International Journal of Uncertainty,
Fuzziness and Knowledge-Based Systems
Vol. 15, No. 5 (2007) 539–557

World Scientific
www.worldscientific.com

A MULTI-OBJECTIVE GENETIC ALGORITHM FOR TUNING AND RULE SELECTION TO OBTAIN ACCURATE AND COMPACT LINGUISTIC FUZZY RULE-BASED SYSTEMS*

R. ALCALÁ*, M. J. GACTO† and F. HERRERA‡

Dept. Computer Science and A.I., University of Granada,
Granada, E-18071, Spain,
** alcala@decsai.ugr.es*
† herrera@decsai.ugr.es
‡ mjgacto@ugr.es

J. ALCALÁ-FDEZ

Dept. Computer Science, University of Jaen, Jaen, E-23071, Spain
jalcala@ujaen.es

Received 2 January 2007
Revised 18 August 2007

This work proposes the application of Multi-Objective Genetic Algorithms to obtain Fuzzy Rule-Based Systems with a better trade-off between interpretability and accuracy in linguistic fuzzy modelling problems. To do that, we present a new post-processing method that by considering selection of rules together with tuning of membership functions gets solutions only in the Pareto zone with the highest accuracy, i.e., containing solutions with the least number of possible rules but still presenting high accuracy. This method is based on the well-known SPEA2 algorithm, applying appropriate genetic operators and including some modifications to concentrate the search in the desired Pareto zone.

Keywords: Multi-Objective Genetic Algorithms; Linguistic Modelling; Interpretability-Accuracy Trade-Off; Rule Selection; Tuning of Membership Functions.

1. Introduction

One of the aims in focusing the research in the Linguistic Fuzzy Modelling area in recent years is the trade-off between interpretability and accuracy.[1] Of course, the ideal thing would be to satisfy both criteria to a high degree, but since they are contradictory issues generally it is not possible.

*Supported by the Spanish Project TIN-2005-08386-C05-01.

A widely-used approach to improve the accuracy of linguistic Fuzzy Rule-Based Systems (FRBSs) is the *tuning* of Membership Functions (MFs),[1–10] which refines a previous definition of the Data Base (DB) once the rule base has been obtained. Although tuning usually improves the system performance, sometimes a large number of rules is used to reach an acceptable degree of accuracy. In this case, some works[1,5] consider the selection of rules together with the tuning of MFs but only considering performance criteria.

In this contribution, we focus on this problem by using Genetic Algorithms as a tool for evolving the MFs parameters and rule base size and by coding all of them (rules and parameters) in the same chromosome. Since the problem presents multi-objective nature we could consider the use of Multi-Objective Genetic Algorithms (MOGAs)[11–15] to obtain a set of solutions with different degrees of accuracy and number of rules by using both characteristics as objectives.

Although there are some works in the literature using MOGAs to improve the difficult trade-off between interpretability and accuracy of FRBSs,[16–25] practically all these works were applied to classification problems trying to obtain the complete Pareto (set of non-dominated solutions with different trade-off) by selecting or learning the set of rules better representing the example data, i.e., improving the system classification ability and decreasing the system complexity but not considering learning or tuning of the fuzzy system parameters (which involves another type of Pareto front, a more complicated search space and therefore needs different considerations respect to the works in the existing literature).

In this way, our main interest is to design an appropriate MOGA for this type of problem due to the fact that standard MOGAs can present some problems. As said, MOGAs are generally based on obtaining a set of non-dominated solutions. However, in this case, there are solutions that are not interesting although they are in the Pareto frontier. For example, non-dominated solutions with a small number of rules and high error are not interesting since they have not the desired trade-off between accuracy and interpretability. Furthermore, the existence of these kinds of solutions favours the selection of solutions with very different number of rules and accuracy to apply the crossover operator, which gives results with poor accuracy (the tuning parameters would be very different and the crossover would not have any sense except for exploring new combinations of rules).

In our proposal, we concentrate the search in the Pareto zone with still accurate solutions trying to obtain the least number of possible rules. To do that, we propose a modification of the well-known SPEA2[26] (Strength Pareto Evolutionary Algorithm 2) that considering the rule selection together with the tuning of MFs concentrates the search in the Pareto zone having accurate solutions with the least number of possible rules, the Accuracy-Oriented SPEA2 (SPEA2$_{ACC}$). Besides, we have performed the same modification and experiments with NSGA-II[27] (Nondominated Sorting Genetic Algorithm II), showing that this approach is not the most adequate for this problem.

This paper is arranged as follows. First, a brief summary of different proposals to improve the balance between interpretability and accuracy is presented, specially taking into account those considering MOGAs for this purpose. In section 3, we present a study of the estimated Pareto frontier for this problem (tuning and rule selection). SPEA2$_{ACC}$ algorithm is introduced in Section 4 together with the modifications proposed on SPEA2 and the genetic operators considered. Section 5 shows an experimental study of the proposed methods in a real-world problem. Finally, Section 6 gives some conclusions.

2. Interpretability-Accuracy Trade-off of FRBSs

Fuzzy Modelling (FM) usually comes with two contradictory requirements to the obtained model: the *interpretability*, capability to express the behaviour of the real system in an understandable way, and the *accuracy*, capability to faithfully represent the real system. Since they are contradictory issues, more priority has generally been given to one of them (defined by the problem nature), leaving the other one in the background. Two FM approaches arise depending on the main objective to be considered:

- *Linguistic FM*, mainly developed by means of linguistic (or Mamdani) FRBSs,[28,29] which is focused on the interpretability.
- *Precise FM*, mainly developed by means of Takagi-Sugeno FRBSs,[30] which is focused on the accuracy.

Regardless of the approach, a common scheme has been considered to attain the desired balance between interpretability and accuracy (Figure 1 graphically shows this operation mode):

(1) Firstly, the main objective (interpretability or accuracy) is tackled defining a specific model structure to be used, thus setting the FM approach.

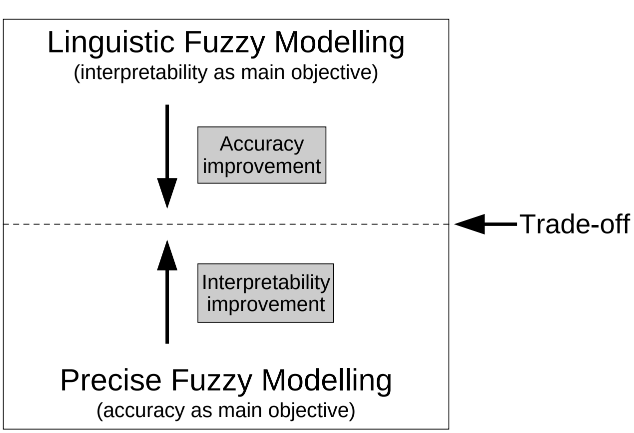

Fig. 1. Improvements of interpretability and accuracy in fuzzy modelling.

(2) Then, the modelling components (the model structure and/or the modelling process) are improved by means of different mechanisms to compensate for the initial difference between both requirements. Thus, accuracy improvements are proposed in linguistic FM at the cost of part of the interpretability whilst interpretability improvements are proposed in precise FM at the cost of part of the accuracy.

Actually, the interpretability-accuracy trade-off is a very important branch of research nowadays.[1,31] Focusing on *Linguistic FM with improved accuracy*[1] (still nearer of the interpretability) we can find many examples in the existing literature. This approach has been performed by learning/tuning the MFs by defining their parameters or shapes,[2–10] their types (triangular, trapezoidal, etc.),[32] or their context (defining the whole semantics),[5,33,34] learning the granularity (number of linguistic terms) of the fuzzy partitions,[33,35] or extending the model structure by using linguistic modifiers,[5,36,37] weights (importance factors for each rule),[2,38–40] or hierarchical architectures (mixing rules with different granularities),[38,42] among others. The main problem of these approaches is that although the system accuracy can be greatly improved (e.g., with a simple tuning of MFs), the original interpretability of the linguistic models is lost to some degree giving way to more complex systems or less interpretable rule structures.

Additionally, although rule base reduction[5,41,42] and input variable selection[43,44] processes improve the interpretability, they can also help to improve the accuracy when redundancy and inconsistency criteria are considered (but usually these improvements are not very significant).

Within the framework of linguistic FM (without improved accuracy), a new and most recent possibility is the use of Multi-Objective Evolutionary Algorithms (MOEAs)[11–15] to improve the difficult trade-off between interpretability and accuracy of FRBSs, considering different performance and interpretability measures as objectives.[16–25] Since this problem presents a multi-objective nature the use of these kinds of algorithms to obtain a set of solutions with different degrees of accuracy and interpretability is an interesting way to work. All of the works in this recent topic only consider quantitative measures of the system complexity (number of rules, number of characteristics in the antecedents, etc.) in order to improve the interpretability of such systems, rarely considering qualitative measures. Furthermore, we can point out that practically all these methods were applied to classification problems for rule selection or rule learning, without considering learning or tuning the MFs or more flexible rule representations, i.e., performing Linguistic FM with improved interpretability to obtain a set of solutions with different trade-offs but nearer the interpretability than the accuracy.

In this way, our main aim in this contribution will be to attain the desired balance by maintaining the improved accuracy that a tuning of MFs could give but trying to obtain more compact models by using MOGAs if it is possible, i.e., to obtain simpler and still accurate linguistic fuzzy models by also considering a tuning of the system parameters. This way to work represents a more complex search space and therefore needs a deeper analysis of the Pareto frontier and different considerations respect to the MOGAs in the existing literature.

3. Interpretability-Accuracy Pareto Frontier by Selecting Rules and Tuning Membership Functions

In this section, we present a study of the kinds of solutions we could find in the optimal Pareto frontier when the system error and the number of rules (both considered as objectives) are optimized by tuning the MFs and selecting the most promising rules. In this way, we can obtain an approximation of the optimal Pareto that can help to determine the desired Pareto zone.

Tuning of MFs usually needs an initial model with large number of rules to get an appropriate level of accuracy. Generally, to obtain a good number of initial rules, methods ensuring covering levels higher than needed are used. In this way, we could obtain rules that being needed at first could be unnecessary once the tuning is applied or rules that could impede the tuning of the remaining ones in order to obtain the global optimum in terms of the accuracy (better configuration of rules to get the minimum error after tuning of the parameters). Thus, we can find the following types of rules respect to this global optimum in the complete set of rules: *Bad Rules* (erroneous or conflicting rules) that degrade the system performance (rules that are not included in the most accurate final solution); *Redundant or Irrelevant Rules* that do not significantly improve the system performance; *Complementary Rules* that complement some others slightly improving the system performance; and *Important Rules* that should not be removed to obtain a reasonable system performance. Obviously, this is a simplification of the problem by only considering in principle the most accurate solution in order to have an idea of the shape of the optimum Pareto. On the other hand, to determine those types of rules in advance is impossible since it directly depends on each concrete configuration of rules and still more on the optimal configuration of the MF parameters for each rule configuration. Therefore, this is impossible to establish any criteria that could be used in the search process.

However, by taking into account the possible existence of these kinds of rules, different rule configurations and different tuning parameters, we can estimate the following zones in the space of the objectives:

- Zone with Bad Rules, which contains solutions with bad rules. In this zone, the Pareto front does not exist given that removing these kinds of rules would improve the accuracy and these solutions would be dominated by others.
- Zone with Redundant or Irrelevant Rules, which is comprised of solutions without bad rules but still maintaining redundant or irrelevant rules. By deleting these kinds of rules the accuracy would be practically the same.
- Zone with Complementary Rules, comprised of solutions without any bad or redundant rule. By removing these rules the accuracy would be slightly decreased.

- Zone with Important Rules, which contains solutions only comprised of essential rules. By removing these kinds of rules the accuracy is really affected.

In Figure 2, we can find an approximation of the optimal Pareto in the problem of tuning and rule selection with the double objective of simplicity and accuracy. This figure shows the different zones in the space of the objectives together with the desired Pareto zone to find solutions with good trade-off. This zone corresponds to the zone of complementary rules, i.e., we would like to delete all the possible rules but without seriously affecting the accuracy of the model finally obtained.

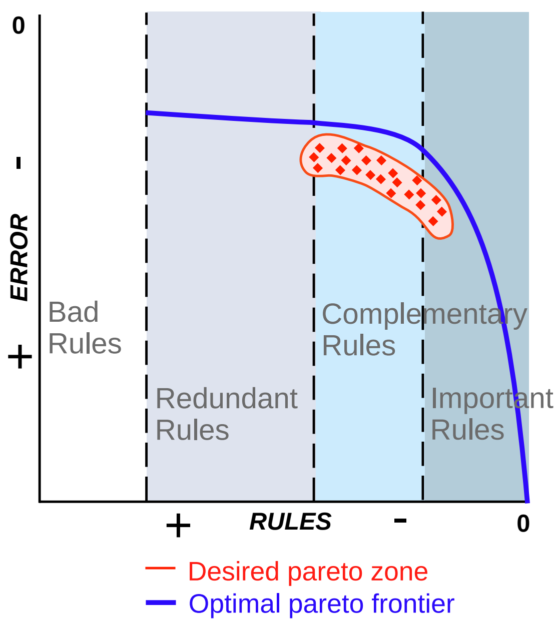

Fig. 2. Estimation of the pareto frontier considering rule selection and tuning of parameters.

Taking into account what we previously exposed, the MOGA should not obtain all the Pareto front since it is difficult to obtain accurate solutions by favouring the crossing of solutions with very different rule configurations (those in the Pareto), which try to obtain the optimum by learning very different parameters for the MFs. In the next section, we present a modification of SPEA2[26] with the main aim of guiding the search towards the desired zone.

4. A Proposal to Evolve Accuracy-Oriented Pareto Sets: the SPEA2$_{ACC}$ Algorithm

This section presents a new algorithm to get solutions with high accuracy and the least possible number of rules by performing rule selection together with a tuning of the MF parameters. In this way, since this algorithm is based on the well known SPEA2[26] we firstly introduce the basis of this algorithm. Then we describe the changes for guiding the search towards the desired Pareto zone and the main components needed to apply this algorithm to this specific problem: the coding scheme and the genetic operators.

4.1. *SPEA2 Basis*

The SPEA2 algorithm[26] (*Strength Pareto Evolutionary Algorithm 2 for multi-objetive optimization*) is one of the most used techniques for solving problems with multi-objective nature. This algorithm was designed to overcome the problems of its predecessor, the SPEA algorithm.[15] In contrast with SPEA, SPEA2: (1) incorporates a fine-grained fitness assignment strategy which takes into account for each individual the number of individuals that it dominates and the number of individuals by which it is dominated; (2) uses the nearest neighbour density estimation technique which guides the search more efficiently; (3) has an enhanced archive truncation method which guarantees the preservation of boundary solutions. Next, we briefly describe the complete SPEA2 algorithm.

SPEA2 uses a fixed population and archive size. The population forms the current base of possible solutions, while the archive contains the current solutions. The archive is constructed and updated by copying all non-dominated individuals in both archive and population into a temporary archive. If the size of this temporary archive differs from the desired archive size, individuals are either removed or added as necessary. Individuals are added by selecting the best dominated individuals, while the removal process uses a heuristic clustering routine in the objective space. The motivation for this is that one would like to try to ensure that the archive contents represent distinct parts of the objective space.

The fitness assignment strategy takes into account both dominating and dominated solutions for each individual. Let P_t and $\overline{P}_t$ denote the population and the archive respectively, each individual i in $P_t + \overline{P}_t$ is assigned a strength value $S(i)$, the number of solutions it dominates,

$$S(i) = \| \{ j \mid j \in P_t + \overline{P}_t \ \wedge \ i \succ j \} \| \tag{1}$$

where $\| \cdot \|$ represents the cardinality of a set, $+$ stands for multiset union and the symbol $\succ$ corresponds to the Pareto dominance relation. Based on the value of $S(i)$, a raw fitness value, $R(i)$, is given to the individual i,

$$R(i) = \sum_{j \in P_t + \overline{P}_t,\ j \succ i} S(j) \tag{2}$$

It is important to notice that fitness is to be minimized here, i.e., $R(i) = 0$ corresponds to a nondominated individual, while a high $R(i)$ value means that i is dominated by many individuals (which in turn dominate many other individuals). This scheme is illustrated in Figure 3. The final fitness value is assigned by adding a

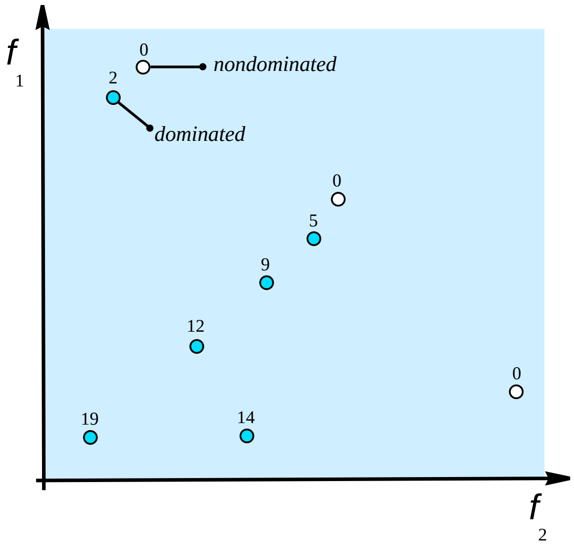

Fig. 3. The raw SPEA2 fitness values for a maximization problem with two objectives f_1 and f_2.

density value. The density function value, $D(i)$, is estimated in the objective space,

$$D(i) = \frac{1}{\delta_i^k + 2} \tag{3}$$

where δ_i^k denotes the k-th nearest distance for the ith individual among P_t and $\overline{P}_t$ in objective space. k is usually set as $\sqrt{N + \overline{N}}$ truncated to an integer, where N is the population size and $\overline{N}$ the archive size. Finally, the fitness value for the i-th individual is calculated as,

$$F(i) = R(i) + D(i) \tag{4}$$

From the definition above, a better solution will be assigned a smaller fitness value. Finally, when selecting individuals for participating in the next generation all candidates are selected from the archive using a binary tournament selection scheme.

According to the descriptions of the authors in,[26] the outline of the SPEA2 algorithm is:

Input: N (population size),
$\overline{N}$ (external population size),
T (maximum number of generations).
Output: A (non-dominated set).

(1) Generate an initial population P_0 and create the empty external population $\overline{P}_0 = \emptyset$.
(2) Calculate fitness values of individuals in P_t and $\overline{P}_t$.

(3) Copy all non-dominated individuals in $P_t \cup \overline{P}_t$ to $\overline{P}_{t+1}$. If $|\overline{P}_{t+1}| > \overline{N}$ apply truncation operator. If $|\overline{P}_{t+1}| < \overline{N}$ fill with dominated in $P_t \cup \overline{P}_t$.
(4) If $t \geq T$, return A and stop.
(5) Perform binary tournament selection with replacement on $\overline{P}_{t+1}$ in order to fill the mating pool.
(6) Apply recombination and mutation operators to the mating pool and set P_{t+1} to the resulting population. Go to step 2 with $t = t + 1$.

4.2. *The SPEA2$_{ACC}$ algorithm*

In the following, the main aspects and components needed to design the proposed algorithm are explained. They are:

- Modifications Applied on SPEA2 to guide the search.
- Coding scheme and initial gene pool.
- Objectives considered for chromosome evaluation.
- Crossover and mutation operators.

4.2.1. *Modifications applied on SPEA2*

In order to focus the search on the desired Pareto zone, high accuracy with least possible number of rules, we propose two main changes on the SPEA2 algorithm with the aim of giving more selective pressure to those solutions that have a high accuracy. The proposed changes are described next:

- A restarting operator is applied exactly at the mid of the algorithm, by maintaining the most accurate individual as the sole individual in the external population ($\overline{P}_{t+1}$ with size 1) and obtaining the remaining individuals in the population (P_{t+1}) with the same rule configuration of the best individual and tuning parameters generated at random within the corresponding variation intervals. This operation is performed in step 4 then returning to step 2 with $t = t + 1$. In this way, we concentrate the search only in the desired pareto zone (similar solutions in a zone with high accuracy).
- In each stage of the algorithm (before and after restarting), the number of solutions in the external population ($\overline{P}_{t+1}$) considered to form the mating pool is progressively reduced, by focusing only on those with the best accuracy. To do that, the solutions are sorted from the best to the worst (considering accuracy as sorting criterion) and the number of solutions considered for selection is reduced progressively from 100% at the beginning to 50% at the end of each stage.

Besides, we have to highlight that the way to create the solutions of the initial population for the part of rule selection is a very important factor. Usually, a Genetic Algorithm generates the initial population totally at random (random selection of the initial rules). However, in this case, to get solutions with a high accuracy we should not lose rules that could present a positive cooperation once their MF parameters have been evolved. The best way to do this is to start with solutions selecting all the possible rules which favours a progressive extraction of bad rules (those that do not improve with the tuning of parameters), only by means of the mutation at the beginning and then by means of the crossover.

4.2.2. *Coding scheme and initial gene pool*

A double coding scheme for both *rule selection* (C_S) and *tuning* (C_T) is used:

$$C^p = C_S^p C_T^p$$

- For the C_S part, the coding scheme consists of binary-coded strings with size m (with m being the number of initial rules). Depending on whether a rule is selected or not, values '1' or '0' are respectively assigned to the corresponding gene.

$$C_S^p = (c_{S1}, \ldots, c_{Sm}) \mid c_{Si} \in \{0, 1\} \ .$$

- For the C_T part, a real coding is considered, being m^i the number of labels of each of the n variables comprising the DB.

$$C_i = (a_1^i, b_1^i, c_1^i, \ldots, a_{m^i}^i, b_{m^i}^i, c_{m^i}^i), \ \ i = 1, \ldots, n \ ,$$
$$C_T^p = C_1 C_2 \ldots C_n \ .$$

The initial population is obtained in the following way:

(1) For the C_T part the initial DB is included as first individual. The remaining individuals are generated at random within the corresponding variation intervals. Such intervals are calculated from the initial DB. For each MF $C_i^j = (a^j, b^j, c^j)$, the variation intervals are calculated in the following way:

$$[I_{a^j}^l, I_{a^j}^r] = [a^j - (b^j - a^j)/2, a^j + (b^j - a^j)/2] \tag{5}$$
$$[I_{b^j}^l, I_{b^j}^r] = [b^j - (b^j - a^j)/2, b^j + (c^j - b^j)/2] \tag{6}$$
$$[I_{c^j}^l, I_{c^j}^r] = [c^j - (c^j - b^j)/2, c^j + (c^j - b^j)/2] \tag{7}$$

(2) For the C_S part all genes take value '1' in all the individuals of the initial population in order to favour a progressive extraction of the worst rules.

4.2.3. *Objectives*

Two objectives are minimized to get the desired trade-off: the number of rules (interpretability) and the Mean Squared Error (accuracy),

$$\text{MSE} = \frac{1}{2 \cdot |E|} \sum_{l=1}^{|E|} (F(x^l) - y^l)^2,$$

with $|E|$ being the size of a data set E, $F(x^l)$ being the output obtained from the FRBS decoded from the mentioned chromosome when the l-th example is considered and y^l being the known desired output. The fuzzy inference system considered to obtain $F(x^l)$ is the *centre of gravity weighted by the matching* strategy as defuzzification operator and the *minimum t-norm* as implication and conjunctive operators.

4.2.4. *Crossover and mutation operators*

The crossover operator depends on the chromosome part where it is applied:

- In the C_T part, the BLX-0.5[45] crossover is used. This operator is based on the the concept of environments (the offspring are generated around one parent). These kinds of operators present a good cooperation when they are introduced within evolutionary models forcing the convergence by pressure on the offspring. Figure 4 depicts the behaviour of this operator, which allow the offspring genes to be around a wide zone determined by both parent genes.

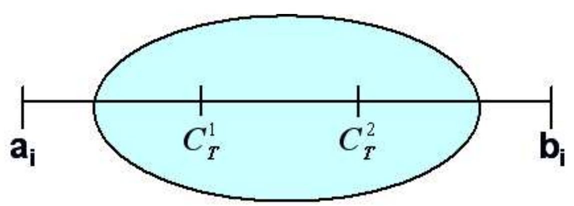

Fig. 4. Scheme of the behaviour of the BLX-α operator.

  The BLX is described as follows. Let us assume that $X = (x_1 \cdots x_n)$ and $Y = (y_1 \cdots y_n)$, $(x_i, y_i \in [a_i, b_i] \subset \Re, i = 1 \cdots n)$, are two real-coded chromosomes that are going to be crossed. The BLX operator (with $\alpha = 0.5$) generates one descendent $Z = (z_1, \cdots, z_g)$ where z_i is a randomly (uniformly) chosen number from the interval $[l_i, u_i]$, with $l_i = max\{a_i, c_{min} - I\}$, $u_i = min\{b_i, c_{max} + I\}$, $c_{min} = min\{x_i, y_i\}$, $c_{max} = max\{x_i, y_i\}$ and $I = (c_{max} - c_{min}) \cdot \alpha$.
- In the C_S part, the HUX[46] crossover is used. The HUX crossover exactly interchanges the mid of the alleles that are different in the parents (the genes to be crossed are randomly selected among those that are different in the parents). This operator ensures the maximum distance of the offspring to their parents (exploration). Figure 5 depicts the behaviour of this operator.

Finally, four offspring are generated by combining the two from the C_S part with the two from the C_T part (the two with the best accuracy are considered to be included in the population). The mutation operator changes a gene value at random in the C_S and C_T parts (one in each part) with probability P_m.

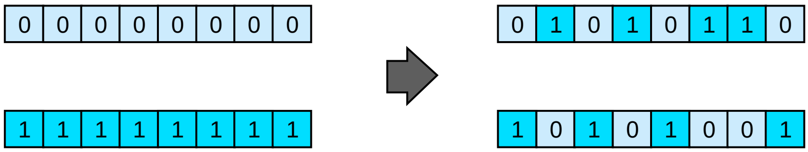

Fig. 5. Scheme of the behaviour of the HUX operator.

Table 1. Methods considered for comparison.

Ref.	Méthod	Description
47	**WM**	Wang & Mendel algorithm
5	**WM+T**	Tuning of Parameters
5	**WM+S**	Rule Selection
5	**WM+TS**	Tuning and Rule Selection
26	**SPEA2**	SPEA2 Algorithm
—	$\mathbf{SPEA2}_{ACC}$	Accuracy-Oriented SPEA2
27	**NSGAII**	NSGA-II algorithm
—	$\mathbf{NSGAII}_{ACC}$	Accuracy-Oriented NSGA-II

5. Experiments

To evaluate the usefulness of the method proposed, SPEA2$_{ACC}$, we have considered a real-world problem[49] with 4 input variables that consists of estimating the maintenance costs of medium voltage lines in a town. The methods considered for the experiments are briefly described in Table 1. WM[47] method is considered to obtain the initial rule base to be tuned. T and S methods perform the tuning of parameters and rule selection respectively. TS indicates tuning together with rule selection in the same algorithm. All of them consider the accuracy of the model as the sole objective. The remaining are MOGAs with and without the proposed modifications (all of them perform rule selection with tuning of parameters considering two objectives, accuracy and number of rules). However, we have to highlight that all of them consider the same population initialization, i.e., they start considering all the candidate rules for the initial individuals in order to see better the influence of the changes applied on the original SPEA2.

The linguistic partitions are comprised by *five linguistic terms* with triangular shape. The values of the input parameters considered by all the MOGAs studied are presented as follows: population size of 200, external population size of 61 (in the case of SPEA2 and SPEA2$_{ACC}$), 50,000 evaluations and 0.2 as mutation probability per chromosome.

5.1. *Problem description*

In Spain, electrical industries do not charge the energy bill directly to the final user, but they share the ownership of an enterprise (called R.E.E., Red Eléctrica

Table 2. Electrical problem characteristics.

Input variable X_1:	Sum of the lengths of all streets in the town
Input variable X_2:	Total area of the town
Input variable X_3:	Area that is occupied by buildings
Input variable X_4:	Energy supply to the town
Output variable Y:	Maintenance costs of the medium voltage lines
Number of examples:	1,059
Domain of X_1:	[0, 11]
Domain of X_2:	[0.15, 8.55]
Domain of X_3:	[1.64, 142.5]
Domain of X_4:	[1, 165]
Range of Y:	[0, 8546.03]

Española) which gets all payments and then distributes them according to some complex criteria (amount of power generation of every company, number of customers, etc.).

In the last years, some of these companies have asked for the rules to be revised. One of the proposed modifications involved a redistribution of the maintenance costs of the network. To compute the maintenance costs of town medium voltage lines, there is a need to know which would be the total line length if the installation made had been the optimal one. Clearly, it is impossible to obtain this value by directly measuring it, since the medium voltage lines existing in a town have been installed incrementally, according to its own electrical needs in each moment.

For this reason, the consideration of models becomes useful to compute the maintenance costs of the medium voltage electrical network in a town.[48,49] These estimations allow electrical companies to justify their expenses. Moreover, the model must be able to explain how a specific value is computed to a certain town. Our objective will be to relate the *maintenance costs of the medium voltage lines* with the following four variables: *sum of the lengths of all streets in the town*, *total area of the town*, *area that is occupied by buildings*, and *energy supply to the town*. We will deal with estimations of minimum maintenance costs based on a model of the optimal electrical network for a town in a sample of 1,059 towns. Table 2 presents a summary of the main characteristics of the problem.

To develop the different experiments, we consider a *5-folder cross-validation model*, i.e., 5 random partitions of data each with 20% (4 of them with 211 examples and one of them with 212 examples)[a], and the combination of 4 of them (80%) as training and the remaining one as test. For each one of the 5 data partitions, the tuning methods have been run 6 times, showing for each problem the average results of a total of 30 runs. In the case of methods with multi-objective approach, the averaged values are calculated considering the most accurate solution from each Pareto obtained. In this way, the multi-objective algorithms can be compared

[a]These data sets are available at: *http://decsai.ugr.es/∼casillas/fmlib/*.

Table 3. Results obtained by the studied methods.

Method	**#R**	**MSE**$_{tra}$	σ_{tra}	**t-test**	**MSE**$_{tst}$	σ_{tst}	**t-test**
WM	65	57605	2841	+	57934	4733	+
WM+T	65	18602	1211	+	22666	3386	+
WM+S	40.8	41086	1322	+	59942	4931	+
WM+TS	41.9	14987	391	+	18973	3772	+
NSGAII	41.0	14488	965	+	18419	3054	+
NSGAII$_{ACC}$	48.1	16321	1636	+	20423	3138	+
SPEA2	**33**	13272	1265	+	17533	3226	+
SPEA2$_{ACC}$	34.5	**11081**	1186	*	**14161**	2191	*

with several single objective based methods. This way to work differs from the previous works in the specialized literature, in which one or several Pareto fronts are presented and an expert should then select one solution. Our main aim following this approach is to compare the same algorithm by only considering an accuracy objective (WM+TS) with the most accurate solution found by the multi-objective ones in order to see if the Pareto fronts obtained are not only wide but also optimal (similar solutions to that obtained by WM+TS should be included in the final Pareto).

5.2. *Results and analysis*

The results obtained by the analyzed methods are shown in Table 3, where $\#R$ stands for the number of rules, MSE$_{tra}$ and MSE$_{tst}$ respectively for the average error obtained over the training and test data, σ for the standard deviation and *t-test* for the results of applying a *test t-student* (with 95 percent confidence) in order to ascertain whether differences in the performance of the proposed approach are significant when compared with that of the other algorithms in the table. The interpretation of this column is:

$\star$ represents the best average result.
\+ means that the best result has better performance than that of the corresponding row.

Analysing the results showed in Table 3 we can highlight the following facts:

- SPEA2$_{ACC}$ gets an important reduction of the mean square error respect to that obtained by the classic methods and NSGA-II. Furthermore, this algorithm improves the results obtained by SPEA2 with only 1.5 more rules.
- The models obtained by SPEA2$_{ACC}$ seem to show very good trade-off between interpretability and accuracy.

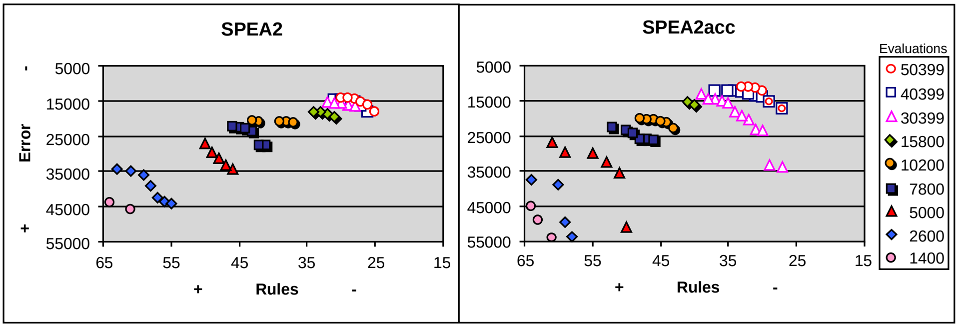

Fig. 6. Pareto fronts of $SPEA2$ and $SPEA2_{acc}$.

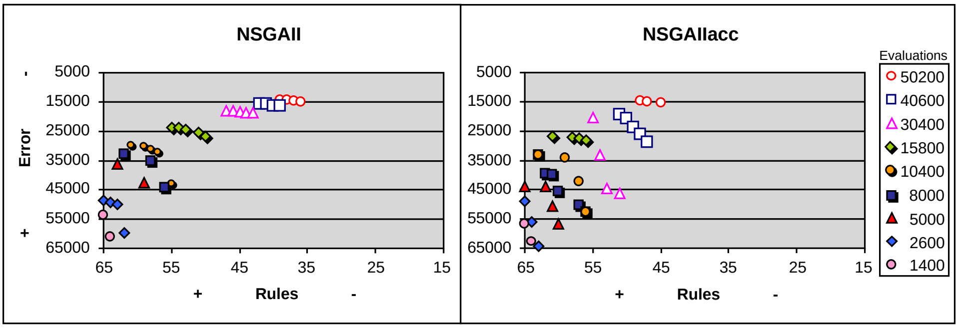

Fig. 7. Pareto fronts of $NSGAII$ and $NSGAII_{acc}$.

- NSGAII and NSGAII$_{ACC}$ present a not so good performance in this particular problem because of the crowding operator which makes very difficult to concentrate the search in the desired Pareto zone.

Moreover, notice the large search space that involves this problem. There are some initial rules that should be removed since they do not cooperate in a good way with the remaining ones. Even in the case of only considering an accuracy-based objective (WM+TS), the large search space that supposes the tuning of parameters makes very difficult to remove these kinds of rules since bad rules are tuned together with the remaining ones searching for their best cooperation. The use of a multi-objective approach favours a better selection of the ideal number of rules, preserving some rule configurations until the rule parameters are evolved to dominate solutions including bad rules.

In Figures 6 and 7, we can see the Pareto evolution for each algorithm. In figure 6, we can observe that SPEA2$_{ACC}$ mainly explores in the mid part of the evolution (before applying the restarting operator) in order to finally focusing on a specific zone of the Pareto. After restarting, the Pareto is extended in order to continue concentrating the search on the Pareto zone presenting solutions with less number of rules but still accurate.

In the remaining methods, Figures 6 and 7, we can see as the Pareto moves along without having a big extension, which does not allow to obtain very good results even in the case of NSGA-II.

6. Conclusions

Taking into account the results showed in the previous section, we can conclude that the models obtained by the proposed method present a better trade-off between interpretability and accuracy than the remaining ones. By searching for a good configuration of rules (only removing rules with little importance) and by tuning the parameter for a small set of rules, the proposed algorithm has obtained models even with a better accuracy than those obtained by methods only guided by measures of accuracy. In this way, the results obtained have shown that the use of MOEAs can represent a way to obtain even more accurate and simpler linguistic models than those obtained by only considering performance measures.

On the other hand, the proposed algorithm (SPEA2$_{ACC}$) could be of interest in problems that, although presenting a multi-objective nature, need as solution not all the Pareto frontier but only a specific area of it.

References

1. J. Casillas, O. Cordón, F. Herrera and L. Magdalena (Eds.), Accuracy improvements in linguistic fuzzy modelling, *Studies in Fuzziness and Soft Computing* **129** (Springer, Heidelberg, Germany, 2003).
2. R. Alcalá, J. Alcalá-Fdez, J. Casillas, O. Cordón and F. Herrera, Hybrid learning models to get the interpretability-accuracy trade-off in fuzzy modeling, *Soft Computing* **10:9** (2006) 717–734.
3. R. Alcalá, J. Alcalá-Fdez and F. Herrera, A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection, *IEEE Transactions on Fuzzy Systems* (2007), in press.
4. R. Alcalá, J. Alcalá-Fdez, M. J. Gacto and F. Herrera, Rule base reduction and genetic tuning of fuzzy systems based on the linguistic 3-tuples representation, *Soft Computing* **11:5** (2007) 401–419.
5. J. Casillas, O. Cordón, M. J. del Jesus and F. Herrera, Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction, *IEEE Trans. on Fuzzy Systems* **13:1** (2005) 13–29.
6. O. Cordón and F. Herrera, A three-stage evolutionary process for learning descriptive and approximate fuzzy logic controller knowledge bases from examples, *International Journal of Approximate Reasoning* **17:4** (1997) 369–407.
7. H. B. Gürocak, A genetic-algorithm-based method for tuning fuzzy-logic controllers, *Fuzzy Sets and Systems* **108:1** (1999) 39–47.
8. Y. Jin, W. von Seelen and B. Sendhoff, On generating FC^3 fuzzy rule systems from data using evolution strategies, *IEEE Transactions on Systems, Man, and Cybernetics — Part B: Cybernetics* **29:4** (1999) 829–845.
9. B. D. Liu, C. Y. Chen and J. Y. Tsao, Design of adaptive fuzzy logic controller based on linguistic-hedge concepts and genetic algorithms, *IEEE Transactions on Systems, Man, and Cybernetics — Part B: Cybernetics* **31:1** (2001) 32–53.

10. D. Nauck and R. Kruse, Neuro-fuzzy systems for function approximation, *Fuzzy Sets and Systems* **101:2** (1999) 261–271.
11. C. A. Coello, D. A. Van Veldhuizen and G. B. Lamont, *Evolutionary algorithms for solving multi-objective problems* (Kluwer Academic Publishers, New York, 2002).
12. C. Coello, Recent trends in evolutionary multiobjective optimization, in A. Abraham, L. Jain, R. Goldberg (Ed.), *Evolutionary Multiobjective Optimization: Theoretical Advances and Applications* (Springer-Verlag, London, 2005) 7–32.
13. C. Coello, G. Toscano and E. Mezura, Current and future research trends in evolutionary multiobjective optimization, in: M. Graña, R. Duro, A. d'Anjou and P. P. Wang (Eds.), *Information Processing with Evolutionary Algorithms: From Industrial Applications to Academic Speculations* (Springer-Verlag, London, 2005) 213–231.
14. C. Coello, Evolutionary multiobjective optimization: A historical view of the field, *IEEE Computational Intelligence Magazine* **1:1** (2006) 28–36.
15. E. Zitzler and L. Thiele, Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach, *IEEE Transactions on Evolutionary Computation* **3:4** (1999) 257–271.
16. O. Cordon, F. Herrera, M. J. del Jesus and P. Villar, A multiobjective genetic algorithm for feature selection and granularity learning in fuzzy-rule based classification systems, *Proceedings of IX IFSA World Congress and XX NAFIPS International Conference* (Vancouver, Canada, 2001) 1253–1258.
17. H. Ishibuchi, T. Murata and I. B. Turksen, Selecting linguistic classification rules by two-objective genetic algorithms, *Proceedings of IEEE-SMC'95* (Vancouver, Canada, 1995) 1410–1415.
18. H. Ishibuchi and T. Murata, Minimizing the fuzzy rule base and its performance by a multi-objective genetic algorithm, *Proceedings of 6th FUZZ-IEEE* (Barcelona, Spain, 1997) 259–264.
19. H. Ishibuchi, T. Murata and M. Gen, Performance evaluation of fuzzy rule-based classification systems obtained by multi-objective genetic algorithms, *Computers & Industrial Engineering* **35:3-4** (1998) 575–578.
20. H. Ishibuchi, T. Nakashima and T. Murata, Three-objective genetics-based machine learning for linguistic rule extraction, *Information Sciences* **136** (2001) 109–133.
21. H. Ishibuchi, T. Nakashima and T. Murata, Multiobjective Optimization in Linguistic Rule Extraction from Numerical, *Proceedings of First International Conference on Evolutionary Multi-Criterion Optimization, Lecture Notes in Computer Science* **1993** (Zurich, Switzerland, 2001) 588–602.
22. H. Ishibuchi and T. Yamamoto, Interpretability Issues in Fuzzy Genetics-Based Machine Learning for Linguistic Modelling, in: J. Lawry, J. G. Shanahan and A. L. Ralescu (Eds.), *Modelling with Words: Learning, Fusion, and Reasoning within a Formal Linguistic Representation Framework, Lecture Notes in Computer Science 2873* (Springer-Verlag, Berlin, Heidelberg, 2003) 209–228.
23. 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* **141:1** (2004) 59–88.
24. H. Ishibuchi and Y. Nojima, Accuracy-complexity tradeoff analysis by multiobjective rule selection, *Proceedings of ICDM 2005 Workshop on Computational Intelligence in Data Mining* (Houston, Texas, 2005) 39–48.
25. K. Narukawa, Y. Nojima and H. Ishibuchi, Modification of evolutionary multiobjective optimization algorithms for multiobjective design of fuzzy rule-based classification systems, *Proceedings of 2005 IEEE International Conference on Fuzzy Systems* (Reno, USA, 2005) 809–814.

26. E. Zitzler, M. Laumanns and L. Thiele, SPEA2: Improving the strength pareto evolutionary algorithm for multiObjetive optimization, *Proc. Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems*, Barcelona, Spain, 2001, 95–100.
27. K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. on Evolutionary Computation* **6:2** (2002) 182–197.
28. E. H. Mamdani, Applications of fuzzy algorithms for control a simple dynamic plant, *Proceedings of the IEEE* **121:12** (1974) 1585–1588.
29. E. H. Mamdani and S. Assilian, An experiment in linguistic synthesis with a fuzzy logic controller, *International Journal of Man-Machine Studies* **7** (1975) 1–13.
30. T. Takagi and M. Sugeno, Fuzzy identification of systems and its application to modeling and control, *IEEE Transactions on Systems, Man, and Cybernetics* **15** (1985) 116–132.
31. J. Casillas, O. Cordón, F. Herrera and L. Magdalena (Eds.), *Interpretability issues in fuzzy modeling*, Studies in Fuzziness and Soft Computing **128** (Springer, Heidelberg, Germany, 2003).
32. Y. Shi, R. Eberhart and Y. Chen, Implementation of evolutionary fuzzy systems, *IEEE Transactions on Fuzzy Systems* **7:2** (1999) 109–119.
33. R. Alcalá, J. Alcalá-Fdez, F. Herrera and J. Otero, Genetic learning of accurate and compact fuzzy rule based systems based on the 2-tuples linguistic representation, *International Journal of Approximate Reasoning* **44:1** (2007) 45–64.
34. W. Pedrycz, R. R. Gudwin and F. A. C. Gomide, Nonlinear context adaptation in the calibration of fuzzy sets, *Fuzzy Sets and Systems* **88:1** (1997) 91–97.
35. J. Espinosa and J. Vandewalle, Constructing fuzzy models with linguistic integrity from numerical data-AFRELI algorithm, *IEEE Transactions on Fuzzy Systems* **8:5** (2000) 591–600.
36. O. Cordón, M. J. del Jesus and F. Herrera, Genetic learning of fuzzy rule-based classification systems cooperating with fuzzy reasoning methods, *International Journal of Intelligent Systems* **13** (1998) 1025–1053.
37. A. González and R. Pérez, A study about the inclusion of linguistic hedges in a fuzzy rule learning algorithm, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **7:3** (1999) 257–266.
38. R. Alcalá, J. R. Cano, O. Cordón, F. Herrera, P. Villar and I. Zwir, Linguistic modeling with hierarchical systems of weighted linguistic rules, *International Journal of Approximate Reasoning* **32:2-3** (2003) 187–215.
39. J. S. Cho and D. J. Park, Novel fuzzy logic control based on weighting of partially inconsistent rules using neural network, *Journal of Intelligent and Fuzzy Systems* **8** (2000) 99–110.
40. N. R. Pal and K. Pal, Handling of inconsistent rules with an extended model of fuzzy reasoning, *Journal of Intelligent and Fuzzy Systems* **7** (1999) 55–73.
41. T. P. Hong and C. Y. Lee, Effect of merging order on performance of fuzzy induction, *Intelligent Data Analysis* **3:2** (1999) 139–151.
42. H. Ishibuchi, K. Nozaki, N. Yamamoto and H. Tanaka, Selecting fuzzy if-then rules for classification problems using genetic algorithms, *IEEE Transactions on Fuzzy Systems* **3:3** (1995) 260–270.
43. T. P. Hong and J. B. Chen, Finding relevant attributes and membership functions, *Fuzzy Sets and Systems* **103:3** (1999) 389–404.
44. H. M. Lee, C. M. Chen, J. M. Chen and Y. L. Jou, An efficient fuzzy classifier with feature selection based on fuzzy entropy, *IEEE Transactions on Systems, Man, and*

Cybernetics — Part B: Cybernetics **31:3** (2001) 426–432.
45. L. J. Eshelman and J. D. Schaffer, Real-coded genetic algorithms and interval-schemata, *Foundations of Genetic Algorithms* **2** (1993) 187–202.
46. L. J. Eshelman, The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination, *Foundations of Genetic Algorithms* **1** (1991) 265–283.
47. L. X. Wang and J. M. Mendel, Generating fuzzy rules by learning from examples, *IEEE Trans. on Systems, Man, and Cybernetics* **22:6** (1992) 1414–1427.
48. O. Cordón, F. Herrera and L. Sánchez, Computing the Spanish Medium Electrical Line Maintenance Costs by means of Evolution-Based Learning Processes, *Proc. International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems*, Castellón, Spain, 1998, 478–486.
49. O. Cordón, F. Herrera and L. Sánchez, Solving electrical distribution problems using hybrid evolutionary data analysis techniques, *Applied Intelligence* **10** (1999) 5–24.