Evol. Intel. (2009) 2:39–51
DOI 10.1007/s12065-009-0026-z

SPECIAL ISSUE

Rule base and adaptive fuzzy operators cooperative learning of Mamdani fuzzy systems with multi-objective genetic algorithms

Antonio A. Márquez · Francisco A. Márquez · Antonio Peregrín

Received: 1 July 2009 / Revised: 24 September 2009 / Accepted: 25 September 2009 / Published online: 24 October 2009

Abstract In this paper, we present an evolutionary multi-objective learning model achieving cooperation between the rule base and the adaptive fuzzy operators of the inference system in order to obtain simpler, more compact and still accurate linguistic fuzzy models by learning fuzzy inference adaptive operators together with rules. The multi-objective evolutionary algorithm proposed generates a set of fuzzy rule based systems with different trade-offs between interpretability and accuracy, allowing the designers to select the one that involves the most suitable balance for the desired application. We develop an experimental study testing our approach with some variants on nine real-world regression datasets finding the advantages of cooperative compared to sequential models, as well as multi-objective compared with single-objective models. The study is elaborated comparing different approaches by applying non-parametric statistical tests for pair-wise. Results confirm the usefulness of the proposed approach.

Keywords Linguistic fuzzy modelling · Interpretability-accuracy trade-off · Multi-objective genetic algorithms · Adaptive inference system · Adaptive defuzzification · Rule learning

A. A. Márquez · F. A. Márquez · A. Peregrín (✉)
Information Technologies Department, University of Huelva, Huelva, Spain
e-mail: peregrin@dti.uhu.es

A. A. Márquez
e-mail: amarquez@dti.uhu.es

F. A. Márquez
e-mail: alfredo.marquez@dti.uhu.es

1 Introduction

Many automated techniques to extract a proper set of fuzzy rules from numerical data are proposed in the literature. Most of these techniques usually aim to improve the performance associated with the prediction error without paying special attention to the system interpretability, an essential aspect of Fuzzy Rule-Based Systems (FRBSs) [8, 9, 33]. In recent years, the problem of finding the right trade-off between interpretability and accuracy, in spite of the original nature of fuzzy logic, has produced a growing interest in methods which take both aspects into account [8, 9]. Of course, the ideal scenario would be to satisfy both criteria to a high degree, but since they are contradictory issues, generally it is not possible. One way to achieve this is to improve the system accuracy while trying to maintain the interpretability to an acceptable level [9].

In this framework, adaptive inference systems and defuzzification methods have acquired greater importance [3, 12, 28] since they can find a way to defuzzify the contribution of each rule appropriately [12] and improve the cooperation between rules [3]. They can also specifically adapt the behaviour of the fuzzy operators to the rule base (RB) or learn the RB and the fuzzy operators jointly [28], obtaining a positive synergy between both elements that lets the system reach greater levels of accuracy, while not only maintaining but also improving the interpretability.

Recently, the use of Multi-objective Evolutionary Algorithms (MOEA) has been applied to improve the aforementioned trade-off between interpretability and accuracy of linguistic fuzzy systems [1, 2, 11, 17, 24, 27, 29]. Some of them obtain the complete Pareto (the set of non-dominated solutions with different trade-offs) by selecting or learning the set of rules which best represents the example data, i.e., improving the system accuracy and decreasing the FRBS complexity. In [1, 2, 17] the authors also propose tuning the membership functions together with the rule selection to obtain simpler and still accurate linguistic fuzzy models.

Following these ideas on the use of adaptive fuzzy operators and MOEAs to improve the trade-off between interpretability and accuracy, in [27] we present a MOEA capable of learning the fuzzy inference operators (both inference and defuzzification) and performing rule selection [22–25] for Mamdani fuzzy linguistic systems jointly. The proposed model aimed to achieve a positive synergy based on the cooperation between the fuzzy operators and the RB to improve the accuracy while simplifying the RB, i.e., improving the complexity-based interpretability [33].

Our main aim in this work was to extend the aforementioned model, proposing a new family of models that also look for cooperation between the fuzzy operators and RB based on the learning of both of them jointly and maintaining the multi-objective evolutionary model philosophy to obtain a set of solutions with different balance between accuracy and interpretability.

To this end, we employed two different MOEA models to concurrently perform the rule learning (also including rule selection) and learning fuzzy operator parameters (inference system and defuzzification method) with the following two objectives: system error and number of rules. The MOEAs selected are based on the well-known SPEA2 [34] and NSGA-II [13] algorithms. They improve their search ability by incorporating a method for guiding the search towards the desired Pareto zone.

We tested the proposed approach using nine real-world regression datasets, and compared them with sequential multi-objective based approaches and accuracy based single objective approaches [3, 27, 28]. We used non-parametric statistical tests for pair-wise comparison taking into consideration three representative points from the Pareto fronts obtained for the MOEAs.

The next section describes the adaptive inference and defuzzification method. Section 3 shows the RB learning and rule selection. Section 4 is devoted to describing the MOEA learning proposal, explaining its main characteristics and the genetic operators considered. Section 5 develops an experimental study and shows the results obtained. Finally, Sect. 6 presents some concluding remarks. Additionally, an appendix describes the Wilcoxon signed-rank test used in our study.

2 Adaptive fuzzy operators

In this section we describe the adaptive inference system as well as the adaptive defuzzification method used in our learning proposal. From now on, we will name them together as Adaptive Fuzzy Operators (AFO).

2.1 Adaptive inference system

Linguistic FRBSs for system modelling use IF–THEN rules in the following form:

R_i : If X_{i1} is A_{i1} and. . .and X_{im} is A_{im} then Y is B_i

with $i = 1$ to N, where N stands for the number of rules of the RB, X_{i1} to X_{im} and Y for the input and output variables, respectively, and A_{i1} to A_{im} and B_i for the involved antecedents and consequent labels, respectively.

The expression of the Compositional Rule of Inference in fuzzy modelling with punctual fuzzification is the following: $\mu_B\ (y) = \mathrm{I}\ (C\ (\mu_{\mathrm{A1}}\ (x_1),\ldots,\ \mu_{\mathrm{Am}}\ (x_{\mathrm{m}})),\ \mu_B\ (y))$, where $\mu_B\ (\cdot)$ is the membership function of the inferred consequent, $I(\cdot)$ is the implication operator, $C(\cdot)$ is the conjunction operator, $\mu_{\mathrm{Ai}}(x_{\mathrm{i}})$ are the values of the matching degree of each input of the system with the membership functions of the rule antecedents, and $\mu_B(\cdot)$ is the rule consequent.

Those operators, the conjunction ($C(\cdot)$) and the implication operator ($I(\cdot)$) are suitable for parameterization in order for the inference system to be adapted. Our previous studies in [3] show that using models based on the adaptive conjunction is a more valuable option than those based on the adaptive implication operator. Hence, we selected the adaptive conjunction in this study in order to insert parameters in the inference system.

Taking into account the aforementioned studies in [3], we have selected the Dubois adaptive t-norm with a separate connector for every rule, the expression for which is shown in (1).

$$T_{\mathrm{Dubois}}(x, y, \alpha) = \frac{x \cdot y}{\mathrm{Max}(x, y, \alpha)},\ (0 \le \alpha \le 1) \tag{1}$$

This adaptive t-norm showed the highest accuracy in previous studies, compared with Frank and Dombi t-norms and is more efficiently computed. The use of an adaptive t-norm for the antecedent connection seeks better performance than traditional t-norms. Dubois t-norm performs between minimum ($\alpha = 0$) and algebraic product ($\alpha = 1$).

2.2 Adaptive defuzzification method

There are various tendencies in the development of adaptive defuzzification methods reported in the literature. These employ one or more parameters in their expression to modify the behaviour of the defuzzifier or, in most cases, to achieve higher accuracy.

Following the studies developed in [12], in this work we consider applying the defuzzification function to the fuzzy set inferred by each rule (obtaining a characteristic value) and computing them by a weighted average operator, because of its fine performance, efficiency and easier implementation. This way of working is named FITA (First Infer, Then Aggregate) [5].

We also consider the use of a product functional term of the matching degree between the input variables and the rule antecedent fuzzy sets (h_i), $f(h_i) = h_i \cdot \beta_i$ where β_i corresponds to one parameter for each rule R_i, $i = 1$ to N, in the RB, as it is more efficiently computed and obtains similar results to other functions [12]. The adaptive defuzzification formula selected is shown in (2),

$$y_0 = \frac{\sum_i^N h_i \cdot \beta_i . V_i}{\sum_i^N h_i . \beta_i}, \tag{2}$$

where V_i represents a characteristic value of the fuzzy set inferred from rule R_i, the Maximum Value or the Gravity Centre (GC), the latter being the one selected in this paper.

The product functional term with a different parameter for each rule has the effect of weighted rules. This value associated with the rule indicates the importance of that rule for the inference process.

3 Rule base learning

This section introduces the fuzzy rule selection technique and the RB learning approach used in our study.

3.1 Fuzzy rule selection

Fuzzy rule set reduction techniques try to minimize the number of rules of a given FRBS while maintaining (or even improving) the system performance. To do so, erroneous and conflicting rules that degrade the performance are eliminated, obtaining a more cooperative fuzzy rule set and, as a result, potentially improving system accuracy. Furthermore, in many cases the accuracy is not the only requirement of the model and the interpretability also becomes an important aspect. Reducing the model complexity is a way to improve the system readability, i.e., a compact system with few rules generally requires less effort for its interpretation. Fuzzy rule set reduction techniques are usually applied as a post-processing stage, once an initial fuzzy rule set has been extracted.

One of the most used fuzzy rule set reduction techniques is rule selection. This approach involves obtaining an optimal subset of fuzzy rules from a previous fuzzy rule set by selecting some of them. We may find several methods for rule selection, with different search algorithms that look for the most successful combination of fuzzy rules [21–25]. In [26], an interesting heuristic rule selection procedure is proposed where, by means of statistical measures, a relevance factor is computed for each fuzzy rule composing the FRBSs to subsequently select the most relevant ones. These kinds of techniques for rule selection could be easily combined with other post-processing techniques to obtain more compact and accurate FRBSs. In this way, some works have considered the selection of rules together with the learning of parametric operators of inference system by coding all of them (rules and parameters) in the same chromosome [27].

In this work, we employ this methodology when using single-objective models without implicit rule selection. However, when the model uses the rule learning joined with rule selection, another mechanism is used, which is described next.

3.2 Rule base learning algorithms

The linguistic RB learning used in this work is based on the ad-hoc data driven methodology named Cooperative Rules (COR) [7]. This methodology manages a set of consequent label sets (one per rule). Instead of selecting the consequent with the best performance in each subspace as usual (Wang and Mendel [31]), the COR methodology considers the possibility of using another consequent, different from the best, which allows the FRBS to be more accurate thanks to having an 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. *Construction of the search space*—This obtains a set of candidate consequents for each rule.
2. *Selection of the most cooperative fuzzy rule set*—This performs a combinatorial search among these sets seeking 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 search technique because it is effective and quick.

A description of the COR-based rule generation process is shown in the following steps:

Inputs:

- An input–output data set $E = \{e_l, \ldots\ldots, e_M\}$, with $e_l = (x_{l1}, \ldots\ldots, x_{lm}, y_{l1}, \ldots\ldots, y_{lp})$, $l \in \{1, \ldots\ldots, N\}$, N being the data set size, and n (m) being the number of input (output) variables—representing the behaviour of the problem being solved.

- A fuzzy partition of the variable spaces. In our case, uniformly distributed fuzzy sets are regarded. Let A_i be the set of linguistic terms of the ith input variable, with $i \in \{1,\ldots\ldots,m\}$, and Bj be the set of linguistic terms of the jth output variable, with $j \in \{1,\ldots\ldots, p\}$, with $|Ai|$ ($|Bj|$) being the number of labels of the ith (jth) input (output) variable.

Algorithm:

1. Search space construction:

1.1 Define the fuzzy input subspaces containing positive examples: To do so, we should define the positive example set ($E+(\mathrm{Ss})$) for each fuzzy input subspace $\mathrm{Ss} = (A_{s1},\ldots\ldots, A_{si},\ldots\ldots, A_{sm})$, with $A_{si} \in A_i$ being a label, $s \in \{1,\ldots\ldots, N_s\}$, and $N_s = \prod n_i = 1$, $|A_i|$ being the number of fuzzy input subspaces. In this paper, we use the following:

$$E^+(S_s) = \Big\{e_l \in E | \forall\, i \in \{1,\ldots\ldots,m\},\ \forall A_i^{'} \in A_{i,}\, \mu_{A_i^S(x_i^l)} \geq \mu_{A_i'(x_i^l)}\Big\}$$

with $\mu_{A_i^S(.)}$ being the membership function associated with the label A_{si}. Among all the N_s possible fuzzy input subspaces, consider only those containing at least one positive example. To do so, the set of subspaces with positive examples is defined as $S^+ = \{S_h | E+(S_h) \neq \varnothing\}$.

1.2 Generate the set of candidate rules in each subspace with positive examples: Firstly, the candidate consequent set associated with each subspace containing at least an example, $S_h \in S^+$, is defined. In this paper, we use the following:

$$C(S_h) = \Big\{\big(B_{1kh},\ldots\ldots,B_{pkh}\big) \in B_1x\ldots xB_p | \exists\, e_l \in E+(S_h) \text{ where } \forall j \in \{1,\ldots\ldots,p\}, \forall B'_j \in B_j, \mu_{B_j^{k_h}(y_j^l)} \geq \mu_{B_j'(y_j^l)}\Big\}$$

Then, the candidate rule set for each subspace is defined as $C_R(S_h) = \{R_{kh} = [\text{IF } X_1 \text{ is } A_{h1} \text{ and}\ldots \text{ and } X_m \text{ is } A_{hm} \text{ THEN } Y_1 \text{ is } B_{1kh} \text{ and}\ldots \text{ and } Y_p \text{ is } B_{pkh}]$ such that $B_{kh} = (B_{1kh},\ldots\ldots, B_{pkh}) \in C(S_h)\}$. To allow COR to reduce the initial number of fuzzy rules, the special element $R_\varnothing$ (which means "don't care") is added to each candidate rule set, i.e., $C_R(S_h) = C_R(S_h) \cup R_\varnothing$. If it is selected, no rules are used in the corresponding fuzzy input subspace. In this work we denote it as COR-S.

2. Selection of the most cooperative fuzzy rule set—This stage is performed by running a combinatorial search algorithm to look for the combination $R_B = \{R_1 \in C_R(S_1),\ldots\ldots, R_h \in C_R(S_h),\ldots\ldots,R_{|S+1|} \in C_R(S_{|S+1|})\}$ with the best accuracy. Since the search space tackled is usually large, approximate search techniques should be used. As mentioned before, in this work we use this technique for both single and multi-objective algorithms when we employ selection and rule learning jointly.

4 Rule base and inference system cooperative learning with multi-objective algorithms

This section describes the evolutionary multi-objective model proposed in this work. As previously mentioned, our aim is to obtain a set of fuzzy systems with different trade-offs between accuracy and interpretability, using AFO and RB learning (including rule selection). To do this, we exploit two specific MOEAs considering a threefold coding scheme [coding the parameters of the AFO (inference systems and defuzzification) and RB]. They are two specific MOEAs based on the well-known and representative second generation ones, SPEA2 [34] and NSGA-II [13].

4.1 Improvements for SPEA2 and NSGA-II

The SPEA2 algorithm [34] (*Strength Pareto Evolutionary Algorithm for Multi-objective Optimization*) is characterized by two aspects: a *fitness* assignment strategy, which takes into account both dominating and dominated solutions for each individual, and a density function, estimated by employing the nearest neighbourhood, which guides the search more efficiently.

NSGA-II algorithm [13] is a parameterless approach with several interesting principles: a binary tournament selection based on fast non-dominated sorting, an elitist strategy and a crowding distance method to estimate the diversity of a solution.

Following the experiences in [27], we propose to guide the searching process of SPEA2 and NSGA-II employing a method called Guided Domination Approach [6], which gives priority to the accuracy objective through a weighted function of the objectives. Focusing the searching process in this way, we can reduce the effort of the search, and a better precision in the non-dominated solutions can be obtained, because the searching effort is concentrated in a more interesting and reduced zone of the Pareto, the density of the obtained solutions being higher. The weighted function of the objectives is defined in (3),

$$\Omega_i(f(x)) = f_i(x) + \sum_{j=1,\, j\neq i}^{M} a_{ij} f_j(x), \quad i = 1,2,\ldots,M \tag{3}$$

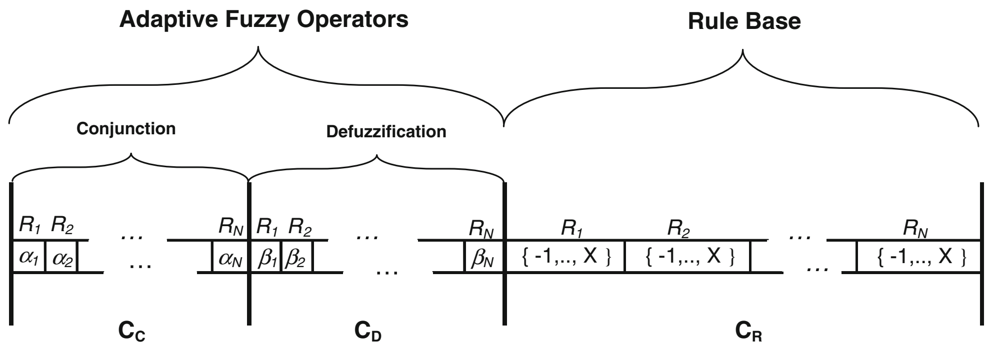

Fig. 1 Coding scheme for the MOEA with N rules

where a_{ij} is the amount of gain in the jth objective function for a loss of one unit in the ith objective function, and M is the number of objectives. The above set of equations require fixing the matrix a, which has a one in its diagonal elements. This method redefines the domination concept as follows:

A solution $x^{(1)}$ dominates another solutions $x^{(2)}$, if $\Omega_i(f(x^{(1)})) \leq \Omega_i(f(x^{(2)}))$ for all $i = 1,2,\ldots, M$, and the strict inequality is satisfied at least for one objective.

Thus, if we have two fitness functions ($M = 2$), the two weighted functions are showed in (4).

$$\Omega_1(f_1,f_2) = f_1 + a_{12}f_2,\ \Omega_2(f_1,f_2) = a_{21}f_1 + f_2 \tag{4}$$

We would like to emphasize that this approach can be viewed as a modified domination principle. This procedure can be viewed as a multi-objective optimization approach with the original domination principle acting on a linearly transformed set of objective functions. A little thought will reveal that the above definition of domination on the objective vector f is the same as the original domination definition on the transformed vector Ω. This approach is different to one that would have considered only single-objective algorithms using these linear combinations [6], because in this case we obtain a set of FRBSs in a single run.

4.2 The multi-objective evolutionary approach

In this section we describe some components of the MOEAs employed.

The evolutionary model uses a chromosome with a threefold coding scheme ($C_C + C_D + C_R$) (Fig. 1) where:

- C_C encodes the α_i parameters of the conjunction connective. They are N real coded parameters (genes), one for each rule, R_i, of the linguistic RB. Each gene can take any value in the interval [0, 1], that is, between the minimum and the algebraic product.
- C_D encodes the β_i parameters of the defuzzification. They are N real coded parameters, one for each rule, of the linguistic RB. Each gene can take any value in the interval [0, 10]. This interval has been selected according to the study developed in [12]. It allows attenuation as well as enhancement of the matching degree.
- C_R encodes the learning RB. It is an integer string of N genes, each one representing a candidate rule consequent of the initial RB. Furthermore, depending on whether a rule is selected or not, the value '−1' is assigned to the corresponding gene.

The initial population is randomly initialized in the fuzzy operators part with the exception of a single chromosome:

- C_C with the N genes is initiated to 0 in order to make Dubois t-norm equivalent to Minimum t-norm initially.
- C_D also with the N genes is initiated to 1 with the objective of beginning like the standard WCOA method.

The initial population in the fuzzy rule part, C_R, is initialized following these two exceptions:

- A single chromosome with the N rules obtained by the WM-method [31], that is, with all the genes initialized to correspondent consequent.
- Default chromosomes randomly initiated with all rules activated. In this case, in order to achieve solutions with a high accuracy we should not lose rules that could present a positive cooperation once their parameters have been evolved. The best way to do this is to start with solutions that select all the possible rules. This favours a progressive extraction of bad rules (those that do not improve with the tuning of parameters).

The crossover operator employed by the fuzzy operators part is BLX-0.5 [15] while the one used for the rule learning part is HUX [16].

Finally, four offspring are generated by combining the two from the C_R part with the two from the operators part (the two best replace their parents). The mutation operator changes a gene value at random in the C_R and operators part (one in each part) with probability 0.2.

In this work, the fitness is based on the interpretability (using the number of rules) and on the accuracy (using the error measure). Both must be minimized.

5 Experimental study

In order to analyse the proposed methods, we have used nine real-world problems with different complexities (different numbers of variables and available data). Table 1 summarizes the main characteristics of the nine datasets selected from the KEEL project webpage [4] (http://www.keel.es) where they can be downloaded.

This section is organized as follows:

First, we describe the experimental set-up in Sect. 5.1. Then, Sect. 5.2 analyses the behaviour of the proposed approach. In order to better analyse theses MOEAs, they have also been compared with their single-objective counterpart. Finally, Sect. 5.3 shows some representative plots of the Pareto fronts obtained.

5.1 Experimental set-up

Methods considered for the experiments are briefly described in Table 2, where the well-known ad-hoc data-driven learning algorithms of Wang and Mendel (WM) [31] and COR [7], denoted as RB_{WM} and RB_{COR}, are applied to obtain an initial set of candidate linguistic rules. The initial linguistic partitions are comprised of five linguistic terms in the case of datasets with less than nine variables and three linguistic terms in the remaining ones (which helps to obtain a more reasonable number of rules in the main datasets). If rule selection is performed after RB learning, we denote it as $RB_{WM} + S$ when using WM or $RB_{COR} + S$ with COR.

Table 1 Data sets considered for the experimental study

Datasets	Name	Variables	Patterns
Plastic strength	PLA	3	1,650
Quake	QUA	4	2,178
Electrical maintenance	ELE	5	1,056
Abalone	ABA	9	4,177
Stock prices	STP	10	950
Ankara weather	WAN	10	1,609
Izmir weather	WIZ	10	1,461
Mortgage	MOR	16	1,409
Treasure	TRE	16	1,409

Generally, we use "+" to denote "sequence" and "−" to denote "together", so if rule selection is performed jointly with COR, we use $RB_{COR\text{-}S}$. $AFO - S$ which means adaptive fuzzy operators learning together with rule selection, consequently $RB_{WM} + AFO - S$ denotes the use of RB obtained with WM and the single-objective evolutionary learning of the AFO together with rule selection. $RB_{WM} + (AFO - S)_{SP2}$, and $RB_{WM} + (AFO - S)_{NS\text{-}II}$ are the MOEAs that learn the AFO and perform rule selection together (with the models based on SPEA2 and NSGA-II, respectively). The new methods proposed in this work are $(AFO - RB_{COR\text{-}S})_{SP2}$ and $(AFO - RB_{COR\text{-}S})_{NS\text{-}II}$ that is, the use of the MOEAs based on SPEA2 and NSGA-II to learn the AFO and the $RB_{COR\text{-}S}$ jointly in the same evolutionary multi-objective process.

In all experiments, we adopted a 5-folder cross-validation model, i.e., five random partitions of the data each with 20% of the patterns of the data set, and used four folds for training and one for testing. For each of the data partitions, the learning methods were run 6 times using different seeds for the random number generator. For each data set, we therefore consider the average results of 30 runs. In the case of methods with multi-objective approach, for each data set and for each trial we generate the approximated Pareto front. Then, we focus on three representative points: the first (the most accurate), the median and the last (the least accurate). For each dataset, we compute the mean values and the standard deviations over the 30 trials of the mean square error (MSE) on the training and test sets and the number of rules (#R) in the FBRSs. For the single objective-based approaches, we compute the mean values over the 30 solutions obtained for each dataset.

This way to work was also employed in [27] in order to compare the single objective methods with the multi-objective ones considering only the accuracy objective, letting us see that the Pareto fronts are not only wide but also optimal. The MSE is computed with the expression (5),

$$MSE(S)_B = \frac{\frac{1}{2}\sum_{k=1}^{P}(y_k - S(x_k))^2}{P} \tag{5}$$

where S denotes the fuzzy model whose inference system uses the Dubois adaptive t-norm as conjunction operator showed in expression (1), inference operator minimum t-norm, and the adaptive defuzzification method showed in expression (2). This measure uses a set of system evaluation data formed by P pairs of numerical data $Zk = (x_k, y_k)$, $k = 1,\ldots, P$, with x_k being the values of the input variables, and y_k being the corresponding values of the associated output variables.

In order to assess whether there are significant differences among the results, we adopted statistical analysis

Table 2 Methods considered for comparison

Reference	Method	Description
[31]	RB_{WM}	RB obtained with WM
[10]	$RB_{WM} + S$	RB obtained with WM and then, rule selection
[7]	RB_{COR}	RB obtained with COR
[7]	$RB_{COR\text{-}S}$	RB obt. with COR using the rule selection
–	$RB_{COR} + S$	RB obt. with COR and then, rule selection
[28]	$RB_{WM} + AFO - S$	RB obt. with WM and then, AFO with rule selection learning using a single objective CHC based model
[28]	$AFO - RB_{COR\text{-}S}$	AFO and $RB_{COR\text{-}S}$ jointly using a single objective CHC based model
[27]	$RB_{WM} + (AFO - S)_{SP2}$	AFO with rule selection using SPEA2, using a RB obt. with WM
[27]	$RB_{WM} + (AFO - S)_{NS\text{-}II}$	AFO with rule selection using NSGA-II, using a RB obt. with WM
–	$(AFO - RB_{COR\text{-}S})_{SP2}$	AFO with RB learning based on COR-S using SPEA2
–	$(AFO - RB_{COR\text{-}S})_{NS\text{-}II}$	AFO with RB learning based on COR-S using NSGA-II

[18–20] and in particular non-parametric tests, according to the recommendations made in [14, 19], where a set of simple, safe and robust non-parametric tests for statistical comparisons of classifiers were introduced. In particular for pair-wise comparison we use the Wilcoxon signed-rank test [30, 32]. A detailed description of this test is presented in the Appendix. To perform the test, we used a level of confidence $\alpha = 0.1$. In particular, the Wilcoxon test is based on computing the differences on two sample means (typically, mean test errors obtained by a pair of different algorithms on different datasets). In the classification framework, these differences are well defined since these errors are in the same domain. In the regression framework, to make the differences comparable, we propose to adopt a normalized difference DIFF, defined as:

$$DIFF = \frac{MSE_{TST}(Other) - MSE_{TST}(Reference)}{MSE_{TST}(Other)} \quad (6)$$

This difference expresses the improvement percentage of the reference algorithm on the other one.

The results of the initial FRBSs obtained by WM and COR are presented in Table 3 as reference data where the number of rules (#*R*) are the same for both methods due to the fact that fuzzy input subspaces were likewise obtained. The values of the parameters considered by the single-objective methods are: the population size was 50, 240,000 evaluations were performed, 0.6 was the crossover probability and 0.2 was the mutation probability per chromosome. The values of the parameters considered by the MOEAs are: the population size was fixed to 200, the external population size was 61, 240,000 evaluations were performed and 0.2 was the mutation probability.

Table 3 Initial results obtained by WM and COR

Datasets	#*R*	RB_{WM}		RB_{COR}	
		MSE_{tra}	MSE_{tst}	MSE_{tra}	MSE_{tst}
PLA	14.8	3.434	3.557	1.477	1.480
QUA	53.6	0.0258	0.0267	0.0178	0.0189
ELE	65.0	56,135	56,359	50,711	54,585
ABA	68.2	8.407	8.424	3.047	3.058
STP	122.8	9.074	9.042	5.254	5.279
WAN	156.0	16.063	16.403	7.150	7.357
WIZ	104.8	6.945	7.139	4.931	5.083
MOR	77.6	0.985	0.973	0.646	0.653
TRE	75.0	1.636	1.632	1.519	1.524

The parameters a_{12}, a_{21} used by the Guided Domination Approach of the MOEAs were determined after several tests and fixed to 0 and 16, respectively, i.e., strongly focused on the accuracy.

5.2 Results and analysis

This section shows and analyses the results of the proposed cooperative methods $(AFO - RB_{COR\text{-}S})_{SP2}$, $(AFO - RB_{COR\text{-}S})_{NS\text{-}II}$ against sequential $RB_{WM} + (AFO - S)_{SP2}$, $RB_{WM} + (AFO - S)_{NS\text{-}II}$ and the single objective counterpart $RB_{WM} + AFO - S$ and $AFO - RB_{COR\text{-}S}$. Tables 4, 5, 6 and 7 show the results obtained by these methods in the three representative points of the accuracy-interpretability plane when we used MOEAs with best accuracy highlighted in bold. For the single objective-based approaches only the most accurate points are shown.

Table 4 Results obtained by the sequential models compared with their single objective counterparts and reference models

Dataset	Method	MAX INT			MEDIAN INT/ACC			MAX ACC		
		#R	MSE_{tra}	MSE_{tst}	#R	MSE_{tra}	MSE_{tst}	#R	MSE_{tra}	MSE_{tst}
PLA	RB_{WM} + S	–			–			11.6	2.403	2.477
	RB_{WM} + AFO − S	–			–			12.9	1.783	1.886
	RB_{WM} + (AFO − S)$_{SP2}$	11.3	1.651	1.722	12.1	1.641	1.712	13.3	1.639	**1.711**
	RB_{WM} + (AFO − S)$_{NS\text{-}II}$	11.4	1.654	1.722	12.1	1.646	1.715	13.3	1.642	1.713
QUA	RB_{WM} + S	–			–			29.9	0.022	0.023
	RB_{WM} + AFO − S	–			–			34.6	0.020	0.022
	RB_{WM} + (AFO − S)$_{SP2}$	25.3	0.0205	0.0221	27.7	0.0205	0.0220	30.5	0.020	0.022
	RB_{WM} + (AFO − S)$_{NS\text{-}II}$	25.7	0.0206	0.0222	27.9	0.0205	0.0220	30.8	0.020	**0.021**
ELE	RB_{WM} + S	–			–			41.8	41,642	44,037
	RB_{WM} + AFO − S	–			–			50.8	21,804	26,054
	RB_{WM} + (AFO − S)$_{SP2}$	44.1	22,812	26,448	45.2	22,723	26,201	46.9	22,663	**25,840**
	RB_{WM} + (AFO − S)$_{NS\text{-}II}$	44.2	23,595	27,639	45.0	23,595	27,639	45.5	23,542	26,593
ABA	RB_{WM} + S	–			–			25.6	5.083	5.049
	RB_{WM} + AFO − S	–			–			36.5	4.578	**4.600**
	RB_{WM} + (AFO − S)$_{SP2}$	20.8	4.627	4.654	22.8	4.617	4.648	25.2	4.614	4.647
	RB_{WM} + (AFO − S)$_{NS\text{-}II}$	21.5	4.654	4.685	23.5	4.645	4.680	25.2	4.635	4.670
STP	RB_{WM} + S	–			–			36.3	2.624	2.786
	RB_{WM} + AFO − S	–			–			61.4	1.364	**1.434**
	RB_{WM} + (AFO − S)$_{SP2}$	36.8	1.391	1.472	41.3	1.397	1.456	46.3	1.376	1.450
	RB_{WM} + (AFO − S)$_{NS\text{-}II}$	38.1	1.419	1.492	42.9	1.406	1.475	45.1	1.403	1.466
WAN	RB_{WM} + S	–			–			55.8	6.567	7.181
	RB_{WM} + AFO − S	–			–			90.0	3.566	4.340
	RB_{WM} + (AFO − S)$_{SP2}$	67.1	3.680	4.267	70.7	3.666	4.270	74.8	3.662	**4.265**
	RB_{WM} + (AFO − S)$_{NS\text{-}II}$	69.5	3.763	4.374	73.5	3.748	4.337	77.7	3.743	4.335
WIZ	RB_{WM} + S	–			–			46.0	3.036	3.506
	RB_{WM} + AFO − S	–			–			71.4	0.782	1.257
	RB_{WM} + (AFO − S)$_{SP2}$	55.5	0.858	1.282	59.7	0.840	1.254	64.5	0.834	**1.184**
	RB_{WM} + (AFO − S)$_{NS\text{-}II}$	57.7	0.903	1.456	61.6	0.887	1.414	61.6	0.888	1.388

Table 5 Results obtained by the sequential models compared with their single objective counterparts and reference models (highest data sets)

Dataset	Method	MAX INT			MEDIAN INT/ACC			MAX ACC		
		#R	MSE_{tra}	MSE_{tst}	#R	MSE_{tra}	MSE_{tst}	#R	MSE_{tra}	MSE_{tst}
MOR	RB_{WM} + S	–			–			23.2	0.204	0.213
	RB_{WM} + AFO − S	–			–			41.7	0.082	**0.097**
	RB_{WM} + (AFO − S)$_{SP2}$	29.8	0.095	0.105	32.4	0.094	0.104	35.6	0.094	0.104
	RB_{WM} + (AFO − S)$_{NS\text{-}II}$	30.7	0.101	0.113	33.2	0.100	0.111	36.1	0.100	0.111
TRE	RB_{WM} + S	–			–			23.3	0.342	0.374
	RB_{WM} + AFO − S	–			–			41.4	0.078	**0.091**
	RB_{WM} + (AFO − S)$_{SP2}$	30.0	0.103	0.116	33.1	0.100	0.105	36.6	0.100	0.111
	RB_{WM} + (AFO − S)$_{NS\text{-}II}$	33.6	0.113	0.123	33.6	0.109	0.120	37.1	0.109	0.119

Model behaviour can be observed, but it is easier to reach conclusions viewing Tables 8, 9 and 10, which show the results of the Wilcoxon test on the MSE_{tst} and the #R for the most accurate point (MAX ACC).

Table 8 shows the results when comparing the sequential models, single objective against multi-objective. The accuracy is similar among them, but the multi-objective is more interpretable. The null hypothesis for Wilcoxon test

Table 6 Results obtained by the cooperative models compared with their single objective counterparts and reference models

Dataset	Method	MAX INT			MEDIAN INT/ACC			MAX ACC		
		#R	MSE_{tra}	MSE_{tst}	#R	MSE_{tra}	MSE_{tst}	#R	MSE_{tra}	MSE_{tst}
PLA	RB_{COR} + S	–			–			14.8	1.477	1.480
	$RB_{COR\text{-}S}$	–			–			14.8	1.477	1.480
	AFO − $RB_{COR\text{-}S}$	–			–			14.5	1.110	1.149
	(AFO − $RB_{COR\text{-}S})_{SP2}$	12.9	1.109	1.146	13.5	1.106	1.143	14.5	1.105	**1.145**
	(AFO − $RB_{COR\text{-}S})_{NS\text{-}II}$	12.9	1.119	1.156	13.5	1.113	1.149	14.5	1.111	1.149
QUA	RB_{COR} + S	–			–			45.8	0.0178	0.0189
	$RB_{COR\text{-}S}$	–			–			42.5	0.0178	0.0186
	AFO − $RB_{COR\text{-}S}$	–			–			48.0	0.0170	0.0186
	(AFO − $RB_{COR\text{-}S})_{SP2}$	25.1	0.0177	0.0190	31.4	0.0172	0.0186	38.3	0.0171	**0.0185**
	(AFO − $RB_{COR\text{-}S})_{NS\text{-}II}$	25.9	0.0177	0.0190	31.8	0.0173	0.0187	38.1	0.0172	**0.0185**
ELE	RB_{COR} + S	–			–			44.7	40,763	43,228
	$RB_{COR\text{-}S}$	–			–			40.8	38,153	38,926
	AFO − $RB_{COR\text{-}S}$	–			–			53.9	18,981	**21,122**
	(AFO − $RB_{COR\text{-}S})_{SP2}$	37.6	21,648	24,975	41.1	20,366	23,141	45.1	19,959	22,585
	(AFO − $RB_{COR\text{-}S})_{NS\text{-}II}$	38.0	22,240	25,099	41.7	20,835	23,445	46.1	20,331	22,763
ABA	RB_{COR} + S	–			–			44.4	2.829	2.908
	$RB_{COR\text{-}S}$	–			–			39.0	2.737	2.760
	AFO − $RB_{COR\text{-}S}$	–			–			50.3	2.345	**2.465**
	(AFO − $RB_{COR\text{-}S})_{SP2}$	29.9	2.443	2.557	32.7	2.425	2.548	36.1	2.417	2.541
	(AFO − $RB_{COR\text{-}S})_{NS\text{-}II}$	31.2	2.466	2.586	34.1	2.447	2.568	37.2	2.442	2.553
STP	RB_{COR} + S	–			–			52.9	2.222	2.433
	$RB_{COR\text{-}S}$	–			–			51.2	2.185	2.450
	AFO − $RB_{COR\text{-}S}$	–			–			72.5	1.101	1.188
	(AFO − $RB_{COR\text{-}S})_{SP2}$	41.9	1.116	1.207	46.2	1.106	1.192	50.9	1.103	**1.187**
	(AFO − $RB_{COR\text{-}S})_{NS\text{-}II}$	45.1	1.143	1.225	50.0	1.131	1.210	55.4	1.128	1.204
WAN	RB_{COR} + S	–			–			73.4	4.113	4.534
	$RB_{COR\text{-}S}$	–			–			63.0	3.926	4.562
	AFO − $RB_{COR\text{-}S}$	–			–			101.9	1.279	1.755
	(AFO − $RB_{COR\text{-}S})_{SP2}$	68.5	1.208	1.642	75.6	1.161	1.601	83.0	1.148	**1.590**
	(AFO − $RB_{COR\text{-}S})_{NS\text{-}II}$	68.5	1.287	1.769	75.5	1.234	1.674	83.1	1.219	1.654
WIZ	RB_{COR} + S	–			–			52.1	3.136	3.637
	$RB_{COR\text{-}S}$	–			–			46.8	2.948	3.369
	AFO − $RB_{COR\text{-}S}$	–			–			75.2	0.691	**0.952**
	(AFO − $RB_{COR\text{-}S})_{SP2}$	51.3	0.769	1.388	58.2	0.717	1.189	65.7	0.704	1.101
	(AFO − $RB_{COR\text{-}S})_{NS\text{-}II}$	53.0	0.781	1.206	59.4	0.738	1.145	66.3	0.725	1.076

was rejected with a very small p-value for the #R, which supports our conclusion with a high degree of confidence. However, they are similar in relation to the accuracy objective. In this case, the null hypothesis associated with the Wilcoxon test is accepted. Thus, we cannot conclude that the results achieved by multi-objective methods are different on MSE_{tst} even if they are better more times (R^+), because the difference is short.

Table 9 is similar to Table 8, but when comparing cooperative models, that is, multi-objective approaches obtain similar accuracy to single-objective. Thus, single objective models, accuracy-oriented, using the same number of evaluations as multi-objective ones obtain very good accuracy, but their interpretability is far away from the more accurate solution of the front of the Pareto of the multi-objective method. Therefore, this is an important advantage of the use of multi-objective approaches.

Table 10 shows an interesting comparison between sequential and cooperative, both for single objective and multi-objective. The null hypothesis associated with the Wilcoxon test is now rejected ($p < \alpha$) in both cases. However, the result is different among #R and MSE_{tst}.

Table 7 Results obtained by the cooperative models compared with their single objective counterparts and reference models (highest data sets)

Dataset	Method	MAX INT			MEDIAN INT/ACC			MAX ACC		
		$\#R$	MSE_{tra}	MSE_{tst}	$\#R$	MSE_{tra}	MSE_{tst}	$\#R$	MSE_{tra}	MSE_{tst}
MOR	$RB_{COR} + S$	–			–			26.6	0.1413	0.1466
	$RB_{COR\text{-}S}$	–			–			25.0	0.135	0.140
	$AFO - RB_{COR\text{-}S}$	–			–			50.2	0.0412	**0.049**
	$(AFO - RB_{COR\text{-}S})_{SP2}$	36.6	0.048	0.053	39.8	0.046	0.052	43.5	0.046	0.052
	$(AFO - RB_{COR\text{-}S})_{NS\text{-}II}$	36.5	0.052	0.058	40.1	0.050	0.056	43.9	0.050	0.056
TRE	$RB_{COR} + S$	–			–			25.8	0.337	0.364
	$RB_{COR\text{-}S}$	–			–			25.0	0.311	0.342
	$AFO - RB_{COR\text{-}S}$	–			–			43.1	0.082	**0.097**
	$(AFO - RB_{COR\text{-}S})_{SP2}$	30.6	0.102	0.121	33.8	0.099	0.117	37.6	0.098	0.116
	$(AFO - RB_{COR\text{-}S})_{NS\text{-}II}$	30.7	0.113	0.130	33.8	0.110	0.125	37.3	0.108	0.124

Table 8 Wilcoxon test to compare sequential models: single-objective $RB_{WM} + AFO - S$ (R^+) compared with multi-objective $RB_{WM} + (AFO - S)_{SP2\ and\ NS\text{-}II}$ (R^-) on $\#R$ and MSE_{tst}

Comparison	Measure	R^+	R^-	Hypothesis ($\alpha = 0.1$)	p-value
$RB_{WM} + AFO - S$ vs. $RB_{WM} + (AFO - S)_{SP2}$	MSE_{tst}	19	17	Accepted	1.000
	$\#R$	1	44	Rejected	0.011
$RB_{WM} + AFO - S$ vs. $RB_{WM} + (AFO - S)_{NS\text{-}II}$	MSE_{tst}	32	13	Accepted	0.260
	$\#R$	1	44	Rejected	0.011

Table 9 Wilcoxon test to compare cooperative models: single-objective $AFO - RB_{COR\text{-}S}$ (R^+) compared with the proposed multi-objective $(AFO - RB_{COR\text{-}S})_{SP2\ and\ NS\text{-}II}$ (R^-) on $\#R$ and MSE_{tst}

Comparison	Measure	R^+	R^-	Hypothesis ($\alpha = 0.1$)	p-value
$AFO - RB_{COR\text{-}S}$ vs. $(AFO - RB_{COR\text{-}S})_{SP2}$	MSE_{tst}	32	13	Accepted	0.260
	$\#R$	0	36	Rejected	0.008
$AFO - RB_{COR\text{-}S}$ vs. $(AFO - RB_{COR\text{-}S})_{NS\text{-}II}$	MSE_{tst}	25	11	Accepted	0.327
	$\#R$	0	36	Rejected	0.008

Table 10 Wilcoxon test to compare sequential (R^+) compared with cooperative (R^-) models, among single objective ($RB_{WM} + AFO - S$ vs. $AFO - RB_{COR\text{-}S}$) and among multi-objective ($RB_{WM} + (AFO - S)_{SP2\ and\ NS\text{-}II}$ vs. $(AFO - RB_{COR\text{-}S})_{SP2\ and\ NS\text{-}II}$) on $\#R$ and MSE_{tst}

Comparison	Measure	R^+	R^-	Hypothesis ($\alpha = 0.1$)	p-value
$RB_{WM} + AFO - S$ vs. $AFO - RB_{COR\text{-}S}$	MSE_{tst}	1	44	Rejected	0.011
	$\#R$	45	0	Rejected	0.008
$RB_{WM} + (AFO - S)_{SP2}$ vs. $(AFO - RB_{COR\text{-}S})_{SP2}$	MSE_{tst}	1	44	Rejected	0.011
	$\#R$	44	1	Rejected	0.011
$RB_{WM} + (AFO - S)_{NS\text{-}II}$ vs. $(AFO - RB_{COR\text{-}S})_{NS\text{-}II}$	MSE_{tst}	1	44	Rejected	0.011
	$\#R$	44	1	Rejected	0.011

So, we can conclude that the results achieved by these methods are statistically different on the MSE_{tst} and on the $R\#$. Thus, cooperative models are more accurate than sequential ones, while regarding interpretability, sequential models are more interpretable. This is why $RB_{COR\text{-}S}$ produces more accurate RBs than $RB_{WM} + S$, thanks to the cooperation among rules, whereas the sequential model uses a more effective rule selection mechanism that lets it give more compact RBs.

Therefore, we can conclude that the proposed cooperative multi-objective model gives:

- Higher interpretability than single-objective approaches with similar accuracy.

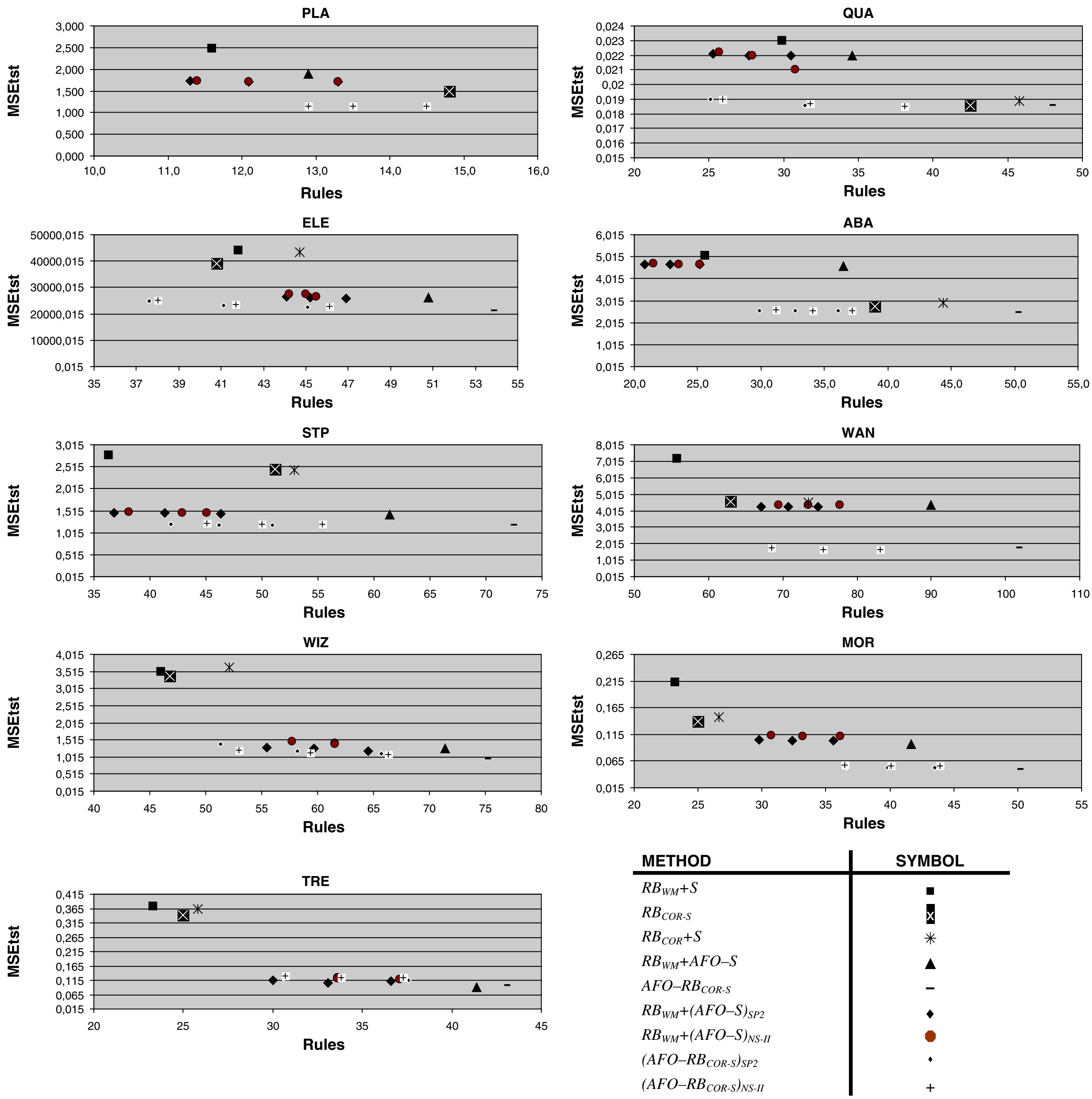

Fig. 2 Averaged Pareto fronts obtained in all the problems

- Higher accuracy than sequential multi-objective approaches.

5.3 Graphical analysis of the Pareto fronts

Now, for each dataset and for each MOEA, we plot on the accuracy-complexity plane the centroids of the first (the most accurate), median and last (the least accurate) solutions obtained on the training and test sets in the different trials of the algorithms, since we performed 30 trials with different training and test partitions, and it would not be readable to show all the Pareto fronts. Thus, we plot the MAX ACC, the MEDIAN ACC/INT and the MAX INT points for each MOEA and for each dataset in Tables 2 and 3. We also show the solutions generated by single-objective methods.

Viewing Fig. 2, we can point out the following:

- Results presented by the proposed cooperative models, both based on the two evolutionary algorithms are, in general, below the other models of the study, that is, they are more accurate.

- However, the three solutions of the MOEAs, the MAX ACC, the MEDIAN ACC/INT and the MAX INT, seems to be nearby, and placed in the more accurate zone. The reason is that we are not using standard SPEA2 and NSGA-II but an improved accuracy version of them based on the Guided Domination Approach depicted in Sect. 4.1. The search process is focused in the reduced Pareto front zone with higher accuracy, and the results obtained are the extremes and central point of the desired zone.
- The cooperation between AFO and the RB makes the proposed cooperative models show different solutions, more accurate than the AFO with sequential models, as was also concluded in Sect. 5.2.

6 Conclusions

In the framework of the trade-off between accuracy and interpretability, the use of MOEAs gives a set of solutions with different levels of conciliation between both features. In this work, we have proposed a multi-objective evolutionary learning model with the two objectives of system error and number or rules, where the adaptive fuzzy operators, including inference and defuzzification, are learnt together with the RB. This fact allows both elements to cooperate, allowing a set of solutions with several optimal trade-offs, with more accuracy than the sequential multi-objective models. It also significantly improves the interpretability against the single objective, accuracy-based models, while giving a similar level of accuracy. These results were achieved developing an experimental study with nine real-world regression data sets comparing the proposed methodology with sequential multi-objective based approaches and single objective accuracy-based approaches with the same number of evaluations for the evolutionary search process. Non-parametric statistical test for pair-wise were used, considering three representative points from the Pareto. The multi-objective algorithms employed are based on two well-known second generation ones with an additional feature to focus the search process to the desired zone, in this case it is centred to the more accurate region of the Pareto front.

Therefore, we consider the proposed methodology as an interesting approach in the design of FRBSs from examples scenario, as it does not suppose a longer developing time and provides high quality solutions.

Acknowledgments Paper supported in part by the Spanish Ministry of Education and Science under grant no. TIN2008-06681-C06-06, and the Andalusian government under grant no. P07-TIC-03179.

Appendix

On the use of Wilcoxon signed-rank test

The Wilcoxon signed-rank test is a pair-wise test that aims to detect significant differences between two sample means: it is analogous to the paired *t*-test in non-parametric statistical procedures. If these means refer to the outputs of two algorithms, then the test practically assesses the reciprocal behaviour of the two algorithms [29, 31]. Let d_i be the difference between the performance scores of the two algorithms on the *i*th out of N_{ds} datasets. The differences are ranked according to their absolute values; average ranks are assigned in case of ties. Let R^+ be the sum of ranks for the datasets on which the first algorithm outperformed the second, and R^- the sum of ranks for the contrary outcome. Ranks of $d_i = 0$ are split evenly among the sums; if there is an odd number of them, one is ignored:

$$R^+ = \sum_{d_i > 0} \mathrm{rank}(d_i) + \frac{1}{2}\sum_{d_i=0} \mathrm{rank}(d_i),$$

$$R^- = \sum_{d_i < 0} \mathrm{rank}(d_i) + \frac{1}{2}\sum_{d_i=0} \mathrm{rank}(d_i),$$

Let T be the smaller of the sums, $T = \min(R^+;R^-)$. If T is less than, or equal to, the value of the distribution of Wilcoxon for N_{ds} degrees of freedom (Table B.12 in [31]), the null hypothesis of equality of means is rejected.The Wilcoxon signed-rank test is more sensitive than the *t*-test. It assumes commensurability of differences, but only qualitatively: greater differences still count for more, which is probably desired, but the absolute magnitudes are ignored. From the statistical point of view, the test is safer since it does not assume normal distributions. Also, the outliers (exceptionally good/bad performances on a few datasets) have less effect on the Wilcoxon test than on the *t*-test. The Wilcoxon test assumes continuous differences di, therefore they should not be rounded to one or two decimals, since this would decrease the test power due to a high number of ties. When the assumptions of the paired *t*-test are met, the Wilcoxon signed-rank test is less powerful than the paired test. On the other hand, when the assumptions are violated, the Wilcoxon test can be even more powerful than the *t*-test. This allows us to apply it to the means obtained by the algorithms in each dataset, without any assumption about the distribution of the obtained results.

References

1. Alcalá R, Alcala-Fdez J, Gacto MJ, Herrera F (2007) A multi-objective evolutionary algorithm for rule selection and tuning on

fuzzy rule-based systems. In: Proceedings of Fuzz-IEEE'07 international conference on fuzzy systems. London, UK, pp 1367–1372
2. Alcalá R, Gacto MJ, Herrera F, Alcala-Fdez J (2007) A multi-objective genetic algorithm for tuning and rule selection to obtain accurate and compact linguistic fuzzy rule-based systems. Int J Uncertain Fuzzin Knowl Based Syst 15(5):539–557
3. Alcalá-Fdez J, Herrera F, Márquez F, Peregrín A (2007) Increasing fuzzy rules cooperation based on evolutionary adaptive inference systems. Int J Intell Syst 22(9):1035–1064
4. Alcalá-Fdez J, Sánchez L, Garcia S, Del Jesus M, Ventura S, Garrell J, Otero J, Romero C, Bacardit J, Rivas V, Fernandez J, Herrera F (2009) Keel: a software tool to assess evolutionary algorithms to data mining problems. Soft Comput 13(3):307–318
5. Buckley JJ, Hayashi Y (1994) Can approximate reasoning be consistent? Fuzzy Sets and Syst. 65(1):13–18
6. Branke J, Kaubler T, Schmeck H (2000) Guiding multi-objective evolutionary algorithms towards interesting region. Technical Report No. 399, Institute AIFB, University of Karlsruhe, Germany
7. Casillas J, Cordón O, Herrera F (2002) COR: a methodology to improve ad hoc data-driven linguistic rule learning methods by inducing cooperation among rules. IEEE Trans Syst Man Cybern 32(4):526–537
8. Casillas J, Cordón O, Herrera F, Magdalena L (eds) (2003) Interpretability issues in fuzzy modelling. Studies in fuzziness and soft computing. Springer, Heidelberg, p 128
9. Casillas J, Cordón O, Herrera F, Magdalena L (eds) (2003) Accuracy improvements in linguistic fuzzy modelling. Studies in fuziness and soft computing. Springer, Heidelberg, p 129
10. Casillas J, Cordón O, Del Jesus MJ, Herrera F (2005) Genetic tuning of fuzzy rule deep structures preserving interpretability for linguistic modelling. IEEE Trans Fuzzy Syst 13(1):13–29
11. Cococcioni M, Ducange P, Lazzerini B, Marcelloni F (2007) A pareto-based multi-objective evolutionary approach to the identification of Mamdani fuzzy systems. Soft Comput 11:1013–1031
12. Cordón O, Herrera F, Márquez FA, Peregrín A (2004) A study on the evolutionary adaptive defuzzification methods in fuzzy modelling. Int J Hybrid Intell Syst 1(1):36–48
13. Deb K, Agrawal S, Pratab A, Meyarivan T (2002) A fast and elitist multi-objective genetic algorithm: NSGA-II. IEEE Trans Evol Comput 6(2):182–197
14. Demšar J (2006) Statistical comparisons of classifiers over multiple data sets. J Mach Learn Res 7:1–30
15. Eshelman LJ (1991) The CHC adaptive search algorithm: how to have safe search when engaging in nontraditional genetic recombination. Found Genet Algorithms 1:265–283
16. Eshelman LJ, Schaffer JD (1993) Real-coded genetic algorithms and interval-schemata. Found Genetic Algorithms 2:187–202
17. Gacto MJ, Alcalá R, Herrera F (2009) Adaptation and application of multi-objective evolutionary algorithms for rule reduction and parameter tuning of fuzzy rule-based systems. Soft Comput 13:419–436
18. García S, Herrera F (2008) An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons. J Mach Learn Res 9:2579–2596
19. García S, Fernández A, Luengo J, Herrera F (2009) A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability. Soft Comput 13(10):959–977
20. García S, Molina D, Lozano M, Herrera F (2009b) A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study on the CEC'2005 special session on real parameter optimization. J Heuristics (in press) doi: 10.1007/s10732-008-9080-4
21. Herrera F, Lozano M, Verdegay JL (1998) A learning process for fuzzy control rules using genetic algorithms. Fuzzy Sets and Systems 100:143–158
22. Ishibuchi H, Murata T, Turksen IB (1997) Single-objective and two objective genetic algorithms for selecting linguistic rules for pattern classification problems. Fuzzy Sets Syst 89(2):135–150
23. Ishibuchi H, Nozaki K, Yamamoto N, Tanaka H (1995) Selecting fuzzy if–then rules for classification problems using genetic algorithms. IEEE Trans Fuzzy Syst 3(3):260–270
24. Ishibuchi H, Yamamoto T (2004) Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining. Fuzzy Sets Syst 141(1):59–88
25. Krone A, Krause H, Slawinski T (2000) A new rule reduction method for finding interpretable and small rule bases in high dimensional search spaces. In: Proceedings of the 9th IEEE international conference on fuzzy systems. San Antonio, TX, USA, pp 693–699
26. Krone A, Taeger H (2001) Data-based fuzzy rule test for fuzzy modelling. Fuzzy Sets Syst 123(3):343–358
27. Márquez AA, Márquez FA, Peregrín A (2008) Cooperation between the inference system and the rule base by using multi-objective genetic algorithms. In: Proceedings of third international of hybrid artificial intelligence system, (HAIS'08). Burgos, Spain, pp 739–746
28. Márquez FA, Peregrín A, Herrera F (2007) Cooperative evolutionary learning of fuzzy rules and parametric aggregation connectors for Mamdani linguistic fuzzy systems. IEEE Trans Fuzzy Syst 15(6):1162–1178
29. Narukawata K, Nojima Y, Ishibuchi H (2005) Modification of evolutionary multi-objective optimization algorithms for multi-objective design of fuzzy rule-based classification systems. In: Proceedings of the 2005 IEEE international conference on fuzzy systems. Reno, pp 809–814
30. Sheskin DJ (2003) Handbook of parametric and nonparametric statistical procedures. CRC Press, Boca Raton
31. Wang LX, Mendel JM (1992) Generating fuzzy rules by learning from examples. IEEE Trans Syst Man Cybern 22(6):1414–1427
32. Wilcoxon F (1945) Individual comparisons by ranking methods. Biometrics 1:80–83
33. Zhou SM, Gan JQ (2008) Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modelling. Fuzzy Sets Syst 159(23):3091–3131
34. Zitzler E, Laumanns M, Thiele L (2001) SPEA2: improving the strength pareto evolutionary algorithm for multi-objective optimization. In: Evolutionary methods for design, optimization and control with applications to industrial problems (EUROGEN'01), pp 95–100