

# Parallel Evolutionary Multiobjective Methodology for Granularity and Rule Base Learning in Linguistic Fuzzy Systems

Juan M. Bardallo, Miguel A. De Vega, Francisco A. Márquez, Antonio Peregrín

**Abstract—** In this paper we present a parallel evolutionary multi-objective methodology for granularity and rule-based learning for Mamdani Fuzzy Systems. The proposed methodology produces a set of solutions with different trade-off between accuracy and interpretability, based on searching the number of labels and the fuzzy rules, and also makes a variable selection. This process is achieved by exploiting present parallel computer systems allowing it to deal with more complex models.

## I. INTRODUCTION

THE main objective in system modeling is to develop reliable and understandable models. Interpretability and accuracy are usually contradictory requirements in the design of linguistic fuzzy models (FMs). Recent research on genetic fuzzy systems [1] has focused on methods aimed at generating fuzzy rule-based systems (FRBS) with an appropriate trade-off between accuracy and interpretability [2], [3].

Recently, Multi-objective Evolutionary Algorithms (MOEAs) has been used to improve the aforementioned trade-off between interpretability and accuracy of linguistic fuzzy systems [4] - [10]. Some of them achieve the Pareto (the set of non-dominated solutions with different trade-off) by selecting [9], [10] the set of rules best representing the example data, i.e., providing a set of solutions with different balance between the complexity of the fuzzy rule base (RB) and system accuracy.

On the other hand, there are many works devoted to the design of different elements of fuzzy systems, such as the derivation of the linguistic RB [11] - [15], the tuning of the meanings of the linguistic values used in the rules [16], [17], the learning of the number of the labels for each variable (granularity) [18], [19], and the setup of the inference system [20] - [22] and defuzzification method [23] between others, or the combination of some of these techniques [24], [25]. However the combination of different elements increases the search space, in particular when we deal with complex models with large data sets. The effect produced by the size of the data in the algorithms is called the scaling problem.

Taking the foregoing into account, in this work, we

present a simple evolutionary multi-objective methodology which aims to learn the granularity and the RB, but which also has the capability to reduce the number of variables concerned. The evolutionary methodology proposed generates a set of FRBSs with different optimal trade-off between accuracy and interpretability. The problem search space is composed of all the possible combinations of granularities for the involved variables and all their associated RBs, which is a huge search space especially if there are a lot of variables. We exploit the power of parallel systems to deal with the scaling problem. Nowadays, the use of a collection of computers to achieve a greater amount of computational resources has become more popular as they are much more cost-effective than single computers of comparable speed. The proposed methodology deals with large search spaces and large data sets, taking advantage of this type of computer clusters or multi-core processor technology in a simple way, obtaining good quality practical results.

In order to explain how this is achieved, Section II describes the proposed model, Section III shows the experimental study developed, and finally Section IV presents some concluding remarks.

## II. PROPOSED MODEL

This section describes the parallel methodology proposed in this work. First, we explain the global mechanism and afterwards, two subsections depict the Central and Subordinate Nodes respectively.

In [19], the influence of fuzzy partition granularity in FRBSs performance was studied, showing that the use of an appropriate number of linguistic terms for each variable influences the accuracy of the model. On the other hand, the interpretability [2] is influenced by several factors: the number of terms (the fewer, the better); the number of variables involved (as it is sometimes possible to have a model with few variables that is still accurate); and additionally the number of rules (since a compact RB is preferable in order to achieve a more interpretable system).

As mentioned above, the proposed model includes the learning of the granularity, the RB and variable selection. In fact, the RB depends on the labels chosen for each involved variable. Thus, the proposed methodology performs a main search of the granularity and variable selection, and carries out a subordinate search of the associated RB for each granularity combination candidate in the main search process.

This work was supported by Projects TIN2008-06681-C06-06 and P07-TIC-03179.

J. M. Bardallo, M. A. De Vega, F. A. Márquez and A. Peregrín are with the Department of Information Technologies, University of Huelva, 21819 Palos de la Fra. Huelva, Spain (e-mail: juanmanuel.bardallo@sic.uhu.es; miguelangelde.vega@sic.uhu.es; alfredo.marquez@dti.uhu.es; peregrin@dti.uhu.es).

In short, the model proposed consists of a multi-objective algorithm designed to find a set of granularity combinations of the variables, and then perform a simple subordinate parallel search to find a compact RB for the proposed granularity.

This is implemented using a node that performs the multi-objective algorithm. This node is a single unit which we have named the Central Node, and connects to a set of subsidiary nodes that perform the parallel search, and which we have named Subordinate Nodes.

The search space for the granularity is not huge, while the subordinate search in the parallel nodes is expensive in terms of computational cost, and it is achieved in parallel form. Communication times are somewhat slower than subordinate search processes, so the model can operate with parallel computing resources without considerable drawbacks.

#### A. Central Node: Multi-objective Schema

As we mentioned above, the Central Node performs a multi-objective search in order to find the granularity of the variables. The number of labels for each variable is limited in order to avoid higher values that may negatively influence interpretability; this has the added benefit of reducing the search space.

The variable selection is implemented through the granularity mechanism: we let the null value pay no attention to the variable concerned.

The multi-objective schema has been implemented using one of the most representative second generation MOEA: the NSGA-II [26]. It is one of the most well-known and frequently-used MOEAs for general multi-objective optimization in the literature. It is a parameterless approach with many interesting principles: a binary tournament selection based on a fast non-dominated sorting, an elitist strategy and a crowding distance method to estimate the diversity of a solution. A fuller description may be found in [26].

The coding scheme used in the Central Node is the one shown in Fig. 1. Every chromosome of the multi-objective algorithm encodes the number of linguistic labels for each variable. More specifically, each chromosome has got as many genes as the problem has variables. Hence, genes are integer values that indicate the number of linguistic terms of their respective variables, likewise for antecedents and consequents, as can be seen in Fig.1. The minimum and maximum number of linguistic terms that each variable may have is predetermined, and they are uniformly distributed in the universe of the variable.

Genes in Central Node are randomly initiated to a value between the predetermined minimum and maximum values. Value “0” indicates that the variable represented by this gene can be ignored. The way it performs the variable selection is given below.

The two objectives considered by the MOEA are the error (as a measure of accuracy) and the number of rules (as a measure of interpretability) of the RB obtained.

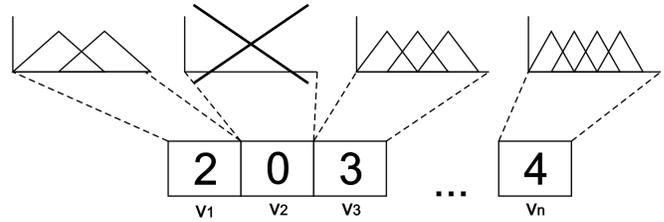


Fig. 1. MOEA chromosome coding

#### B. Subordinate Nodes: Rule-based Learning

The parallel Subordinate Nodes receive the number of linguistic terms for each variable and perform the search for the RBs, returning to the Central Node the RB and its accuracy computed with the training dataset.

A large number of methods have been proposed to automatically generate fuzzy rules from numerical data. Usually, they use complex rule generation mechanisms such as neural networks [27], [28] or genetic algorithms (GAs) [29]. In contrast to these, a family of efficient methods guided by covering criteria of the data in the example set, called ad hoc data-driven methods, has been proposed in the literature [12], [30]–[32].

These methods are suitable for use with the linguistic RB learning in the Subordinate Nodes. However, we have chosen to use those based on ad-hoc data driven methods for their efficiency and simplicity. Specifically, we propose using the COR [11] methodology because it manages a set of consequent labels (one set per rule), and instead of selecting the consequent with the highest performance in each subspace as usual (Wang and Mendel, WM [12]), considers the possibility of using another consequent, different from the best one, when the alternative confers greater accuracy on the FRBS thanks to having a RB with best cooperation. For this purpose, COR performs a combinatorial search among the candidate rules looking for the set of consequents which globally achieves the best accuracy.

COR consists of two stages:

- 1) *Search space construction*—It obtains a set of candidate consequents for each rule.
- 2) *Selection of the most cooperative fuzzy rule set*—It performs a combinatorial search among these sets looking for the combination of consequents with the best global accuracy.

In order to perform this combinatorial search, an *explicit enumeration* or an *approximate search technique* can be considered. In this work, we use a generational GA as the search technique for its effectiveness.

The coding scheme used to encode COR, is an integer string of N genes, each one representing a candidate rule consequent of the initial RB. Additionally, to obtain solutions with higher interpretability, we allow a value “don’t care” into the set of consequents. Thus, if this value is selected, the rule is eliminated, effectively performing a rule selection mechanism.

Regarding the Subordinate Node, this uses three stages: first, it generates the RB using a WM algorithm as, in the next step, COR needs this information; second, COR algorithm obtains the rules with the best cooperation; and third, it computes the accuracy over the training data set, and sends the results to the Central Node.

The coding scheme used to encode COR depends on the number of rules found by WM.

The evaluation of the population of the MOEA in the Central Node is carried out in the Subordinate Nodes for every chromosome in its population. If possible, each Subordinate Node corresponds to a parallel computer or core processor, but if the number of Subordinate Nodes is higher than the number of parallel computers or core processors, more than one chromosome can be assigned to each parallel compute unit, so as to distribute the load evenly between all the computers.

### III. EXPERIMENTAL STUDY

In order to analyze the practical behavior of the proposed methodology, we built several FMs in two real-world problems [33], [34], with different complexities (different number of variables and amount of data).

Table I summarizes the different models used in the experimental study, where  $WM_N$  [12] and  $COR_N$  [11] methods are considered as reference,  $N$  being the granularity selected in all the variables. The  $M_N$  methods are the proposed multi-objective models, where  $N$  is the maximum number of labels used by the algorithm. We also decided to test the use of the WM methodology in the Subordinate Nodes, which are denoted by  $M_{N-COR}$  and  $M_{N-WM}$ .

TABLE I.  
METHODS CONSIDERED FOR COMPARISON

Ref.	Method	Description
[12]	$WM_N$	Wang & Mendel algorithm. $N$ labels for each variable.
[11]	$COR_N$	COR algorithm with rule selection and $N$ labels for each variable.
-	$M_{N-WM}$	Parallel multi-objective evolutionary methodology with Wang & Mendel. Variables will have up to $N$ labels
-	$M_{N-COR}$	Parallel multi-objective evolutionary methodology with COR and rule selection. Variables will have up to $N$ labels

#### A. Applications Selected and Comparison Methodology

To evaluate the goodness of the proposed approaches, two real-world problems with different complexities were considered:

- An electrical distribution problem [33] that consists in estimating the maintenance costs of medium voltage lines in a town (1059 cases; 4 input continuous variables; one output variable).
- The Ankara Weather dataset [34] that concerns the task of trying to predict the mean temperature in Ankara, Turkey (1609 cases; 9 input continuous variables; one output variable).

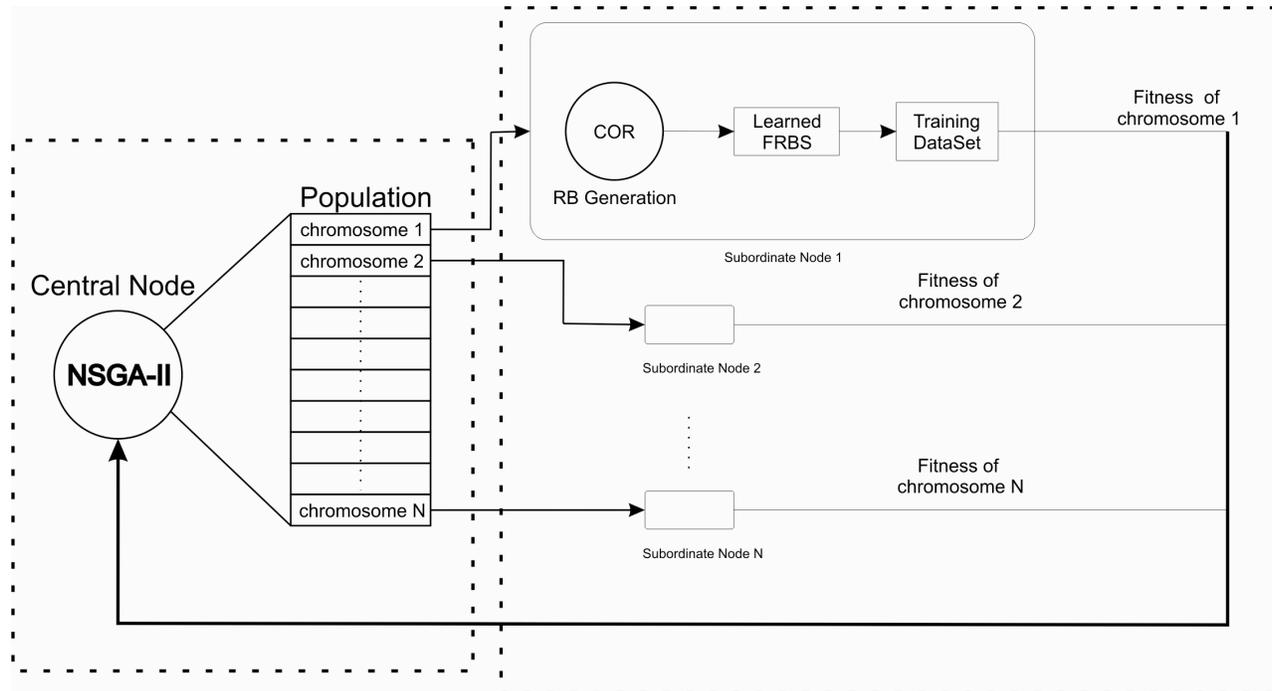


Fig. 2. Detailed global scheme

We considered a 5-fold cross-validation model, i.e., 5 random partitions of the data each with 20% (4 with 211 examples, and one with 212 examples, for the electrical problem, and 4 with 322 examples, and one with 321 examples for the Ankara Weather problem) and the combination of 4 of them (80%) as training, with the remaining one as test. We achieved a total of 30 trials for each model, by running the learning methods for each one of the data partitions 6 times with different seeds for the random number generator. We show the average values of the mean square error (MSE), as a standard performance measure (with expression (1)),

$$MSE(S_B) = \frac{1}{2} \frac{\sum_{k=1}^P (y_k - S(x_k))^2}{P}, \quad (1)$$

where  $S$  denotes the fuzzy model selected, computed considering the most accurate solution from each Pareto obtained with the multi-objective algorithm. This measure uses a set of system evaluation data formed by  $P$  pairs of numerical data  $Z_k = (x_k, y_k)$ ,  $k=1, \dots, P$ , with  $x_k$  being the values of the input variables, and with  $y_k$  being the corresponding values of the associated output variables.

### B. Questions Related to the Genetic Algorithms

The multi-objective algorithm in the Central Node used several configurations depending on the maximum number of linguistic terms set up. When a maximum value of 5 labels was selected in the electrical distribution problem, the population was fixed to 25 chromosomes and the number of generations was 30. On the other hand, when the maximum value of 9 labels for each variable was selected in the electrical distribution problem - just as when the maximum value of 4 labels for each variable was selected in the Ankara Weather problem - the population was fixed at 60 chromosomes and the number of generations was 75. The crossover operator used was HUX- $\alpha$  with  $\alpha = 0.5$ . The mutation operator used was the classic mutation operator with a probability of 0.2. The initial population was randomly initialized within the minimum to maximum range of values for the partitions.

The COR method used in the Subordinate Nodes was set up with a population of 61 chromosomes and 300 generations in all cases. The selection operator was based on Baker's stochastic universal sampling with elitism. The crossover operator used was the two point crossover with a probability of 0.6. The initial population was initialized with the  $N$  rules obtained by the WM method.

To compare the results obtained we also used non-parametric tests, according to the recommendations made in Demšar [35] which suggests a set of simple, safe and robust non-parametric tests for statistical comparisons of algorithms, one of which is the Wilcoxon signed-ranks test [36] we use in this work. It is analogous to the paired t-test in non-parametrical statistical procedures.

We used a server farm of 16 compute nodes based on dual CPU Intel Xeon 3 GHz each with 2 GB of RAM, interconnected with a Gigabit Ethernet network.

### C. Results and Analysis

The results obtained are shown in Table II and III for the electrical distribution problem and in Table IV for the Ankara Weather, where MSE are the average MSE, for training and test, and Wil-test are the results of applying the Wilcoxon signed-ranks test [36] (with 95% confidence), with the following interpretation: \* represents the best average result (control algorithm); + means that the best result has better performance than that of the corresponding row; and finally #R is the average number of rules.

It is important to note that Tables II, III and IV only show the FRBS with the best accuracy of the Pareto front for each multi-objective model. Viewing these Tables we can point out the following:

- The results obtained by the parallel evolutionary multi-objective methodology  $M_{N-COR}$  is in two cases noticeably lower in accuracy (Tables III and IV, viewing Wilcoxon signed ranks test results) as well as in the number of rules, than the reference models,  $WM_N$  and  $COR_N$  (these methods are placed as orientation: we know they cannot be compared because they are using different search spaces and computing resources). The  $M_{N-WM}$  did not always achieve greater accuracy (Table IV), but the best accurate result it finds always has a significant reduction in the number of rules.
- Specifically, viewing Table II,  $M_{5-COR}$  does not show significant accuracy improvements over of  $M_{5-WM}$ , but the number of rules decreases from 10 to 7.
- Comparing Table II with III, that is, letting the algorithm use more linguistic terms (from a maximum of 5 in Table II to 9 in Table III), the accuracy is better, as was expected, but it is due to a significant increase in the number of rules generated based on more linguistic terms.
- The behavior of the parallel evolutionary multi-objective with parallel evaluation methodology presented does not look to be negatively affected by the larger data set, Ankara Weather, against the smaller electrical distribution one, regardless of being limited to a maximum of only 4 linguistic terms.

TABLE II.  
RESULTS OBTAINED FOR THE ELECTRICAL DISTRIBUTION APPLICATION  
WITH A MAXIMUM OF 5 LABELS

Method	Training		Test		#R
	MSE	Wil-test	MSE	Wil-test	
$WM_5$	56135.75	+	56359.42	+	65
$COR_5$	39652.82	+	41952.33	+	43
$M_{5-WM}$	20854.01	=	20826.72	=	10
$M_{5-COR}$	20505.56	*	20134.06	*	7

TABLE III.  
RESULTS OBTAINED FOR THE ELECTRICAL DISTRIBUTION APPLICATION  
WITH A MAXIMUM OF 9 LABELS

Method	Training		Test		#R
	MSE	Wil-test	MSE	Wil-test	
WM <sub>9</sub>	32408.20	+	37549.06	+	124
COR <sub>9</sub>	21318.85	+	24741.38	+	106
M <sub>9-WM</sub>	19606.84	+	19874.90	+	21
M <sub>9-COR</sub>	<b>15334.73</b>	*	<b>17135.04</b>	*	47

TABLE IV.  
RESULTS OBTAINED FOR THE ANKARA WEATHER APPLICATION WITH A  
MAXIMUM OF 4 LABELS

Method	Training		Test		#R
	MSE	Wil-test	MSE	Wil-test	
WM <sub>4</sub>	8.135657	+	8.886438	+	264
COR <sub>4</sub>	3.379250	+	3.986888	+	157.8
M <sub>4-WM</sub>	4.086000	+	4.074000	+	23.2
M <sub>4-COR</sub>	<b>2.282000</b>	*	<b>3.002000</b>	*	27.2

Table V shows an example of the Pareto front obtained by the model M<sub>9-COR</sub>, (the best accurate model in Table III, but with a huge number of rules). Its columns show the MSE for training and test, the number of rules, and the chromosome achieved, that is, the number of linguistic terms found, where 0 means the variable has been removed. The first 4 genes on the left of the chromosome belong to the antecedents and the last one on the right belongs to the consequent. The Pareto front shows 13 different FRBSs learned. We analyze these in the following way:

- The most accurate FRBSs obtained are similar to the aforementioned mean value in Table III: very accurate but with a lot of rules.
- It shows two solutions with 1 and 2 rules (at the bottom of the table) but with an enormous MSE, so these must be ignored.

- Nevertheless, the solutions with 6 and 7 rules are quite interesting for the lower number of rules and the suitable MSE. Viewing the chromosome, the two first variables have been removed, and the granularity of the antecedents is not high, which is coherent with the low number of rules.
- Many other solutions with different trade-offs have been found among the aforesaid.

TABLE V.  
A PARETO FRONT EXAMPLE OBTAINED BY THE MODEL M<sub>9-COR</sub>

M <sub>9-COR</sub>			
MSE <sub>TRA</sub>	MSE <sub>TST</sub>	#R	Chromosome
13965.67	15673.11	51	2 - 6 - 9 - 5 - 9
15481.69	18184.36	39	4 - 0 - 7 - 5 - 9
15898.54	19192.16	33	3 - 0 - 7 - 5 - 9
16731.81	19859.63	26	2 - 2 - 7 - 5 - 9
16740.49	18314.79	24	0 - 2 - 7 - 5 - 9
17143.90	19095.34	23	0 - 0 - 7 - 5 - 9
18489.08	17716.04	22	2 - 2 - 5 - 5 - 9
18617.83	16398.43	17	0 - 0 - 5 - 5 - 9
18691.96	19950.22	7	0 - 0 - 4 - 2 - 9
19543.62	19214.96	6	0 - 0 - 4 - 2 - 5
43636.01	49413.55	3	0 - 0 - 2 - 2 - 9
298944.75	293681.41	2	0 - 2 - 0 - 0 - 6
7337253.50	7167457.00	1	0 - 0 - 0 - 0 - 6

Fig. 3., shows the Pareto front for training and test, for the electrical distribution problem, for M<sub>9-COR</sub> and M<sub>9-WM</sub>. Comparing both figures, they are similar, so there is no over-fitting.

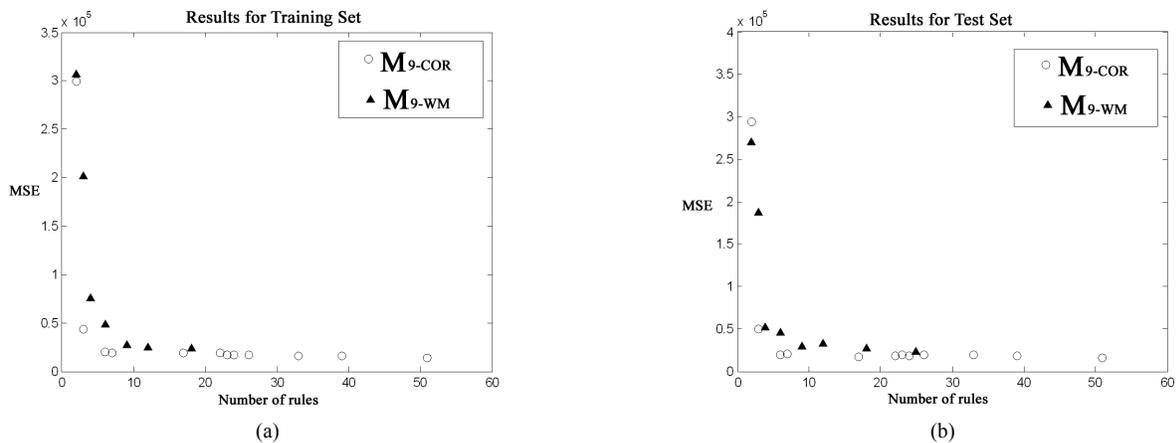


Fig. 3. Example of the Pareto front with WM and COR for the electrical distribution problem. (a) Training (b) Test

## IV. CONCLUSIONS

In the framework of the trade-off between accuracy and interpretability, the use of multi-objective algorithms gives a set of solutions with different levels of conciliation between both features. In this work we have proposed a simple evolutionary parallel multi-objective learning methodology where the number of labels and variable selection are learnt together with the rule base, exploiting the power of current clusters of computers or multi-core processor capabilities to achieve greater amounts of computational power. The parallel methodology proposed deals with more complex models with larger data sets, taking advantage of this type of computing resource, making the design process of linguistic fuzzy systems easier. The experimental study developed shows the good quality results of the proposal both for the accuracy and for the interpretability of the fuzzy systems learned. The authors continue testing the model with many other applications.

## REFERENCES

- [1] F. Herrera. "Genetic Fuzzy Systems: Taxonomy, current research trends and prospects", *Evolutionary Intelligence*, vol. 1, pp. 27–46, 2008.
- [2] J. Casillas, O. Cordon, F. Herrera and L. Magdalena. *Interpretability issues in fuzzy modeling*. Springer-Verlag, 2003.
- [3] J. Casillas, O. Cordon, F. Herrera and L. Magdalena. *Accuracy improvements in linguistic fuzzy modeling*. Springer-Verlag, 2003.
- [4] H. Ishibuchi, "Multiobjective Genetic Fuzzy Systems: review and future research directions", in *Proc of Fuzz-IEEE'07 Int Conf on Fuzzy Syst*, London, 2007, pp. 1–6.
- [5] M. Cococcioni, P. Ducange, B. Lazzarini, and F. Marcelloni, "A pareto-based multi-objective evolutionary approach to the identification of Mamdani fuzzy systems", *Soft Comput*, vol. 11, pp. 1013–1031, 2007.
- [6] 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 Syst*, Vol. 141, pp. 59–88, 2004.
- [7] K. Narukawata, Y. Nojima, and H. Ishibuchi. "Modification of evolutionary Multiobjective optimization algorithms for Multiobjective design of fuzzy rule-based classification systems", in *Proc of Fuzz-IEEE'05 Int Conf on Fuzzy Syst*, Reno, 2005, pp. 809–814.
- [8] R. Alcalá, J. Alcalá-Fdez, M.J. Gacto, and F. Herrera. "A Multi-Objective Evolutionary Algorithm for Rule Selection and Tuning on Fuzzy Rule-Based Systems". *Proc of Fuzz-IEEE'07 Int Conf on Fuzzy Syst*, London, 2007, pp. 1367–1372.
- [9] R. Alcalá, M.J. Gacto, F. Herrera, and J. Alcalá-Fdez. "A Multi-Objective Genetic Algorithm for tuning and Rule Selection to obtain Accurate and Compact Linguistic Fuzzy Rule-Based Systems", *Int J of Uncertainty, Fuzziness and Knowledge-Based Syst*, vol. 15, no. 5, pp. 539–557, 2007.
- [10] A.A. Márquez, F.A. Márquez, and A. Peregrín. "Cooperation between the Inference System and the Rule Base by using Multiobjective Genetic Algorithms". *Proc of HAIS'08 Int Conf of Hybrid Artificial Intelligence Syst*, Burgos, 2008, pp. 739–746.
- [11] J. Casillas, O. Cordon, and F. Herrera. "COR: A methodology to improve ad hoc data-driven linguistic rule learning methods by inducing cooperation among rules", *IEEE Trans Syst, Man, Cybern. Part B: Cybern*, vol. 32, no. 4, pp. 526–537, 2002.
- [12] L.X. Wang, and J.M. Mendel. "Generating fuzzy rules by learning from examples", *IEEE Trans Syst, Man, Cybernetics*, vol. 22, no. 6, pp. 1414–1427, 1992.
- [13] M.J. Del Jesus, P. González, F. Herrera, and M. Mesonero, "Evolutionary fuzzy rule induction process for subgroup discovery: a case study in marketing", *IEEE Trans Fuzzy Syst*, vol. 15, no. 4, pp.578–592, 2007.
- [14] M. Kaya, "Multi-objective genetic algorithm based approaches for mining optimized fuzzy association rules", *Soft Comput*, vol. 10, no. 7, pp. 578–586, 2006.
- [15] P. Thrift, "Fuzzy logic synthesis with genetic algorithms", in *Proc of ICGA'91 Int Conf on Genetic Algorithms*, 1991, pp 509–513
- [16] J. Casillas, O. Cordon, M.J. Del Jesus, and F. Herrera, "Genetic tuning of fuzzy rule deep structures preserving interpretability for linguistic modelling", *IEEE Trans Fuzzy Syst*, vol. 13, no. 1, pp. 13–29, 2005.
- [17] C. Karr, "Genetic algorithms for fuzzy controllers", *AI Expert* vol. 6, no. 2, pp. 26–33, 1991.
- [18] W. H. Au, K. C. C. Chan, and A. K. C. Wong, "A fuzzy approach to partitioning continuous attributes for classification", *IEEE Trans Knowl Data Eng*, vol. 18, no. 5, pp. 715–719, 2006
- [19] O. Cordon, F. Herrera, and P. Villar, "Analysis and guidelines to obtain a good fuzzy partition granularity for fuzzy rule-based systems using simulated annealing", *Int J Approx Reason*, vol. 25, no. 3, pp. 187–215, 2000.
- [20] J. Alcalá-Fdez, F. Herrera, F. Márquez, and A. Peregrín, "Increasing fuzzy rules cooperation based on evolutionary adaptive inference systems", *Int J Intell Syst*, vol. 22 no. 9, pp. 1035–1064, 2007.
- [21] K. A. Crockett, Z. Bandar, J. Fowdar, J. O'Shea, "Genetic tuning of fuzzy inference within fuzzy classifier systems", *Expert Syst Appl*, vol. 23, pp. 63–82, 2006.
- [22] K. A. Crockett, Z. Bandar, and D. Mclean, "On the optimization of T-norm parameters within fuzzy decision trees" in *Proc of Fuzz-IEEE'07 Int Conf on Fuzzy Syst*, London, 2007, pp. 103–108.
- [23] D. Kim, Y. Choi, and S. Lee, "An accurate COG defuzzifier design using Lamarckian co-adaptation of learning and evolution", *Fuzzy Sets and Syst*, vol. 130, no. 2, pp. 207–225, 2002.
- [24] A. Homaifar, and E. McCormick, "Simultaneous design of membership functions and rule sets for fuzzy controllers using genetic algorithms". *IEEE Trans Fuzzy Syst*, vol. 3, no. 2, pp. 129–139, 1995.
- [25] F. A. Márquez, A. Peregrín, and F. Herrera, "Cooperative evolutionary learning of linguistic fuzzy rules and parametric aggregation connectors for Mamdani fuzzy systems", *IEEE Trans Fuzzy Syst*, vol. 15, no. 6, pp. 1168–1178, 2007.
- [26] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. "A fast and elitist multiobjective genetic algorithm: NSGA-II". *IEEE Trans Evolut Comput*, vol. 6, no 2, pp. 182–197, 2002.
- [27] R. Fullér, *Introduction to Neuro-Fuzzy Systems*. Heidelberg, Germany: Springer-Verlag, 2000.
- [28] D. Nauck, F. Klawonn, and R. Kruse, *Foundations of Neuro-Fuzzy Systems*. New York: Wiley, 1997.
- [29] O. Cordon, F. Herrera, F. Hoffmann, and L. Magdalena, *Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*. Singapore: World Scientific, 2001.
- [30] O. Cordon and F. Herrera, "A proposal for improving the accuracy of linguistic modeling", *IEEE Trans Fuzzy Syst*, vol. 8, pp. 335–344, 2000.
- [31] H. Ishibuchi, K. Nozaki, H. Tanaka, Y. Hosaka, and M. Matsuda, "Empirical study on learning in fuzzy systems by rice taste analysis", *Fuzzy Sets and Syst*, vol. 64, pp. 129–144, 1994.
- [32] K. Nozaki, H. Ishibuchi, and H. Tanaka, "A simple but powerful heuristic method for generating fuzzy rules from numerical data", *Fuzzy Sets Syst*, vol. 86, no. 3, pp. 251–270, 1997.
- [33] O. Cordon, F. Herrera, and L. Sánchez., "Solving electrical distribution problems using hybrid evolutionary data analysis techniques", *Appl Intell*, vol. 10, pp. 5–24, 1999.
- [34] M. Erdem Kurul and E. Tuzun. Available:www.wunderground.com; <http://www.wunderground.com/global/stations/17128>.
- [35] J. Demšar, Statistical comparisons of classifiers over multiple data sets, *J Mach Learn Res*, vol. 7, pp. 1–30, 2006.
- [36] D.J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, CRC Press, Boca Raton, FL, 2003.