5

On the Usefulness of MOEAs for Getting Compact FRBSs Under Parameter Tuning and Rule Selection*

R. Alcalá, J. Alcalá-Fdez, M.J. Gacto, F. Herrera

Department of Computer Science and Artificial Intelligence
University of Granada, E-18071 – Granada, Spain
{alcala,jalcala,herrera}@decsai.ugr.es, mjgacto@ugr.es

Summary. In the last years, multi-objective genetic algorithms have been successfully applied to obtain Fuzzy Rule-Based Systems satisfying different objectives, usually different performance measures.

Recently, multi-objective genetic algorithms have been also applied to improve the difficult trade-off between interpretability and accuracy of Fuzzy Rule-Based Systems, obtaining linguistic models not only accurate but also interpretable. It is know that both requirements are usually contradictory, however, a multi-objective genetic algorithm can obtain a set of solutions with different degrees of accuracy and interpretability.

This contribution briefly reviews the state of the art in this very recent topic and presents an approach in order to prove the ability of multi-objective genetic algorithms for getting compact fuzzy rule-based systems under rule selection and parameter tuning, i.e., to obtain linguistic models with improved accuracy and the least number of possible rules. This way to work involves another trade-off degree respect with the works in the existing literature which has been still not explored.

5.1 Introduction

One of the most important areas for the application of Fuzzy Set Theory are Fuzzy Rule-Based Systems (FRBSs). Typically, they have been considered to solve problems in different application domains as classification, regression or control and rule mining (26). There are at least two different kinds of FRBSs in the literature, Mamdani (33) and Takagi-Sugeno (36), which differ on the composition of the rule consequent. The use of one or another depends on the fact that the main requirement is the interpretability or the accuracy of the model, respectively.

Many automatic techniques have been proposed in the literature to extract a proper set of fuzzy rules from numerical data. However, most of these techniques usually try to improve the performance associated to the prediction error without

* Supported by the Spanish CICYT Project TIN2005-08386-C05-01 (KEEL II).

R. Alcalá et al.: *On the Usefulness of MOEAs for Getting Compact FRBSs Under Parameter Tuning and Rule Selection*, Studies in Computational Intelligence (SCI) **98**, 91–107 (2008)
www.springerlink.com

inclusion of any interpretability measure, an essential aspect of FRBSs. In the last years, the problem of finding the right trade-off between interpretability and accuracy, in spite of the original nature of fuzzy logic, has arisen a growing interest in methods which take both aspects into account (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.

Recently, Multi-Objective Evolutionary Algorithms (MOEAs) (5; 12) have been also applied to improve the difficult trade-off between interpretability and accuracy of FRBSs, obtaining linguistic models not only accurate but also interpretable. 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. In this way, we can obtain a set of solutions where each solution tends to satisfy a criterion to a higher extent than another, and considering different performance and interpretability measures an expert can select those solutions satisfying these objectives (accuracy and interpretability measures) to the desired degree.

Most of these works apply MOEAs to obtain accurate but also interpretable Mamdani FRBSs (8; 19; 22; 23; 24; 25; 27; 32) since they are much more interpretable than Takagi-Sugeno ones (29; 30; 38; 39). This contribution briefly reviews the state of the art in this recent topic, analyzing the most representative works of the specialized literature in order to point out the most important aspects that should been taken into account to deal with these kinds of problems. All these works try 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 accuracy and decreasing the FRBS complexity but not considering learning or tuning of the membership function parameters. In this way, this work also presents an approach in order to study the usefulness of multi-objective genetic algorithms to obtain simpler and still accurate linguistic fuzzy models by applying rule selection and a tuning of membership functions, which represents a more complex search space and therefore needs of different considerations respect to the works in the existing literature. Two different MOEAs are considered and studied to perform this task.

This contribution is arranged as follows. The next section analyzes the state of the art on the use of MOEAs to get a better trade-off between interpretability and accuracy of FRBSs in different application areas. In Section 5.3, we present two different algorithms by considering two of the most known MOEAs in order to obtain linguistic models by applying rule selection together with a tuning of parameters. Section 5.4 shows an experimental study of these methods applied to a complex but interesting problem. Finally, Section 5.5 points out some conclusions and further research lines.

5.2 Use of MOEAs to Get the Interpretability-accuracy Trade-off of Fuzzy Systems

Evolutionary algorithms deal simultaneously with a set of possible solutions (the so-called population) which allows us to find several members of the Pareto optimal set in a single run of the algorithm. Additionally, evolutionary algorithms are less susceptible to the shape or continuity of the Pareto front (e.g., they can easily deal with discontinuous and concave Pareto fronts) (5; 6; 12; 41). In this way, MOEAs have proved to be very effective to search for optimal solutions to problems that incorporate multiple performance criteria in competition with each other (41; 42). These kinds of algorithms generate a family of equally valid solutions, where each solution tends to satisfy a criterion to a higher extent than another.

After the first generation of MOEAs for general use (17; 18; 35) (MOGA, NPGA and NSGA respectively) a second generation of MOEAs (3; 4; 9; 10; 11; 13; 14; 31; 40; 41; 43) started when elitism became a standard mechanism (34). The Strength Pareto Evolutionary Algorithm 2 (SPEA2) (43) and the Nondominated Sorting Genetic Algorithm II (NSGA-II) (11; 13) can be considered as the most representative MOEAs of the second generation. However, nowadays NSGA-II is the paradigm within the MOEA research community since the powerful crowding operator that this algorithm incorporates usually allows to obtain the widest Pareto sets in a great variety of problems, and this is a very appreciated property in this framework.

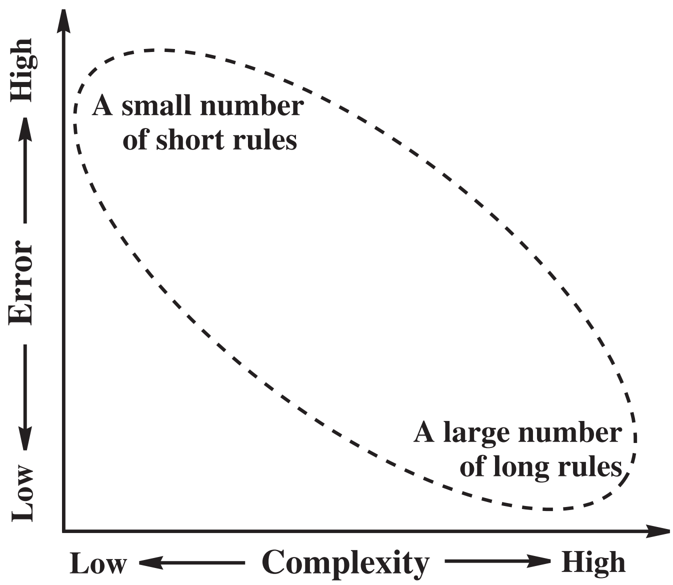

Fig. 5.1. Trade-off between the error and the interpretability of rule sets

MOEAs have been also applied to improve the difficult trade-off between interpretability and accuracy of FRBSs, where each solution in the pareto front represents a different trade-off between interpretability and accuracy (see Figure 5.1). The most continuous and prolific research activity in the application of MOEAs to Mamdani FRBS generation for finding the accuracy-interpretability trade off has been certainly performed by Ishibuchi's group. Since they suggested the idea of a MOEA to find a set of solutions with different trade off by using rule selection and two objectives (19), they have published many works in this interesting field. Earlier works (19; 21; 22) were devoted to the application of simple MOEAs of the first generation to perform a rule selection on an initial set of classification rules involving *"don't care"* conditions and considering two different objectives (classification accuracy and number of rules). Then, a third objective was also included in order to minimize the length of the rules by rule selection (23; 24) or rule learning (23). In (27), they apply a better MOEA, the Multi-Objective Genetic Local Search (20) (MOGLS), following the same approach for rule selection with three objectives. And finally, two algorithms based on a MOEA of the second generation (NSGA-II) have been proposed respectively for rule selection (32) and rule learning (28) considering the same concepts. In the literature, we can also find some papers of other researchers in this topic. For instance in (8), Cordon et al. use MOGA for jointly performing feature selection and fuzzy set granularity learning with only two objectives.

At this point, we can see that all the methods mentioned were applied to classification problems for rule selection or rule learning, without learning or tuning the membership functions that were initially fixed. Most of the works in this topic only consider quantitative measures of the system complexity in order to improve the interpretability of such systems, rarely considering qualitative measures. Moreover, MOEAs considered were slight modifications of MOEAs proposed for general use (MOGA, NSGA-II, etc.) or based on them. Notice that, although NSGA-II improves the results respect to other MOEAs, since to cross non-dominated rule sets with very different numbers of rules and different rule structures (forced by the NSGA-II crowding operator) usually gives a bad accuracy, this MOGA needed of an adaptation to favor the cross of similar solutions in order to also get good results for the accuracy objective (32). The problem is that, although to directly apply this powerful algorithm to this problem gets a wider Pareto front with several good solutions, to improve the accuracy objective is more difficult than simplifying the fuzzy models, by which the Pareto front finally obtained still becomes sub-optimal respect to the accuracy objective.

On the other hand, there are a few works in the framework of fuzzy modeling. In (25), authors show how a simple MOGA can be applied to a three-objective optimization problem (again not considering learning or tuning of parameters). Some applications of MOEAs have been also discussed in the literature to improve the difficult trade-off between accuracy and interpretability of Takagi-Sugeno models. In (29; 30; 38; 39), accuracy, interpretability and compactness have been considered as objectives to obtain interpretable and very accurate Takagi-Sugeno models. However, since Takagi-Sugeno models have a linear function in the consequent part of each fuzzy rule, they are close to accuracy representing another type of trade-off with less interpretable models (25). For this reason, the type of rule most used to achieve the trade-off between accuracy and complexity are the fuzzy rules with linguistic terms in both the antecedent and consequent parts, i.e., Mamdani rules (33).

Table 5.1 presents a resume of the different methods in the existing literature.

Table 5.1. Use of MOEAs for finding the accuracy-interpretability trade off

Year	Ref.	Problem	MOEA/Gen.	#Objs.	RS	FS	RL	LP
MAMDANI LINGUISTIC MODELS (closer to interpretability)								
1995/7/8	(19; 21; 22)	Classification	MOGA/1st	2	√	√	-	-
2001	(8)	Classification	MOGA/1st	2	-	√	√	-
2001	(23; 24)	Classification	MOGA/1st	3	√	√	√	-
2003	(25)	Regression	MOGA/1st	3	√	√	√	-
2004	(27)	Classification	MOGLS/1st	3	√	√	-	-
2005	(32)	Classification	NSGA-II*/2nd	3	√	√	-	-
2007	(28)	Classification	NSGA-II*/2nd	3	-	√	√	-
TAKAGI-SUGENO MODELS (closer to accuracy)								
2001	(29; 30)	Regression	Specific/1st	3	-	-	√	√
2005	(38)	Regression	MOGA*/1st	5	√	√	√	√
2005	(39)	Regression	NSGA-II*/2nd	5	√	√	√	√

RS = Rule Selection, **FS** = Feature Selection, **RL** = Rule Learning, **LP** = Learning/Tuning of parameters.
* based on that algorithm

5.3 Two Different MOEAs for Rule Selection and Tuning of Parameters

As we explain in the previous section most works in the field of fuzzy systems are applied to classification problems by learning or selecting rules, not considering tuning of the membership function parameters. The main reason of this fact is that a tuning of parameters implies a lost of the interpretability to some degree. However, it is known that this way to work greatly improves the performance of the linguistic models so obtained, being another alternative to improve the interpretability-accuracy trade-off. For this reason, we present two different algorithms that focus the research in the linguistic fuzzy modeling area, in order to evaluate the performance of MOEAs in a field which is still less explored, and with the main objective of inject some ideas or recommendations in this open topic (improvement of the interpretability of very accurate models).

The proposed algorithms will perform rule selection from a given fuzzy rule set together with the parametric tuning of the membership functions. To do that, we apply the most used multi-objective algorithms of the second generation, SPEA2 (43) and NSGA-II (13), considering two different objectives, system error and number of rules.

In the next subsections, we present the SPEA2 and NSGA-II algorithms applied for linguistic fuzzy modeling. At first, the common components of both algorithms are proposed and then the main steps and characteristic of both algorithms are described.

5.3.1 Main Components of the Algorithms

As mentioned, we use the well-known SPEA2 and NSGA-II to perform rule selection and tuning of membership functions and with the aim of improving the desired trade-off between interpretability and accuracy. In the following, the components needed to apply these algorithms in this concrete problem are explained. They are coding scheme, initial gene pool, objectives and genetic operators:

- **Coding scheme and initial gene pool**
  A double coding scheme for both *rule selection* (C_S) and *tuning* (C_T) is used:

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

  In the $C_S^p = (c_{S1}, \ldots, c_{Sm})$ 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. In the C_T part, a real coding is considered, being m^i the number of labels of each of the n variables comprising the database,

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

  The initial population is obtained with all individuals having all genes with value '1' in the C_S part. And in the C_T part the initial database is included as first individual. The remaining individuals are generated at random within the corresponding variation intervals. Such intervals are calculated from the initial database. For each membership function $C_i^j = (a^j, b^j, c^j)$, the variation intervals are calculated in the following way:

  $$\begin{array}{r} [I_{a^j}^l, I_{a^j}^r] = [a^j - (b^j - a^j)/2, a^j + (b^j - a^j)/2] \\ [I_{b^j}^l, I_{b^j}^r] = [b^j - (b^j - a^j)/2, b^j + (c^j - b^j)/2] \\ [I_{c^j}^l, I_{c^j}^r] = [c^j - (c^j - b^j)/2, c^j + (c^j - b^j)/2] \end{array}$$

- **Objectives**
  Two objectives are minimized for this problem: the number of rules (interpretability) and the Mean Squared Error (accuracy),

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

with $|E|$ being the size of a data set E, $F(x^l)$ being the output obtained from the FRBS decoded from the said 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 *center of gravity weighted by the matching* strategy as defuzzification operator and the *minimum t-norm* as implication and conjunctive operators.

- **Genetic Operators**
  The crossover operator depends on the chromosome part where it is applied: the BLX-0.5 (16) in the C_T part and the HUX (15) in the C_S part:
  - In the C_T part, the BLX-0.5 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.
    The BLX operator can be described as follows. To ease the description 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 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).

  Finally, four offspring are generated by combining the two from the C_S part with the two from the C_T part (the two best replace to their parent). Moreover, 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.

5.3.2 SPEA2 Based Approach

The SPEA2 algorithm (43) was designed to overcome the problems of its predecessor, the SPEA algorithm (41). 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. Finally, when selecting individuals for participating in the next generation all candidates are selected from the archive using a binary tournament selection scheme.

Considering the components defined and the descriptions of the authors in (43), the SPEA2 algorithm consists of the next steps:

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 (BLX-HUX) and mutation operators to the mating pool and set P_{t+1} to the resulting population. Go to step 2 with $t = t + 1$.

5.3.3 NSGA-II Based Approach

The NSGA-II algorithm (11; 13) is one of the most well-known and frequently-used MOEAs in the literature. As in other evolutionary algorithms, first NSGA-II generates an initial population. Then an offspring population is generated from the current population by selection, crossover and mutation. The next population is constructed from the current and offspring populations. The generation of an offspring population and the construction of the next population are iterated until a stopping condition is satisfied. The NSGA-II algorithm has two features, which make it a high-performance MOEA. One is the fitness evaluation of each solution based on Pareto ranking and a crowding measure, and the other is an elitist generation update procedure.

Each solution in the current population is evaluated in the following manner. First, Rank 1 is assigned to all non-dominated solutions in the current population. All solutions with Rank 1 are tentatively removed from the current population. Next, Rank 2 is assigned to all non-dominated solutions in the reduced current population. All solutions with Rank 2 are tentatively removed from the reduced current population. This procedure is iterated until all solutions are tentatively removed from the current population (i.e., until ranks are assigned to all solutions). As a result, a different rank is assigned to each solution. Solutions with smaller ranks are viewed as being better than those with larger ranks. Among solutions with the same rank, an additional criterion called a crowding measure is taken into account.

The crowding measure for a solution calculates the distance between its adjacent solutions with the same rank in the objective space. Less crowded solutions with larger values of the crowding measure are viewed as being better than more crowded solutions with smaller values of the crowding measure.

A pair of parent solutions are selected from the current population by binary tournament selection based on the Pareto ranking and the crowding measure. When the next population is to be constructed, the current and offspring populations are combined into a merged population. Each solution in the merged population is evaluated in the same manner as in the selection phase of parent solutions using the Pareto ranking and the crowding measure. The next population is constructed by choosing a specified number (i.e., population size) of the best solutions from the merged population. Elitism is implemented in NSGA-II algorithm in this manner.

Considering the components previously defined and the descriptions of the authors in (13), NSGA-II consists of the next steps:

1. A combined population R_t is formed with the initial parent population P_t and offspring population Q_t (initially empty).
2. Generate all non-dominated fronts $F = (F_1, F_2, ...)$ of R_t.
3. Initialize $P_{t+1} = 0$ and $i = 1$.
4. Repeat until the parent population is filled.
5. Calculate crowding-distance in F_i.
6. Include i-th non-dominated front in the parent population.
7. Check the next front for inclusion.
8. Sort in descending order using crowded-comparison operator.
9. Choose the first $(N - |P_{t+1}|)$ elements of F_i.
10. Use selection, crossover (BLX-HUX) and mutation to create a new population Q_{t+1}.
11. Increment the generation counter.

5.4 Experiments: A Case Study on Linguistic Fuzzy Modeling

In this section, we present an example on the use of MOGAs to obtain linguistic models with a good trade-off between interpretability and accuracy in a real-world problem (7) 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 5.2.

WM 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 proposed algorithms (NSGA-II and SPEA2) perform rule selection from a given fuzzy rule set together with the

Table 5.2. Methods considered for comparison

Méthod	Ref.	Description
WM	(37)	Wang & Mendel algorithm
WM+T	(2)	Tuning of Parameters
WM+S	(2)	Rule Selection
WM+TS	(2)	Tuning and Rule Selection
SPEA2	-	Tuning and Rule Selection with SPEA2
NSGA-II	-	Tuning and Rule Selection with NSGA-II

parametric tuning of the membership functions considering two objectives, system error and number of rules.

In the next subsections, we describe this real-world problem and finally we show the results obtained.

5.4.1 Problem Description and Experiments

Estimating the maintenance costs of the medium voltage electrical network in a town (7) is a complex but interesting problem. Since a direct measure is very difficult to obtain, it is useful to consider models. These estimations allow electrical companies to justify their expenses. Moreover, the model must be able to explain how a specific value is computed for 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.

To develop the different experiments, we consider a *5-folder cross-validation model*, i.e., 5 random partitions of data each with 20%, 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 averaged results of a total of 30 runs. In the case of methods with multi-objective approach (SPEA2 and NSGA-II), the averaged values are calculated considering the most accurate solution from each Pareto obtained. In this way, the multi-objective algorithms are compared with several single objective based methods. This way to work differs with the previous works in the specialized literature in which one or several Pareto fronts are presented and an expert should after 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).

The initial linguistic partitions are comprised by *five linguistic terms* with equally distributed triangular shape membership functions. Figure 5.2 shows an example of this type of partition.

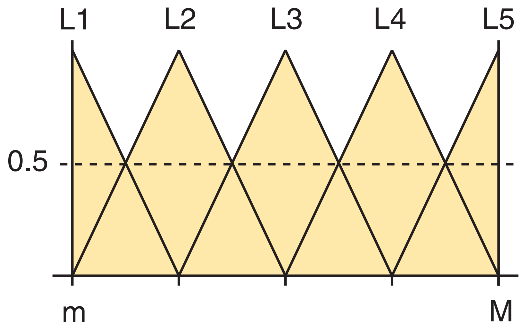

Fig. 5.2. Example of linguistic partition with five linguistic terms

The values of the input parameters considered by MOEAs are shown in the next: population size of 200, external population size of 61 (in the case of SPEA2), 50000 evaluations and 0.2 as mutation probability per chromosome.

5.4.2 Results and Analysis

The results obtained by the analyzed methods are shown in Table 5.3, where $\#R$ stands for the number of rules, MSE$_{tra}$ and MSE$_{tst}$ respectively for the averaged 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 multi-objective approach are significant when compared with that of the other algorithms in the table. The interpretation of this column is:

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

Table 5.3. 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	=
SPEA2	**33**	**13272**	1265	*	**17533**	3226	*
NSGA-II	41.0	14488	965	=	18419	3054	=

Analyzing the results showed in Table 5.3 we can highlight the two following facts:

- NSGA-II obtains the same accuracy and the same number of rules than the models obtained with WM+TS (single objective-based approach) considering the most accurate result of each obtained Pareto. Therefore, we could consider that this algorithm gets good solutions, from the most accurate ones (with more complexity) to the most simple ones (with the worst accuracy).
- The SPEA2 method shows a reduction of the MSE_{tra} and produces more or less the same MSE_{tst} respect to the models obtained by only considering the accuracy objective (WM+TS). Moreover, a considerable number of rules have been removed from the initial FRBS, obtaining simpler models with a similar performance. In this way, the most accurate models obtained by SPEA2 considering a multi-objective approach get a better trade-off between interpretability and accuracy than those obtained by a single objective based algorithm (which theoretically should obtain the most accurate results).

These results are due to 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, 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 favors 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.

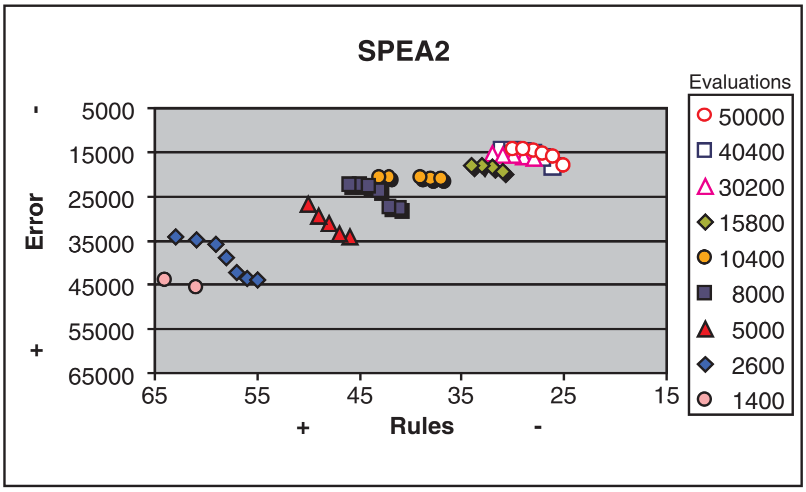

Fig. 5.3. Evolution of the Pareto fronts of SPEA2

On the other hand, NSGA-II tries to obtain a wider Pareto front by crossing very different solutions based on its crowding operator. However, in this problem, it is difficult to obtain accurate solutions by favoring 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 membership functions. This is the reason by which this algorithm does not work as well as SPEA2 in this particular problem.

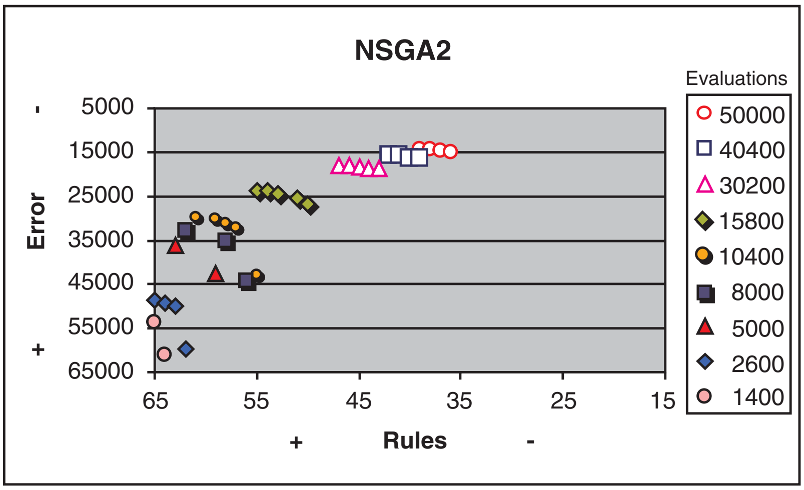

Fig. 5.4. Evolution of the Pareto fronts of NSGA2

All these facts suggest the design of more specific algorithms in order to get even better solutions for these kinds of problems, probably, solutions with a better performance that considering a single objective and with a minor number of rules. In Figures 5.3 and 5.4, we can see the Pareto evolution for each multi-objective algorithm. We can observe as the Pareto moves along without having a wide extension, even in the case of NSGA-II.

5.5 Concluding Remarks

In this work we have analyzed the use of MOEAs to improve the trade-off between interpretability and accuracy of FRBSs. A brief revision of the state of the art in this topic has been performed. From this study we can point out the following facts:

- Most of the contributions in this topic were made in the framework of fuzzy classification, considering Mamdani FRBSs.
- Most of the works only consider quantitative measures of the system complexity to determine the FRBS interpretability.
- None of the works considered a learning or tuning of the membership functions, only performing a rule learning or selection.

- The MOEAs considered were slight modifications of MOEAs proposed for general use (MOGA, NSGA-II, etc.) or adaptations of them for this concrete and difficult problem. It is due to the special nature of this problem, in which to improve the accuracy objective is more difficult than simplifying the fuzzy models, by which the Pareto front finally obtained still becomes sub-optimal respect to the accuracy objective. This specially occurs in algorithms as NSGA-II (32), since to cross non-dominated rule sets with very different numbers of rules and different rule structures (forced by the NSGA-II crowding operator) usually gives a bad accuracy, by which this MOEA needs of an adaptation to favor the cross of similar solutions in order to also get good results for the accuracy objective (28; 32).

On the other hand, this contribution has presented two different algorithms and a case of study on the use of multi-objective genetic algorithms to obtain simpler and still accurate linguistic fuzzy models by also considering a tuning of the system parameters, which represents a more complex search space and therefore needs of different considerations respect to the works in the existing literature. 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. In this case (also performing a tuning of the parameters), the problem of crossing very different solutions with different number of rules and very different parameters becomes more important since to obtain a wide Pareto with the best solutions is practically impossible. Therefore, as further work, more specific algorithms should be proposed in order to get the best possible solutions.

References

[1] Casillas J, Cordón O, Herrera F, Magdalena L (2003) (eds), Interpretability issues in fuzzy modeling. Studies in Fuzziness and Soft Computing 129, Springer, Heidelberg, Germany

[2] Casillas J, Cordón O, del Jesus M J, Herrera F (2005) Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction. IEEE Transactions on Fuzzy Systems 13(1):13–29

[3] Coello Coello C A, Toscano G (2001) A micro-genetic algorithm for multiobjective optimization. First International Conference on Evolutionary Multi-Criterion Optimization, Lecture Notes in Computer Science 1993:126–140

[4] Coello Coello C A, Toscano G (2001) Multiobjective optimization using a micro-genetic algorithm. Proceedings of the Genetic and Evolutionary Computation Conference (San Francisco, California), 274–282

[5] Coello Coello C A, Van Veldhuizen D A, Lamont G B (2002) Evolutionary algorithms for solving multi-objective problems. Kluwer Academic Publishers

[6] Coello Coello C A (2006) Evolutionary multiobjective optimization: A historical view of the field. IEEE Computational Intelligence Magazine 1(1):28–36

[7] Cordón O, Herrera F, Sánchez L (1999), Solving electrical distribution problems using hybrid evolutionary data analysis techniques. Applied Intelligence 10:5–24

[8] Cordon O, Herrera F, del Jesus M J, Villar P (2001) 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, 1253–1258
[9] Corne D, Knowles J, Oates M (2000) The pareto envelope-based selection algorithm for multiobjective optimization. Proceedings of the Parallel Problem Solving from Nature VI Conference, Lecture Notes in Computer Science 1917:839–848
[10] Corne D, Jerram N, Knowles J, Oates M (2001) PESA-II: Region based selection in evolutionary multiobjective optimization. Proceedings of the Genetic and Evolutionary Computation Conference, San Francisco, Califormia, 283–290
[11] Deb K, Agrawal S, Pratab A, Meyarivan T (2000) A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. Proceedings of the Parallel Problem Solving from Nature VI Conference, Lecture Notes in Computer Science 1917, Paris, France, 849–858
[12] Deb K (2001) Multi-objective optimization using evolutionary algorithms, John Wiley & Sons, Chichester, UK
[13] Deb K, Agrawal S, Pratab A, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6(2):182-197
[14] Erickson M, Mayer A, Horn J (2001) The niched Pareto genetic algorithm 2 applied to the design of groundwater remediation systems. Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization, Lecture Notes in Computer Science 1993, London, UK, 681–695
[15] Eshelman L J (1991) The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination. Foundations of genetic Algorithms 1:265–283
[16] Eshelman L J, Schaffer J D (1993) Real-coded genetic algorithms and interval-schemata. Foundations of Genetic Algorithms 2:187–202
[17] Fonseca C M, Fleming P J (1993) Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization. Proceedings of the Fifth International Conference on Genetic Algorithms, San Mateo, California, 416–423
[18] Horn J, Nafpliotis N, Goldberg D E (1994) A niched pareto genetic algorithm for multiobjective optimization. Proceedings of the First IEEE Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence 1:82–87
[19] Ishibuchi H, Murata T, Turksen I B (1995) Selecting linguistic classification rules by two-objective genetic algorithms. Proceedings of IEEE-SMC'95, Vancouver, Canada, 1410–1415
[20] Ishibuchi H, Murata T (1996) Multi-objective genetic local search algorithm. Proceedings of Third IEEE International Conference Evolutionary Computation, Japan, 119-124

[21] Ishibuchi H, Murata T, Turksen I B (1997) Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems. Fuzzy Sets and Systems 89(2):135–150
[22] Ishibuchi H, Murata T, Gen M (1998) Performance evaluation of fuzzy rule-based classification systems obtained by multi-objective genetic algorithms. Computers & Industrial Engineering 35(3-4):575–578
[23] Ishibuchi H, Nakashima T, Murata T (2001) Three-objective genetics-based machine learning for linguistic rule extraction. Information Sciences 136:109–133
[24] Ishibuchi H, Nakashima T, Murata T (2001) Multiobjective optimization in linguistic rule extraction from numerical data. Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization, Lecture Notes in Computer Science 1993:588–602
[25] Ishibuchi H, Yamamoto T (2003) Interpretability issues in fuzzy genetics-based machine learning for linguistic modelling. In Lawry J, Shanahan J G, Ralescu A L (eds) Modelling with Words: Learning, Fusion, and Reasoning within a Formal Liguistic Representation Framework, Lecture Notes in Computer Science 2873, Springer-Verlag, Berlin, Heidelberg, 209–228
[26] Ishibuchi H, Nakashima T, Nii M (2004) (eds) Classification and modeling with linguistic information granules. Advanced Approaches to Linguistic Data Mining Series: Advanced Information Processing, Springer, Berlin
[27] Ishibuchi H, Yamamoto T (2004) Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining. Fuzzy Sets and Systems 141(1):59–88
[28] Ishibuchi H, Nojima Y (2007) Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning. International Journal of Approximate Reasoning 44(1):4–31
[29] Jimenez F, Gomez-Skarmeta A F, Roubos H, Babuska R (2001) Accurate, transparent, and compact fuzzy models for function approximation and dynamic modeling through multi-objective evolutionary optimization. Proceedings of First International Conference on Evolutionary Multi-Criterion Optimization, Lecture Notes in Computer Science, 1993:653–667
[30] Jimenez F, Gomez-Skarmeta A F, Roubos H, Babuska R (2001) A multi-objective evolutionary algorithm for fuzzy modelling. Proceedings of 9th IFSA World Congress and 20th NAFIPS International Conference, Vancouver, Canada, 1222–1228
[31] Knowles J D, Corne D W (2000) Approximating the non dominated front using the Pareto archived evolution strategy. Evolutionary Computation 8(2):149–172
[32] Narukawa K, Nojima Y, Ishibuchi H (2005) Modification of evolutionary multiobjectiveoptimization algorithms for multiobjective design of fuzzy rule-based classification systems. Proceedings of 2005 IEEE International Conference on Fuzzy Systems, 809–814
[33] Mamdani E H, Assilian S (1975) An experiment in linguistic synthesis with a fuzzy logic controller. International Journal of Man-Machine Studies 7:1–13

[34] Rudolph G, Agapie A (2000) Convergence properties of some multi-objective evolutionary algorithms. Proceedings of the 2000 Conference on Evolutionary Computation, 1010–1016
[35] Srinivas N, Deb K (1994) Multiobjective optimization using nondominated sorting in genetic algorithms. Evolutionary Computation 2:221-248
[36] Takagi T, Sugeno M (1985) Fuzzy identification of systems and its application to modeling and control. IEEE Transactions on Systems, Man, and Cybernetics 15(1):116–132
[37] Wang L X, Mendel J M (1992) Generating fuzzy rules by learning from examples. IEEE Transactions on Systems, Man, and Cybernetics 22(6):1414–1427
[38] Wang H L, Kwong S, Jin Y C, Wei W, Man K F (2005) Multi-objective hierarchical genetic algorithm for interpretable fuzzy rule-based knowledge extraction. Fuzzy Sets and Systems 149(1):149–186
[39] Wang H L, Kwong S, Jin Y C, Wei W, Man K F (2005) Agentbased evolutionary approach for interpretable rule-based knowledge extraction. IEEE Trans. on Systems, Man, and Cybernetics - Part C: Applications and Reviews 35(2):143–155
[40] Zitzler E, Thiele L (1998) Multiobjective optimization using evolutionary algorithms–a comparative study. In Eiben A (eds) Parallel Problem Solving from Nature V, Springer-Verlag, Amsterdam, 292–301
[41] Zitzler E, Thiele L (1999) Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach. IEEE Transactions on Evolutionary Computation 3(4):257–271
[42] Zitzler E, Deb K, Thiele L (2000) Comparison of multiobjective evolutionary algorithms: Empirical results. Evolutionary Computation 8(2):173–195
[43] Zitzler E, Laumanns M, Thiele L (2001) SPEA2: Improving the strength Pareto evolutionary algorithm for multiobjective optimization. Proceedings of the Evolutionary Methods for Design, Optimization and Control with Application to Industrial Problems, Barcelona, Spain, 95-100