Linguistic Modifiers to improve the Accuracy-Interpretability Trade-off in Multi-Objective Genetic Design of Fuzzy Rule Based Classifier Systems

Alessandro G. Di Nuovo and Vincenzo Catania
Dipartimento di Ingegneria Informatica e delle Telecomunicazioni
Università degli Studi di Catania, Viale A. Doria 6, 95125 Catania, Italy
Email:{adinuovo,vcatania}@diit.unict.it.

***Abstract*—In the last few years a number of studies have focused on the design of fuzzy rule-based systems which are interpretable (i.e. simple and easy to read), while maintaining quite a high level of accuracy. Therefore, a new tendency in the fuzzy modeling that looks for a good balance between interpretability and accuracy is increasing in importance. In fact, recently multi-objective evolutionary algorithms have been applied to improve the difficult trade-off between interpretability and accuracy. In this paper, we focus both on rule learning and fuzzy memberships tuning proposing a technique based on a multi-objective genetic algorithm (MOGA) to design deep-tuned Fuzzy Rule Based Classifier Systems (FRBCSs) from examples. Our technique generates a FRBCS which includes certain operators (known as linguistic hedges or modifiers) able to improve accuracy without losses in interpretability. In our proposal the MOGA is used to learn the FRBCS and to set the operators in order to optimize both model accuracy and metrics of interpretability, compactness and transparency in a single algorithm. The resulting Multi-Objective Genetic Fuzzy System (MOGFS) is evaluated through comparative examples based on well-known data sets in the pattern classification field.**

I. Introduction

The design of fuzzy rule-based systems comes with two contradictory requirements in the obtained model [1]: the interpretability, i.e. the capability to understand the behavior of the real system, and the accuracy, i.e. the capability to faithfully represent the real system. These objectives are conflicting, which means that improving one of them will probably worsen the other. Such problems are known as multi-criteria optimization problems and their optimal solutions are usually sub-optimal for each objective. For an overview see [2], [3].

Whereas the definition of accuracy in a certain application is straightforward, the definition of interpretability is rather problematic. Most researchers and practitioners would agree in interpretability involving the following aspects [4]:

- The number of rules is enough to be comprehensible.
- The rule premises should be easy in structure and contain only a few input variables.
- The linguistic terms should be intuitively comprehensible.
- The inference mechanism should produce technically and intuitively correct results.

Generally speaking, two techniques can be used to build a fuzzy system. The first consists of extracting the necessary knowledge directly from experts; this approach usually generates easily understandable fuzzy systems, but it is not easy and often not applicable. Another technique is to acquire the knowledge automatically from numerical data, that represent samples or examples of the problem. At present a vast number of algorithms exist for automatic data-based fuzzy modeling; popular approaches belonging to EA family are the *Genetic Fuzzy Systems* (GFSs) [5], which has become an important research area during the last fifteen years. GFS have proved to be capable of building compact and transparent fuzzy models while maintaining a very good level of accuracy [6], [7].

Obtaining high degrees of interpretability and accuracy is a contradictory purpose, and, in practice, one of the two properties prevails over the other. Nevertheless, a new trend in the fuzzy modeling scientific community, that looks for a good balance between interpretability and accuracy is increasing in importance. In situations where the best solution corresponds to a trade-off between the different objectives only a multi-objective algorithm will be able to find it. Multi-Objective Genetic Algorithms are an important research line within GAs due to the fact that to population-based algorithms are capable of capturing a set of non-dominated solutions in a single run of the algorithm. With MOGAs the framework of Pareto optimality is embraced, where the algorithm gives a set of trade-off solutions, called Pareto set, among which it is possible to choose the one that is suitable for a specific task. The use of MOGAs allows both model accuracy and metrics of interpretability to be included.

In this paper we propose a Multi-Objective Genetic Fuzzy System (MOGFS) which is able to learn a fuzzy rule-based classifier system (FRBCS) from a database of numerical examples, optimizing the feature weights, number of fuzzy sets, linguistic hedges, and rule weights. Summarizing, there are four objectives: model accuracy, comprehensibility, complexity, and transparency. To evaluate these objectives we use the number of misclassification, the number of rules, the number of features, and the number of the linguistic hedges. Details

978-0-7695-3872-3/09 $26.00 © 2009 IEEE
DOI 10.1109/ISDA.2009.97

of the evaluation measures used are given in Section V-B.

This contribution is structured as follows. Section II discusses similar works. Section III describes the MOGA algorithm we used. Section IV introduces linguistic hedges and the operators we used. Section V introduces our MOGFS approach. Section VI shows some experimental results obtained. Finally, Section VII concludes the paper.

II. Similar Works

Improvement of the interpretability of rule-based systems is a central issue in recent GFS research, where not only accuracy is receiving attention but also the compactness and interpretability of the rules obtained. In [8], the use of MOGAs was considered one of the most promising future directions of GFSs. In [9] an overview of the GFS field (a taxonomy, current research trends and prospects) is given. In [10] were presented and analyzed six different MOGAs to obtain simpler and still accurate linguistic fuzzy models by performing rule selection and a tuning of the membership functions. Recently, MOGAs have been used in new GFS models and applications. Among others we recall in [11] a Pareto-based approach to the identification of Mamdani fuzzy systems is presented. The approach uses an appropriate implementation of the well-known (2+2)PAES to generate a Pareto set of Mamdani fuzzy systems from numerical data. The solutions are characterized by a high accuracy and a good comprehensibility.

Ishibuchi and Yamamoto, in [12], proposed the idea of using three-objective genetic local search algorithms and rule evaluation measures as rule selection criteria for prescreening candidate fuzzy if-then rules used in rule selection. Their strategy operates in two distinct phases. The first phase generates candidate rules through rule evaluation measures whereas the second step selects rules via multi-objective evolutionary algorithms. The three objectives involved are the classification error to measure accuracy, the number of rules and the conditions within the fuzzy classification rule system to measure its comprehensibility or complexity, respectively. All the objectives have to be minimized.

In [13] Wang *et al.* proposed a new scheme based on a multi-objective hierarchical genetic algorithm (MOHGA) to extract interpretable rule-based knowledge from data. The approach is derived from the use of MOGA, where the genes of the chromosome are arranged into control genes and parameter genes. These genes are in a hierarchical form so that the control genes can manipulate the parameter genes in a more effective manner. The effectiveness of this chromosome formulation enables the fuzzy sets and rules to be optimally reduced. In order to remove the redundancy of the rule base proactively, the authors further apply an interpretability-driven simplification method to newborn individuals. In the Wang *et al.* approach, fuzzy clustering is applied first to generate an initial rule-based model. Then the multi-objective hierarchical genetic algorithm and the recursive least square method are used to obtain the optimized fuzzy models.

In former works [14], [15], we demonstrated that good ways to improve accuracy without losses in interpretability are an appropriate assignment of *feature weights* and the use of *linguistic hedges*. Feature weight assignment can be regarded as a generalization of feature selection, which is useful to prune the useless features in order to reduce system complexity. Using this technique a real number w in the range $[0, 1]$ is assigned to each feature indicating the importance of the feature. When $w_i = 0$ the i-th feature has no importance and it can be erased from the system without performance losses. A linguistic hedge (also known as linguistic modifier [16]) is an operator modifying the shape of membership functions. The linguistic operators we used in this work are detailed in section IV. *Rule weights*, i.e. numbers that describe the certainty of the rules, also proved to be useful to improve system performance, while maintaining its interpretability [17], [12].

III. Multi-Objective Evolutionary Algorithm

For this work we chose SPEA2 [18], which is very effective in sampling from along the entire Pareto-optimal front and distributing the solutions generated over the trade-off surface. SPEA2 is an elitist multiobjective evolutionary algorithm which incorporates a fine-grained fitness assignment strategy, a density estimation technique, and an enhanced archive truncation method.

A. Crossover operator with variable-length chromosome

Crossover operator should be compatible with variable lengths of pair of parent chromosomes, and maintains integrity of their offspring. We apply the modulo crossover operator [19] in our implementation of SPEA2. Consider two chromosomes of length (in bits) l_1 and l_2, where both l_1 and l_2 are integer multiples of the length of a single cluster center. Let $r > 0$ be an integer drawn from a uniformly random distribution. If $c_1 = (r \bmod l_1)$ and $c_2 = (r \bmod l_2)$ are taken as crossover points for two individuals, crossover results in two offspring and the sum of whose lengths remains $(l_1 + l_2)$. Each offspring is then guaranteed to be legal representation of fuzzy classifier system solution.

IV. Fuzzy Sets with Linguistic Hedges

In this paper, we focus on mining fuzzy rules of the following schema:

IF $\mathbf{x_i}$ is G_r THEN $\mathbf{x_i}$ is C_j with b_{ij}.

Here, $\mathbf{x_i}$ is a d-dimensional real vector of useful information about the i-th pattern. G_r is a d-dimensional fuzzy relation. $\mathbf{C}$ is a set of possible classes that the i-th pattern could belong with b_i possibility (possibility like probability is a value between 0 and 1, but the sum of possibility vector $\mathbf{b_i}$ need not be 1). Fuzzy sets, according with their significance, could be associated with linguistic labels, for example we could have the following rule:

$$\begin{array}{c}\texttt{IF weather is cold THEN} \\ \texttt{season is winter with 0.8}\end{array} \quad (1)$$

Certain operators may be included to slightly change the meaning of the linguistic labels involved in a specific linguistic fuzzy rule. As Zadeh highlighted in [16], a way to do so with a minor description loss is to use linguistic hedges. A linguistic hedge (also known as linguistic modifier) is an operator modifying the shape of membership functions. Linguistic hedge operations are classified into three categories: concentration, dilation, and contrast intensification [16]. In this paper only the concentration type and the dilation type hedge operations are employed.

Concentration - Applying a concentration operator to a fuzzy set A results in the reduction in the magnitude of the grade of membership of x in A which is relatively small for those x with a high grade of membership in A and relatively large for those x with a high grade of membership in A and relatively large for those x with low membership. The linguistic-hedge operation of "*concentration x*" defined by Zadeh [16] is :

$$CON(x) \triangleq x^{\alpha}; \alpha > 1 \tag{2}$$

Based on the above definition, a few related linguistic-hedge operations such as *absolutely, very, much more, more and plus* [20] can be defined by specifying the values of α in (2) as 5, 4, 3, 2 and 1.5, respectively.

Dilation - In contrast, the effect of dilation is opposite to that of concentration. The linguistic-hedge operation of "*dilation x*" defined by Zadeh [16] is

$$DIL(x) \triangleq x^{\alpha}; 0 < \alpha < 1 \tag{3}$$

Similarly, some related linguistic-hedge operations such as *minus, more or less and slightly* [20] are defined by specifying the values of $alpha$ in (3) as 0.75, 0.5, and 0.25, respectively. Take the fuzzy set cold as an example. Figure 1 shows the membership functions $\mu_{cold(t)}$, $\mu_{verycold(t)}$ and $\mu_{moreorlesscold(t)}$. Of course, the fact of using linguistic

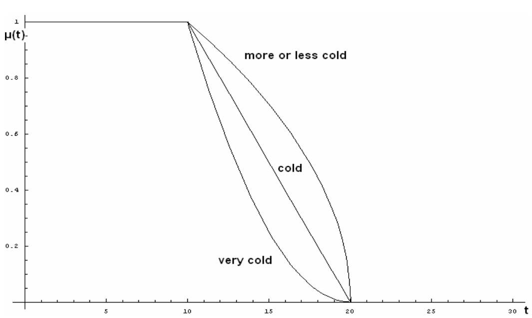

Figure 1. Effects of the fuzzy linguistic hedge very and more or less.

hedges will have a significant influence in the fuzzy rule based system performance since the matching degree of the rule antecedents as well as the output fuzzy set obtained when applying the implication in the inference process will vary. For some proposals that perform this kind of tuning with linguistic hedges, the interested reader can refer to [21] and [22].

V. Proposed Approach

The proposed approach, named MOGFS, is divided into two straightforward stages. The flowchart of the MOGFS is

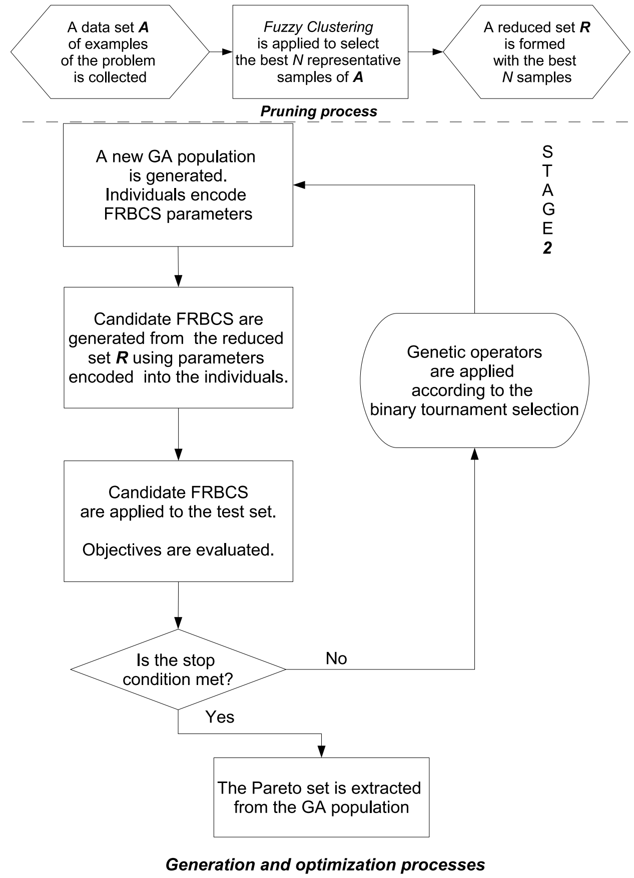

Figure 2. Flowchart of the proposed approach

illustrated in 2. The first stage is the *pruning process*, in which a reduced set R of the best representative samples is extracted from a set A of numerical data. The second stage is composed of two processes, the *generation process*, in which candidate FRBCS are learned with different parameters from the reduced set R using the extension of the Wang and Mendel generating method [23] to classification problem [24], while in the *optimization process* the candidate systems are optimized by means of the SPEA2. The two processes in the second stage are iterated until the stop condition occurs (e.g. until a predefined number of iterations of the SPEA2 is reached). The FRBCS generating method is detailed in Section V-A. System parameters to be optimized are: the feature weights, the rule weights, the number of fuzzy sets for each feature, if linguistic hedges are to be applied or not, and eventually the type of the hedges. These parameters are encoded into the genes of the chromosome, that represent an individual of the population of the SPEA2. Due to the fact that the number of fuzzy sets is a parameter of the system, and that the number of linguistic hedges depends on it, the chromosome length is variable. For this reason we use the variable length crossover operator described in Section III-A. Table I details the chromosome. Note that the total number of fuzzy sets S is the sum of the number of fuzzy sets for each feature (i.e.

the sum of each value of the parameter s for the D features).

Table I
THE CHROMOSOME. D IS THE NUMBER OF FEATURES, R THE NUMBER OF MAXIMUM ALLOWABLE RULES AND S THE TOTAL NUMBER OF FUZZY SETS.

Parameter	No. of Genes	Parameter Space	Type
Feature weights (w)	D	[0,1]	real
Rule weights (r)	R	[0,1]	real
No. of fuzzy sets (s)	D	$2, 3, 5$	integer
Linguistic hedges (α)	$D \times R \times S$	$0.25, 0.50, 0.75$ $1, 1.5, 2, 3, 4, 5$	integer

We have forced the probabilities $p(w = 0)$ and $p(r = 0)$ to be equal to 50% in order to have the same chances either to select or to erase a feature and a rule. $\alpha = 1$ means that the hedge is not applied.

The proposed approach is a GFS belonging to the Pittsburgh family [8], in which each individual represents a rule set. The use of the Wang and Mendel generating method implies that the larger is the set of samples the higher will be the number of possible rules and, consequently, the longer will be the chromosome length. The convergence speed of the GA is influenced by the chromosome length, in fact the longer is the chromosome the slower is the convergence. For this reason, to speed up the convergence, we introduced the pruning process to reduce the number of samples before the execution of the SPEA2. This choice does not affect the overall quality of the solution found as will be shown in Section VI. However, one drawback of the weight specification scheme is that the inherited weights from parents are not always appropriate for their offspring.

The method applied to extract the subset R of the best representative samples is based on fuzzy clustering. The Fuzzy C-Means algorithm [25] is used to partition the whole set into a predefined number of clusters. The subset R is build selecting the patterns which have the highest fuzzy degree of truth with each cluster. Note that there is not a magic number of predefined clusters that is good for all problems, it depends on the characteristics of the problem. In this work it was chosen after a cluster analysis. The optimization phase involves four objectives: model accuracy, comprehensibility, complexity, and transparency. To evaluate these objectives we use the number of misclassifications, the number of rules, the number of features, and the number of the linguistic hedges. Details of the evaluation measures are given in Section V-B. The output of the proposed approach will be a Pareto-set of trade-off FRBCSs.

A. Fuzzy rule based classifier system generating method

The objective of the generating method is to generate from the training data a set of fuzzy rules that describes the relationship between the system and determines a mapping between the feature space and the class set. The generating method used in this work is an extension of Wang and Mendel's rule generation algorithm [23] to the classification problem [24]. It consists of five steps:

- *Step 1* Divides the input space of the given numerical data into fuzzy regions. A membership function is adopted for each fuzzy region. In our experiments we use equidistant membership functions with triangular shapes.
- *Step 2* Generates fuzzy rules from the given data: for each of the D inputs (x_i) the fuzzy set S_i with the highest degree of truth out of those belonging to the term set of the i-th input is selected. After constructing the set of antecedents the consequent is the class to which the pattern belongs.
- *Step 3* Assigns to each of the generated rules a degree equal to the product of the D highest degrees of truth associated with the fuzzy sets chosen S_i.
- *Step 4* Creates a combined fuzzy rule base. In case of conflicts (i.e. repeated rules) they are solved according to rule degrees. More specifically, if the rule base contains two rules with the same antecedents, the degrees associated with the rules are compared and the one with the highest degree wins.

Note that the output space is represented by the class set. This method does not repeat the fuzzy rules. Steps 1 to 4 are iterated with the SPEA2.

When the FRBCS is used to predict the class of an unseen pattern the following straightforward procedure is applied. First it is calculated the degree of truth of the pattern with the R rules. The degree of truth W_j with the j-th rule is defined as the weighted average of the membership degrees (m_i) :

$$W_j = \frac{\sum_{i=1}^{D} m_i \times w_i}{D} \quad (4)$$

where w_i are the feature weights. Then the degree of truth of the pattern with the h-th class C_h is calculated as the weighted average of W_j:

$$C_h = \frac{\sum_{j=1}^{R} W_j \times r_j}{R_h} \quad (5)$$

where r_j are the rule weights (and R_h are the number of rules which have the h-th class as consequent. Finally the pattern is assigned to the class with the highest degree of truth.

B. Evaluation measures

In previous work [12] three objectives are used. The three objectives involved are the classification performance (to be maximized) to measure accuracy, the number of rules and conditions (i.e. fuzzy sets) within the fuzzy classification rule system to measure its comprehensibility and complexity, respectively. Both the last two objectives have to be minimized. In order to compare our experimental results with [12] we chose the same measures used in the previous work to measure accuracy, comprehensibility and complexity. To measure the fourth objective, transparency, we add a new measure: the number of linguistic hedges. Summarizing the following measures are used in our MOGFS:

- *Number of misclassification*, it is used to measure the accuracy of the classification. It is defined as the number of patterns that are in a class different than the one in the reference classification. It ranges from 0 to the total number of patterns.
- *Number of rules*, it counts the fuzzy rules present in the system. It is a measure of the comprehensibility of the system. It can take values from the number of classes to the maximum allowable number of rules (R).
- *Average rule length*, it is the average number of conditions (i.e. fuzzy sets) in the rules. It is a measure of the complexity of the system. Its value depends on the number of selected features (D_s) that is a variable of the system. It can vary from 1 to D_s.
- *Number of linguistic hedges*, it is used to measure transparency of the fuzzy sets. It is defined as the number of linguistic hedges applied to the fuzzy sets of the system. It ranges from 0 to the total number of fuzzy sets S which are present in the fuzzy system.

All the measures used are to be minimized and they are integer numbers. Note that when the number of rules is lower than the number of classes the system has a poor accuracy, because in our model there is a one to one correspondence between rules and classes. For this reason we do not allow to have the number of rules lower than the number of classes.

VI. Performance Evaluation and Applications

In this section we will evaluate the performance of the proposed method on some problems in the University of California machine-learning repository [26]. We considered four well known data sets: the two most popular data sets (the Iris and the Wine data sets) to benchmark the performance of the proposed method.To avoid over-fitting, the leave-one-out procedure was used. The SPEA2 was run ten times and tables show the average performance on the test sets. The population for the genetic algorithm was set as 500 individuals, using a crossover probability of 0.8 and a mutation probability of 0.1. The stop criterion used for SPEA2 it was 1000 generations. After a cluster analysis we chose 19 and 31 as the number of samples in R in the case of Iris data and Wine data, respectively.

A. Iris data set

The most classical data set used to compare classification methods is the Iris data set. The data set comprises 150 flowers belonging to 3 subspecies of Iris: setosa, versicolor and virginica. For each subspecies the data set contains 50 observations of four main features: Sepal Length (SL), Sepal Width (SW), Petal Length (PL) and Petal Width (PW).

Results on the iris data set are summarized in Table II, which reports the set of non-dominated FRBCS with different complexity and accuracy obtained by our MOGFS. The FRBCS are described by their values on the measures of the four objectives. Table II shown also the rule set obtained by

Table II
Comparative study on Iris data. Error rate is on test data (leave-one-out procedure was used)

No. of rules	Avg rule length	Error rate (%)	No. of Hedges
Proposed MOGFS			
9	2.00	**2.0**	16
4	2.00	3.4	8
4	2.00	4.0	6
3	**1.00**	4.6	2
MOGLS - results from [12]			
3	2.00	3.6	0
3	1.66	5.1	0
3	1.33	5.5	0
3	1.00	6.1	0
Wang & Mendel - triangular sets			
43	4.0	3.3	0
15	4.0	6.1	0
8	4.0	16.7	0

MOGLS [12]. The accuracy achieved by the best FRBCS found by our MOGFS is better than the one obtained by MOGLS, and it is as good as one of the best found in literature [27] for FRBCS. It should be stated that the rules of MOGFS with higher classification ability are longer and more complex than those of MOGLS. However the smallest FRBCS found by our MOGFS is as simple as the smallest found by MOGLS, but our FRBCS has higher accuracy. This result is due to the use of two linguistic hedges, which are able to improve accuracy without significant losses in interpretability.

B. Wine data set

These data are the results of a chemical analysis of 178 wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines. These constituents are the attributes of data set, they are all continuous. The number of patterns in three classes are 59, 71 and 48 respectively. The number of samples in R is 31.

Results on the wine data are summarized in Table III in the same manner as Table II in the previous subsection. That is, a tradeoff between accuracy and interpretability of FRBCS is clearly shown by obtained rule sets. As shown in Table III the non-dominated FRBCS obtained by our MOGFS outperform the rule set obtained with MOGLS both in maximum accuracy and interpretability. Here again it should be stated that the rules of MOGFS with higher classification ability are longer and more complex than those of MOGLS. This is due to the fact that MOGLS is a local search algorithm, that found only trade-off classifiers with three rules. It should be noted that, differently from other approaches, our MOGFS does not use the *don't care* condition, which can help to reduce the number of fuzzy sets in the system. Our FRBCS with maximum accuracy is very close to the best one reported in literature [7], where an error rate of 1.1% was obtained with 5 rules and 13 Gaussian-shaped fuzzy sets. The good performance obtained with this data set suggest that the proposed approach has the potential to perform well even with high dimensionality.

Table III
COMPARATIVE STUDY ON WINE DATA. ERROR RATE IS ON TEST DATA (LEAVE-ONE-OUT PROCEDURE WAS USED)

No. of rules	Avg rule length	Error rate (%)	No. of Hedges
Proposed MOGFS			
5	6.00	**1.7**	44
5	5.00	3.3	35
5	4.00	4.5	24
4	4.00	5.6	10
3	3.00	6.7	8
3	2.00	7.3	5
3	**1.00**	10.1	2
MOGLS - results from [12]			
3	2.66	2.8	0
3	2.33	3.9	0
3	2.00	6.1	0
3	1.66	7.2	0
3	1.33	9.9	0
3	1.00	13.7	0
Wang & Mendel			
177	14.0	6.7	0
138	14.0	8.4	0
124	14.0	9.0	0

VII. CONCLUSION

The algorithms proposed in the literature to construct fuzzy systems from examples usually refine a single model iteratively until a compromise between its complexity and its approximation error is found, but we can conclude that this is not an adequate approach because there may exist more than one alternative optimal solution. In this paper, we proposed an idea of using the SPEA2 for the design of fuzzy systems which are interpretable (i.e. simple and easy to read), while maintaining quite a high level of accuracy. Our MOGFS is able to learn FRBCS from a database of numerical examples, optimizing the feature weights, the linguistic hedges, and the rule weights. The output of our MOGFS is a non-dominated set of trade-off solutions among which it is possible to choose the one that is suitable for a specific task. We compared our approach with others to be found in literature in empirical tests on two data sets that summarize the main problems encountered. The tests showed that the proposed approach provides a set of FRBCS that outperform other existing methods. In fact our MOGFS approach leads to FRBCS with a small number of transparent, readable rules, which are less complex than those reported in the literature with comparable or better accuracy.

REFERENCES

[1] J. Casillas, F. Herrera, R. Pérez, M. J. del Jesus, and P. Villar, "Preface: Special issue on genetic fuzzy systems and the interpretability–accuracy trade-off," *Int. J. Approx. Reasoning*, vol. 44, no. 1, pp. 1–3, 2007.

[2] J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, *Accuracy Improvements in Linguistic Fuzzy Modeling*. Springer-Verlag, 2003.

[3] ——, *Interpretability Issues in Fuzzy Modeling*. Springer-Verlag, 2003.

[4] R. Mikut, J. Jäkel, and L. Gröll, "Interpretability issues in data-based learning of fuzzy systems," *Fuzzy Sets and Systems*, vol. 150, pp. 179–197, 2005.

[5] O. Cordón, F. Herrera, F. Hoffmann, and L. Magdalena, *GENETIC FUZZY SYSTEMS Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*, ser. Advances in Fuzzy Systems - Applications and Theory. World Scientific, 2001, vol. 19.

[6] H. Roubos and M. Setnes, "Compact and trasparent fuzzy models and classifiers through iterative complexity reduction," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 4, pp. 516–524, 2001.

[7] X. Chang and J. H. Lilly, "Evolutionary design of a fuzzy classifier from data," *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics*, vol. 34, no. 4, pp. 1894–1906, 2004.

[8] F. Herrera, "Genetic fuzzy systems: Status, critical considerations and future directions," *International Journal of Computational Intelligence Research*, vol. 1, pp. 59–67, 2005.

[9] ——, "Genetic fuzzy systems: taxonomy, current research trends and prospects," *Evolutionary Intelligence*, vol. 1, no. 1, pp. 27–46, 2008.

[10] M. J. Gacto, R. Alcala, and F. Herrera, "Adaptation and application of multi-objective evolutionary algorithms for rule reduction and parameter tuning of fuzzy rule-based systems," *Soft Computing*, vol. 13, pp. 419–436, 2009.

[11] M. Cococcioni, P. Ducange, B. Lanzerini, and F. Marcelloni, "A pareto-based multi-objective evolutionary approach to the identification of mamdani fuzzy systems," *Soft Computing Journal*, vol. 11, pp. 1013–1031, 2007.

[12] H. Ishibuchi and T. Yamamoto, "Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining," *Fuzzy Sets and Systems*, vol. 141, pp. 59–88, 2004.

[13] H. Wang, S. Kwonga, Y. Jin, W. Wei, and K. Man, "Multi-objective hierarchical genetic algorithm for interpretable fuzzy rule-based knowledge extraction," *Fuzzy Sets and Systems*, vol. 149, pp. 149–186, 2005.

[14] A. G. Di Nuovo and V. Catania, "An efficient approach for the design of transparent fuzzy rule-based classifiers," in *Proceedings of 2006 IEEE International Conference on Fuzzy Systems (FUZZIEEE '06)*, Vancouver, Canada, 2006, pp. 6941–6947.

[15] A. G. Di Nuovo and V. Catania, "Genetic tuning of fuzzy rule deep structures for efficient knowledge extraction from medical data," in *2006 IEEE International Conference on Systems, Man, Cybernetics*, Oct. 2006, pp. 5053–5058.

[16] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.

[17] O. Cordón, M. del Jesus, and F. Herrera, "A proposal on reasoning methods in fuzzy rule-based classification systems," *International Journal of Approximate Reasoning*, vol. 20, pp. 21–45, 1999.

[18] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the performance of the strength pareto evolutionary algorithm," in *EUROGEN 2001. Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems*, Athens, Greece, Sep. 2001, pp. 95–100.

[19] R. Srikanth, R. George, and N. Warsi, "A variable-length genetic algorithm for clustering and classification," *Pattern Recognition Letters*, vol. 16, pp. 789–800, 1995.

[20] M.-S. Chen and S.-W. Wang, "Fuzzy clustering analysis for optimizing fuzzy membership functions," *Fuzzy Sets and Systems*, vol. 103, pp. 239–254, 1999.

[21] A. Gonzalez and R. Perez, "A study about the inclusion of linguistic hegdes in a fuzzy rule learning algorithm," *International Journal of Uncertainty, Fuzziness, Knowledge-Based Systems*, vol. 7, pp. 257–266, 1999.

[22] J. Casillas, O. Cordón, M. J. Del Jesús, and F. Herrera, "Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction," *IEEE Transactions on Fuzzy Systems*, vol. 13, pp. 13–29, Feb. 2005.

[23] L.-X. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Transactions on System, Man and Cybernetics*, vol. 22, pp. 1414–1427, 1992.

[24] J. Casillas, O. Cordón, M. J. Del Jesús, and F. Herrera, "Genetic feature selection in a fuzzy rule-based classification system learning process for high-dimensional problems," *Information Sciences*, vol. 136, no. 1-4, pp. 135–157, 2001.

[25] J. C. Bezdek, *Pattern Recognition with fuzzy objective function algorithms*. New York: Plenum Press, 1981.

[26] "University of california machine learning repository," http://www.ics.uci.edu/~mlearn/MLRepository.html.

[27] K. Nozaki, H. Ishibuchi, and K. Tanaka, "Adaptive fuzzy rule-based classification systems," *IEEE Transactions on Fuzzy Systems*, vol. 4, pp. 238–250, 1996.