the framework of FLCS, the IRL method is used for utilization of these new criteria. These criteria consist of different composite measures of precision and recall. The fuzzy versions of precision, recall and fallout were redesigned as criteria for fuzzy rule selection. The well-known composite criteria such as F-measure, φ-coefficient, precision$\times$recall and the average of interpolated and calculated precisions have been also applied to select the best rule within the genetic process of the IRL approach. Extracting an interpretable rule evaluation measure is accomplished by using the φ-measure and the Wang-Mendel method. The extracted Mamdani type FIS called φ-FIS is utilized for rule evaluating. Although the proposed φ-FIS takes more execution time than the other methods, it is not as computationally prohibitive as AIP. The performance and accuracy of fuzzy rules extracted from numerical data were evaluated on three well-known datasets. The results show the effectiveness of these measures. It is notable that none of the criteria is consistently better than others over a particular dataset. The AIP is a more time-consuming measure than others and φ-FIS is the most interpretable one. It is apparent from the results of this paper that composite evaluation criteria from IR can be used successfully in the fuzzy rule learning processes.

REFERENCES

[1] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen and A. I. Verkamo, *Fast discovery of association rules*, In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy (Editors.) Advances in Knowledge Discovery and Data Mining, AAAI Press, (1996), 307-328.

[2] R. Agrawal, and R. Srikant, *Fast algorithms for mining association rules*, Proc. of 20^{th} International Conference on Very Large Data Bases, (1994), 487-499.

[3] Richardo Baeza-Yates and Berthier Ribeiro-Neto, *Modern information retrieval*, New York, ACM Press, Addison-Wesley, 1999.

[4] L. B. Booker, D. E. Goldberg and J. H. Holland, *Classifier systems and genetic algorithms*, Artificial Intelligence, **40** (1989), 235-282.

[5] L. Castillo, A. Gonzanlez and R. Perez, *Including a simplicity criterion in the selection of the best rule in a genetic fuzzy learning algorithm*, Fuzzy Sets and Systems, **120 (2)** (2001), 309-321.

[6] L. Castro, J. J. Castro-Schez and J. M. Zurita, *Use of a fuzzy machine learning technique in the knowledge acquisition process*, Fuzzy Sets and Systems, **123 (3)** (2001), 307-320.

[7] S. M. Chen and C. H. Yu, *A new method to generate fuzzy rules from training instances for handling classification problems*, Cybernetics and Systems, **34** (2003), 217-232.

[8] O. Cordon, F. Herrera, F. Hoffman and L. Magdalena, *Genetic Fuzzy Systems*, World Scientific, 2001.

[9] K. A. De Jong, W. M. Spears and F. D. Gordon, *Using genetic algorithm for concept learning*, Machine Learning, **13** (1993), 161-188.

[10] P. A. Devijver and J. Kittler, *Pattern Recognition: A statistical Approach*, Englewood Cliffs: Prentice Hall, 1982.

[11] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, 1996.

[12] A. Gonzalez and R. Perez, *Completeness and consistency conditions for learning fuzzy rules*, Fuzzy Sets and Systems, **96** (1998), 37–51.

[13] A. Gonzanlez and R. Perez, *SLAVE: A genetic learning system based on an iterative approach*, IEEE Trans. On Fuzzy Systems, **7 (2)** (1999), 176-191.

[14] F. Herrera, M. Lozano and J. L. Verdegay, *Generating rules from examples using genetic algorithms*. Fuzzy Logic and Soft Computing, Word Scientific, (1995), 11-20.

[15] J. H. Holland and Escaping Britleness: *The possibilities of general purpose learning algorithms applied to parallel rule-based systems*, Machine Learning: An AI Approach, Vol. **2**, Morgan-Kaufmann, (1986), 593-623.

[16] H. Ishibuchi and T. Yamaoto, *Comparison of heuristic criteria for fuzzy rule selection in classification problems*, Fuzzy Optimization and Decision Making, **3 (2)** (2004), 119-139.

[17] H. Ishibuchi, T. Yamamoto and T. Nakashima, *Fuzzy data mining: Effect of fuzzy discretization*, Proc. of 1st IEEE International Conference on Data Mining, (2001), 241-248.

[18] C. Z. Janikow, *A knowledge intensive genetic algorithm for supervised learning*, Machine Learning , **13** (1993), 198-228.

[19] D. E. Kraft and A. Bookstein, *Evaluation of information retrieval system: A decision theory approach*, Journal of the American Society for Information Science, **29** (1978), 31-40.

[20] D. J. Newman and S. Hettich, C.L. Blake and C.J. Merz, (1998). UCI Repository of machine learning databases [http://www.ics.uci.edu/~mlearn/MLRepository.html], Irvine, CA: University of California, Department of Information and Computer Science.

[21] J. A. Roubos and M. Setnes, *Compact fuzzy models through complexity reduction and evolutionary optimization*. In FUZZ-IEEE, San Antonio, USA, (2000), 762-767.

[22] J. A. Roubos, M. Setnes and J. Abonyi, *Learning fuzzy classification rules from data*, Developments in Soft Computing, In John, R. And Birkenhead, R. (Editors), Springer - Verlag Berlin/Heidelberg, (2001), 108-115.

[23] M. Setnes, R. Babuska, U. Kaymak and H. R. van Nauta Lemke*, Similarity measures in fuzzy rule base simplification*, IEEE Trans. SMC-B, **28** (1998), 376-386.

[24] S. F. Smith, *A learning system based on genetic adaptive algorithms*, PhD Thesis, University of Pittsburgh, 1980.

[25] C. J. Van Rijsbergen, *Information Retrieval*, Butterworths, 1979.

[26] G. Venturini, *SIA: A Supervised Inductive Algorithm with Genetic Search for Learning Attributes based Concepts*. European Conference on Machine Learning, (1993), 280-296.

[27] L. X. Wang and J. M. Mendel, *Generating fuzzy rules by learning from examples*, IEEE Trans. Syst., Man, Cybern., **22 (6)** (1992), 1414-1427.

MEHDI EFTEKHARI *, DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, SHIRAZ UNIVERSITY, SHIRAZ, IRAN
E-mail address: `eftekhar@shirazu.ac.ir`

MANSOUR ZOLGHADRI JAHROMI, DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, SHIRAZ UNIVERSITY, SHIRAZ, IRAN
E-mail address: `zjahromi@shirazu.ac.ir`

SERAJEDDIN KATEBI, DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, SHIRAZ UNIVERSITY, SHIRAZ, IRAN

E-mail address: **katebi@shirazu.ac.ir**

* CORRESPONDING AUTHOR