

A New Approach to Handle High Dimensional and Large Datasets in Multi-objective Evolutionary Fuzzy Systems

Michela Antonelli, Pietro Ducange, Francesco Marcelloni

Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni
University of Pisa
Pisa, Italy

{m1.antonelli, p.ducange, f.marcelloni@iet.unipi.it}

Abstract—In the framework of multi-objective evolutionary fuzzy systems (MOEFSs), the search space grows as the number of features of the dataset increases, leading to a slow and possibly difficult convergence of the evolutionary algorithm. Furthermore, mainly due to the fitness evaluation, datasets with a large number of instances require very high computational costs. In this paper, we propose a co-evolutionary approach to generate sets of Mamdani fuzzy rule-based systems (MFRBSs) with different trade-offs between accuracy and interpretability. We aim to deal with high dimensional and large datasets and to learn together the rule base (RB) and the membership function parameters. To reduce the search space, we perform the multi-objective evolutionary learning of the RB by selecting reduced sets of rules and conditions from a previously generated RB. Further, to lessen the computational costs, during the multi-objective evolutionary learning process, periodically, a single-objective genetic algorithm evolves a population of reduced training sets. We show the preliminary results obtained by applying our approach to two real world high dimensional and large regression datasets.

Keywords: *multi-objective evolutionary fuzzy systems; high dimensional datasets; large datasets; regression problems*

I. INTRODUCTION

Two main approaches can be exploited to design fuzzy rule-based systems (FRBSs): i) to elicit the knowledge from a human expert and ii) to extract this knowledge from available information (typically, input-output samples). In complex application domains, with a large quantity of information sources, the first approach is difficultly applicable. Thus, the second approach is often the unique alternative.

Several techniques have been proposed in the literature to generate FRBSs from available data [1]-[3]. Among these techniques, multi-objective evolutionary fuzzy systems (MOEFSs) have been successfully used in the last years to generate sets of FRBSs with different trade-offs between accuracy and interpretability [4][5]. MOEFSs exploit multi-objective evolutionary algorithms (MOEAs) to perform the learning [6][7] or the selection [8] of rules, and the tuning of the data base (DB) [9], with a prefixed DB and rule base (RB), respectively. Further, the most recent works describe

approaches to concurrently learn [10]-[13] or select [14]-[16] rules together with some elements of the DB, such as the membership function (MF) parameters and the granularities.

In real word applications, we often have to manage high-dimensional and large datasets. Let us consider, for instance, some applications such as environmental monitoring, traffic control or weather forecasting. Here, a large amount of sensors is typically used, thus generating data characterized by a large number of features. Further, if the sampling frequency is high, the resulting datasets comprise a very large amount of instances.

Dealing with high dimensional and large datasets is a challenge topic in the context of MOEFSs [4]. A high number of dimensions increases the search space, especially in rule learning approaches, thus affecting the performance of the evolutionary algorithm in terms of convergence and solution quality. A large number of instances, as mentioned in [4] and confirmed by the experimental results discussed in several papers [10]-[12], makes the computation of the fitness very time-consuming, thus hindering the full exploitation of the evolutionary process by obliging to limit the number of iterations.

Four main approaches have been followed to reduce these drawbacks in MOEFSs: i) to use data reduction techniques, such as feature selection [17][18] and instance selection [19][20], ii) to exploit parallel evolutionary algorithms [19][21], iii) to use some fitness approximation approaches [22], and iv) to exploit ad-hoc modified MOEAs [23][24].

In this contribution, we introduce a novel approach to generate sets of Mamdani FRBSs (MFRBSs) with different trade-offs between accuracy and RB complexity from high dimensional and large datasets. We aim to concurrently learn the RBs and the parameters of MFs in the same multi-objective evolutionary process.

To reduce the search space in the generation of the rules, we exploit the RB generation method recently introduced by us in [24]: RBs are determined by selecting a reduced number of rules and conditions during the evolutionary process from a previously generated RB. In [24], we used the Wang and

Mendel algorithm [1] and denoted the approach as two-level rule selection (2LRS): at the first level we select the rules and at the second level we select the conditions. In contrast with the RB learning process exploited in our previous works [7][10]-[12], the 2LRS approach allows considerably reducing the search space, especially when the number of features is high, though preserving some good characteristics of a learning approach (actually, once determined the rule at the first level, at the second level the selection of conditions appears more similar to a learning than a selection process). We experimentally verified that the new approach allows the specific MOEA used in the experiments, namely (2+2)M-PAES, to rapidly converge to a set of solutions with trade-offs between accuracy and complexity comparable to the ones computed by the RB learning process of our previous works. In particular, we experimentally verified that the (2+2)M-PAES with the 2LRS approach achieves after 50000 evaluations Pareto front approximations comparable to the ones achieved by the (2+2)M-PAES with the rule learning after 300000 evaluations.

Further, to deal with large datasets, we integrate the 2LRS in our co-evolutionary model proposed in [20]. Here, during the execution of the MOEA, periodically, a single-objective genetic algorithm (SOGA) evolves a population of reduced training sets (RTSs).

We have aim to verify whether the co-evolutionary approach allows generating a set of MFRBSs that are statistically equivalent to those generated using the overall training set (OTS). In order to prove the positive behaviour of the new co-evolutionary approach, we show some preliminary results on two real world high dimensional and large datasets.

The paper is organized as follows: Section II briefly describes the MFRBSs and introduces some notation. In Section III, we describe the co-evolutionary approach aimed at concurrently selecting training instances and performing multi-objective evolutionary learning of MFRBSs. Finally, Section IV shows the experimental results and Section V draws some final conclusions.

II. MAMDANI FUZZY SYSTEMS

Let $\mathbf{X} = \{X_1, \dots, X_f, \dots, X_F\}$ be the set of input variables and X_{F+1} be the output variable. Let U_f , with $f = 1, \dots, F+1$, be the universe of the f^{th} variable. Let $P_f = \{A_{f,1}, \dots, A_{f,T_f}\}$ be a strong fuzzy partition of T_f fuzzy sets on variable X_f . An MFRBS is composed of M rules expressed as:

$$\begin{aligned} R_m: \quad & \text{IF } X_1 \text{ is } A_{1,j_{m,1}} \text{ AND } \dots \text{ AND } X_F \text{ is } A_{F,j_{m,F}} \\ & \text{THEN } X_{F+1} \text{ is } A_{F+1,j_{m,F+1}} \quad (m = 1, \dots, M) \end{aligned} \quad (1)$$

The RB of an MFRBS can be completely described by a matrix $J \in \mathbb{N}^{M \times (F+1)}$, where the generic element $j_{m,f}$, with $m \in \{1, \dots, M\}$ and $f \in \{1, \dots, F+1\}$, indicates that fuzzy set $A_{f,j_{m,f}}$ has been selected for variable X_f in rule R_m . We use

the product and the centre of gravity method as AND logical operator and defuzzification method, respectively.

We adopt triangular fuzzy sets $A_{f,j}$ defined by the tuple $(a_{f,j}, b_{f,j}, c_{f,j})$, where $a_{f,j}$ and $c_{f,j}$ correspond to the left and right extremes of the support of $A_{f,j}$, and $b_{f,j}$ to the core. Since we use strong fuzzy partitions, we assume that $a_{f,1} = b_{f,1}$, $b_{f,T_f} = c_{f,T_f}$, and for $j = 2, \dots, T_f - 1$, $b_{f,j} = c_{f,j-1}$ and $b_{f,j} = a_{f,j+1}$.

To take the “don’t care” condition into account, a new fuzzy set $A_{f,0}$ ($f = 1, \dots, F$) is added to all the F input partitions P_f . This fuzzy set is characterized by an MF equal to 1 on the overall universe. The terms $A_{f,0}$ allow generating rules which contain only a subset of the input variables.

Given a set of N input observations $\mathbf{x}_n = [x_{n,1}, \dots, x_{n,F}]$, with $x_{n,f} \in \mathfrak{R}$, and the set of the corresponding outputs $x_{n,F+1} \in \mathfrak{R}$, $n = 1, \dots, N$, we aim to generate a set of MFRBSs with different trade-offs between accuracy and interpretability. As regards MFRBS interpretability, several definitions can be found in the literature [25][26]. In particular, MFRBS interpretability can be analyzed in terms of RB complexity and DB integrity. In this work, the DB integrity is ensured by using strong fuzzy partitions with a low level of granularity. Thus, for the sake of simplicity, we measure the interpretability in terms of only RB complexity.

III. THE PROPOSED APPROACH

In this work, we exploit both the effectiveness of the 2LRS in managing high dimensional datasets and the efficacy of the evolutionary training set selection (TSS) in reducing the computational complexity when dealing with large datasets. To this aim, we follow the same approach that we have described in [20], in which a co-evolutionary algorithm has been proposed to perform both the TSS and the multi-objective evolutionary RB learning. Unlike in [20], however, we adopt the 2LRS in place of the RB learning.

A. The Co-Evolutionary Algorithm

Fig. 1 shows the execution flow of the co-evolutionary algorithm by using the UML activity diagram notation. In the activity diagram, rectangles represent data objects, rounded rectangles represent activities and diamonds are used as decision and merge points. The algorithm handles two populations: the KB population, aimed at performing the 2LRS and the MF parameter learning, and the reduced training set (RTS) population, aimed at performing the TSS.

First of all, we apply the WM algorithm to the OTS to generate the set of candidate rules. We assume that the linguistic variables are uniformly partitioned with T_f fuzzy sets. Then, TSS is fulfilled by means of an SOGA while the 2LRS and the MF parameter learning are carried out by applying the (2+2)M-PAES algorithm.

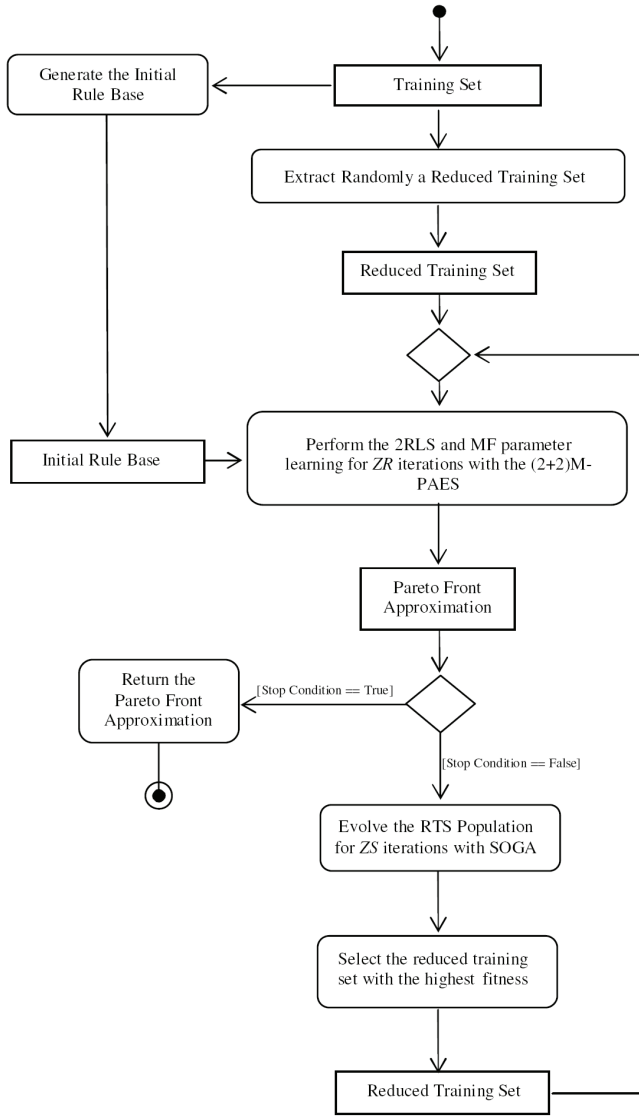


Figure 1. Activity diagram of the co-evolutionary approach

The SOGA and (2+2)M-PAES are cyclically executed one after the other. The RTS population evolves for a fixed number ZS of fitness evaluations in order to search for an RTS which is a good representative of the OTS. Thus, the RTS is used in the (2+2)M-PAES to perform the multi-objective 2RLS and MF parameters learning with different trade-offs between accuracy and RB complexity. (2+2)M-PAES is executed for a fixed number ZR of RB fitness evaluations. The resulting approximated Pareto front (set of non-dominated MFRBSs with different accuracy-complexity tradeoffs) is used in the execution of the SOGA to calculate the fitness function.

At the end of its execution, our algorithm returns a set of non-dominated MFRBSs in the complexity-accuracy plane. Even though several stop conditions can be considered, we terminate the execution of our algorithm when a prefixed number of KB fitness evaluations have been performed. In this way, we are able to carry out comparisons with different

techniques using the same number of KB fitness evaluations.

In the following, we introduce and discuss the 2LRS and the MF parameters learning, including their coding and the mating operators. Finally, we introduce the MOEA exploited in the evolutionary process.

B. The Multi-objective Evolutionary Two-Level Rule Selection and MF Learning

The 2LRS and the MF parameters learning are carried out by (2+2)M-PAES, which determines an approximation of the optimal Pareto front by concurrently minimizing the MSE and the RB complexity.

The MSE is calculated as:

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

where $|E|$ is the size of the dataset, $F(x^l)$ is the output obtained from the MFRBS when the l -th input pattern is considered, and y^l is the desired output. RB complexity is measured as sum of the conditions which compose the antecedents of the rules included in the RB. Thus, low values of complexity correspond to RB characterized by a low number of rules and a low number of input variables really used in each rule.

We indicate with J_{WM} and M_{WM} the RB generated by applying the WM algorithm to the OTS and the number of rules of this RB, respectively. We recall that the antecedent of each candidate rule has one condition for each linguistic variable, since the WM algorithm performs no selection of the conditions. Further, especially when dealing with large and high dimensional datasets, the WM algorithm could generate RBs composed by a huge number of rules. For this reason, in order to generate compact and interpretable RBs, we fix the maximum number M_{MAX} of rules which can compose the RB. Further, the minimum number of rules that must be present in the RB is fixed by the user-defined parameter M_{MIN} .

We exploit a chromosome C^{KB} composed of two parts (C_1^{KB}, C_2^{KB}) , which define the RB and the positions of the centroids of each fuzzy set in each linguistic variable, respectively. The C_1^{KB} part of the chromosome is a vector of M_{MAX} pairs $\mathbf{p}_m = (k_m, \mathbf{v}_m)$, where $k_m \in [0, \dots, M_{WM}]$ identifies the index of the rule in J_{WM} selected for the current RB and $\mathbf{v}_m = [v_{m,1}, \dots, v_{m,F}]$ is a binary vector which indicates, for each condition in the rule, if the condition is present or corresponds to a don't care. In particular, if $k_m = 0$, the m^{th} rule is not included in the RB. In this way we manage to generate RBs with a lower number of rules than M_{MAX} . Further, if $v_{m,f} = 0$, the f^{th} condition of the m^{th} rule is replaced by a don't care condition; otherwise it remains unchanged.

As regards C_2^{KB} , since we adopt strong fuzzy partitions with, for $j = 2, \dots, T_f - 1$, $b_{f,j} = c_{f,j-1}$ and $b_{f,j} = a_{f,j+1}$, in order to define each fuzzy set of the partition it is sufficient to fix the positions of the cores $b_{f,j}$ along the universe U_f of the f^{th} variable (we normalize each variable in $[0,1]$). Since $b_{f,1}$ and b_{f,T_f} coincide with the extremes of the universe, the partition of each linguistic variable X_f is completely defined by $T_f - 2$ parameters. Finally, C_2^{KB} consists of $F + 1$ vectors of $T_f - 2$ real numbers: the f^{th} vector contains the $[b_{f,2}, \dots, b_{f,T_f-1}]$ cores which define the positions of the MFs for the linguistic variable X_f .

To ensure a good integrity level of the MFs, in terms of order, coverage and distinguishability, $\forall j \in [2, T_f - 1]$, we force $b_{f,j}$ to vary in the definition interval $\left[b_{f,j} - \frac{b_{f,j} - b_{f,j-1}}{2}, b_{f,j} + \frac{b_{f,j+1} - b_{f,j}}{2} \right]$.

The evolutionary process is performed by means of the (2+2)M-PAES proposed in [7], where both crossover and mutation operators are independently applied to each part of the chromosome. In particular we apply the one-point crossover to the C_1^{KB} part and the BLX- α crossover, with $\alpha = 0.5$, to the C_2^{KB} part.

Let s_1 and s_2 be two selected parent chromosomes. The common gene for the one point crossover is chosen by extracting randomly a number in $[1, \rho_{MAX} - 1]$, where ρ_{MAX} is the maximum number of rules in s_1 and s_2 . The crossover point is always chosen between two rules and not within a rule. When we apply the one-point crossover to the RB part, we can generate an RB with one or more pairs of equal rules. In this case, we simply eliminate one of the rules from each pair setting the correspondent k_m to zero. This allows us to reduce the total number of rules.

As regards the mutation, for the C_1^{KB} part we have defined three operators. The first step of each operator is to randomly select a pair $\mathbf{p}_m$, i.e. a rule, in the chromosome. The first operator replaces the value of k_m in the selected pair with an integer value randomly generated in $[1, \dots, M_{WM}]$. The second operator removes a rule from the RB by setting k_m of the selected pair to zero. Finally, the third operator modifies the antecedent $\mathbf{v}_m$ of the selected rule by complementing each gene $v_{m,f}$ with a probability equal to P_{cond} ($P_{cond} = 2/F$ in the experiments). After applying the three mutation operators, we remove the duplicate rules in the RB.

The mutation operator applied to C_2^{KB} , first, randomly chooses a variable X_f , $f \in [1, F + 1]$, and a fuzzy set

$j \in [2, T_f - 1]$ and then replaces the value of $b_{f,j}$ with a value randomly chosen within the definition interval of $b_{f,j}$.

The crossover operators are independently applied to C_1^{KB} and C_2^{KB} with probabilities P_{c1}^{KB} and P_{c2}^{KB} , respectively. As regards mutation, the first, second and third operators are applied to C_1^{KB} with probabilities P_{add}^{KB} , $0.1 - P_{add}^{KB}$, and P_{mod}^{KB} , respectively. Finally, the mutation operator is applied to C_2^{KB} with probability P_{m2}^{KB} .

C. The Single-objective Training Set Selection

We split the OTS into B disjoint blocks: each block contains $I = N/B$ instances randomly extracted from the OTS. Each RTS consists of K blocks. Thus, in the SOGA we exploit a chromosome C^{TS} which codifies the K blocks included in the RTS: each gene is an integer which varies from 1 to B and represents the index of the block selected to be included in the RTS.

The aim of the evolutionary process is to find the K blocks which are the best representatives of the OTS. We measure the degree DR of representativeness as follows. First, for each solution C_i^{KB} ($i = 1, \dots, N^{KB}$, where N^{KB} is the number of solutions contained into the (2+2)M-PAES archive), we compute the $MSE/2$ by using the OTS and the RTSs codified in all chromosomes C_j^{TS} ($j = 1, \dots, N^{TS}$, where N^{TS} is the number of individuals in the SOGA population). We refer to these MSEs as $MSE_i/2$ and $MSE_{i,j}/2$, respectively. Then, we compute:

$$DR_j = \frac{1}{N^{KB}} \sum_{i=1}^{N^{KB}} \frac{|MSE_i/2 - MSE_{i,j}/2|}{\max(MSE_i/2, MSE_{i,j}/2)}. \quad (2)$$

The lower the value of DR_j , the more the RTS encoded in C_j^{TS} achieves performance similar to the OTS.

In order to perform the optimization of the TSS population, we exploit a classical SOGA. We start with a randomly generated initial population P_0 . At each iteration t , the fitness function DR_j of each individual j is evaluated and individuals are selected for reproduction by using the roulette wheel selection. To generate the offspring population Q_t of N^{TS} individuals both one-point crossover and mutation are used. The mutation operator, when selected, randomly chooses a gene in $[1, K]$ and changes the value of the gene with an integer randomly chosen in the interval $[1, B]$.

The one-point crossover could generate descendants with one or more pairs of genes which have the same value. This corresponds to have duplicate blocks in the K blocks which compose the RTS. If this occurs, we change randomly the value of one of the equal genes so as to eliminate duplicate blocks. In the experiments, we executed the one-point

crossover and the mutation operators with probabilities P_c^{TS} and P_m^{TS} , respectively.

The new population P_{t+1} coincides with Q_t except for the worst 5% individuals which are replaced with the best 5% individuals of P_t (in practice, we use an elitist strategy where the worst individuals of the offspring population are replaced with the best individuals of the parent population with a percentage of 5%).

IV. EXPERIMENTAL RESULTS

The method has been tested on two regression problems (available at <http://sci2s.ugr.es/keel/category.php?cat=reg#sub2>), namely Elevators and Ailerons. As shown in Table I, the datasets are characterized by a high number of both input variables and input/output patterns.

We compare the results obtained by the proposed co-evolutionary approach (denoted as PAES-SOGA(10%)), which uses the 10% of the OTS, with the ones obtained by executing the (2+2)M-PAES with the 2LRS on the OTS (denoted as PAES). We performed a five-fold cross validation and executed each fold six trials with different seeds (thirty trials in total).

Table II shows the values used for the parameters of PAES and of PAES-SOGA(10%) in the experiments. These parameters have been obtained after a long experimentation. In particular, the choice of the probability values derives from an accurate analysis carried out to balance exploitation and exploration during the evolutionary process.

As stopping condition, we use the maximum number of evaluations. This number of evaluations is fixed to 50000 for both the algorithms. In the case of PAES-SOGA(10%) we consider as total number of evaluations the ones performed on the RTS.

Since the generation of the RBs is based on the selection of rules and conditions from the initial RB computed by the WM algorithm, in Table III we show the mean values $\overline{MSE}_{TR}/2$ and $\overline{MSE}_{TS}/2$, and the standard deviations σ_{TR} and σ_{TS} of the MSE/2 computed on the training and test sets, respectively, and the mean number $\bar{R}$ and the standard deviation σ_R of the rules of the MFRBSs generated by the WM algorithm.

In order to analyze the effectiveness of the evolutionary process when the RTSs are used in place of the OTSs, we adopt the hypervolume indicator, computed by using the MSE/2 calculated on the OTS for both PAES and PAES-SOGA(10%). Thus, for the PAES-SOGA(10%), we re-compute the MSE/2 of the solutions contained in the final archive by using the OTS in place of the RTS. The new values of MSE/2 might make some solution dominated by other solutions in the archive. Thus, before computing the indicator, we select the non-dominated solutions so as to generate again a Pareto front approximation.

TABLE I. DATASETS USED IN THE EXPERIMENTS

Dataset name	Number of patterns	Number of input variables
Elevators (ELV)	16599	18
Ailerons (AIL)	13750	40

TABLE II. VALUES OF THE PARAMETERS USED IN THE EXPERIMENTS

Parameter	Value
AS (2+2)M-PAES archive size	64
N^{TS} SOGA population size	32
ZS Number of evaluations in the SOGA for each cycle	64
ZR Number of evaluations in the (2+2)M-PAES block	640
T_f Number of fuzzy sets in each variable $X_{f,f}$	3
M_{MIN} Minimum number of rules	5
M_{MAX} Maximum number of rules	30
P_{c1}^{KB} Probability of applying the crossover operator to C_1^{KB}	0.2
P_{c2}^{KB} Probability of applying the crossover operator to C_2^{KB}	0.5
P_{m1}^{KB} Probability of applying the mutation to C_1^{KB}	0.1
P_{add}^{KB} Probability of applying the first mutation operator to C_1^{KB}	0.07
P_{mod}^{KB} Probability of applying the third mutation operator to C_1^{KB}	0.7
P_{m2}^{KB} Probability of applying the mutation operator to C_2^{KB}	0.2
P_c^{TS} Probability of applying the crossover operator to C^{TS}	0.1
P_m^{TS} Probability of applying the mutation operator to C^{TS}	0.01
I Number of instances in a block	5

TABLE III. AVERAGE RESULTS ACHIEVED BY THE WM ALGORITHM

	$\overline{MSE}_{TR}/2$ (σ_{TR})	$\overline{MSE}_{TS}/2$ (σ_{TS})	$\bar{R}$ (σ_R)
ELV	1.584E-5 (0.024E-5)	1.596E-5 (0.076E-5)	537.7 (11.9)
AIL	3.19E-8 (0.19E-8)	3.26E-8 (0.23E-8)	1116.8 (51.3)

The hypervolume indicator, whose features and properties have been discussed in detail in [27], measures the hypervolume of the portion of the objective space that is weakly dominated by a Pareto front approximation. The computations of the hypervolume were performed using the performance assessment package provided in the PISA toolkit [28]. In this package, the value of the indicator is normalized in [0,1] and lower values are associated with better performances.

The analysis of the box plots, shown in Figs. 2 and 3, suggests that in both datasets the hypervolumes of the Pareto front approximations generated by the PAES-SOGA(10%) and PAES are quite similar. This is also confirmed by the results of the Wilcoxon signed-rank test [28] executed with significance level $\alpha = 0.05$. Indeed, the two distributions of the hypervolumes associated with the two algorithms result to be statistically equivalent. This means that the evolutionary process, in terms of both exploration and exploitation, is not affected by using the RTSs rather than the OTS.

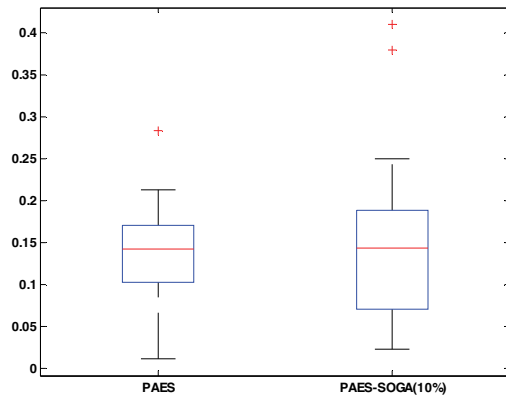


Figure 2. Box plots of the hypervolume indicators for the ELV dataset

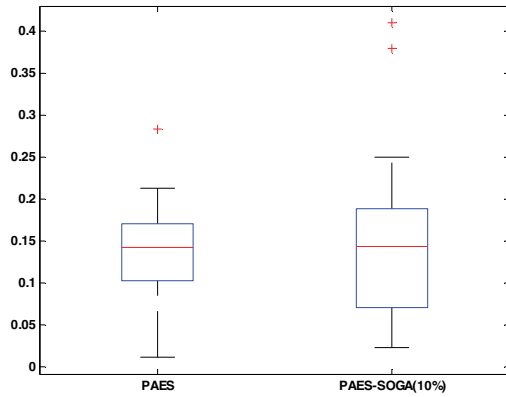


Figure 3. Box plots of the hypervolume indicators for the AIL dataset

To evaluate the MFRBSs generated by PAES and PAES-SOGA(10%) in terms of accuracy and complexity, we applied the procedure introduced in [10]. This procedure analyzes the mean and standard deviation of three representative points of the approximated Pareto front on the plane complexity-MSE/2, namely the first (the most accurate and complex), the median and the last (the least accurate and complex). This three points are referred as FIRST, MEDIAN and LAST, respectively. From an operative point of view, for each of the thirty trails, first we order the solutions of the approximated Pareto front for increasing MSE/2 values, then, we calculate the average values and the corresponding standard deviations over the thirty trials, of the MSE/2 and complexity values related to the three points for both the training and test sets.

In Figures 4 and 5, we plot the average values of the three points for ELV and AIL datasets, respectively, both on the training and test sets. We observe that there do not exist substantial differences between the training and test sets. This highlights that both the algorithms do not suffer from the overfitting problem. This result is particularly significant for PAES-SOGA(10%) which uses an RTS for most of the evolutionary optimization process. Further, though the solutions generated by PAES-SOGA(10%) are characterized by approximately the same MSE/2 as the ones generated by PAES, they result to be generally less complex.

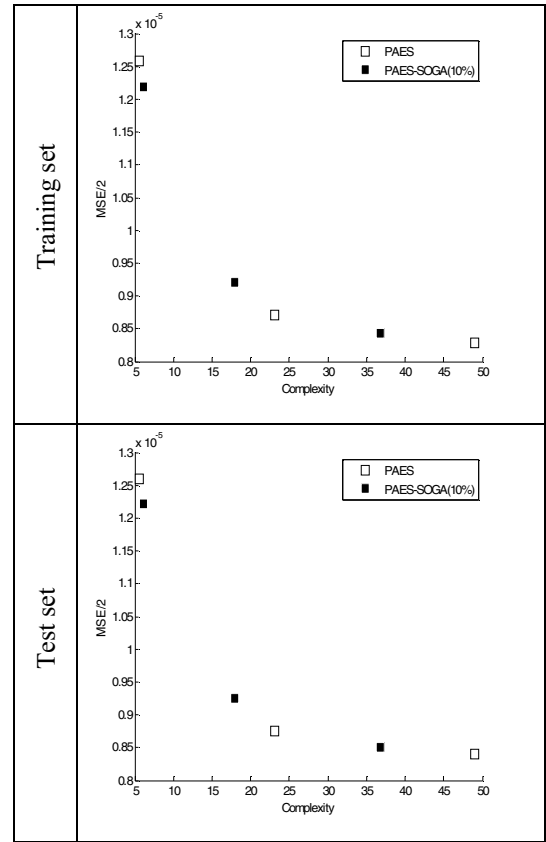


Figure 4. Average values of the FIRST, MEDIAN and LAST solutions for the ELV dataset

Tables IV and V show the mean values and the standard deviations of the MSE/2 on both the training ($\overline{MSE_{TR}}/2$ and σ_{TR}) and test ($\overline{MSE_{TS}}/2$ and σ_{TS}) sets, the average complexity ($\overline{COMP}$ and σ_{COMP}), the average number of rules ($\overline{RULES}$ and σ_{RULES}) and the average number of input variables used in the final KBs ($\overline{VARS}$ and σ_{VARS}). A careful analysis of these tables confirms the great similarity of the two Pareto front approximations generated by PAES and PAES-SOGA(10%).

In order to assess if there exist statistical differences among the MSE/2 corresponding to the FIRST, MEDIAN and LAST solutions generated by the PAES-SOGA(10%) and PAES, we applied the Wilcoxon signed-rank test with 95% confidence (columns **W-t_{TR}** and **W-t_{TS}**).

The interpretation of the **W-t_{TR}** and **W-t_{TS}** columns are as follows:

- * represents the best result (the most accurate);
- + means that the best result has better performance than that of the corresponding row;
- = means that the best result has performance comparable to that of the corresponding row.

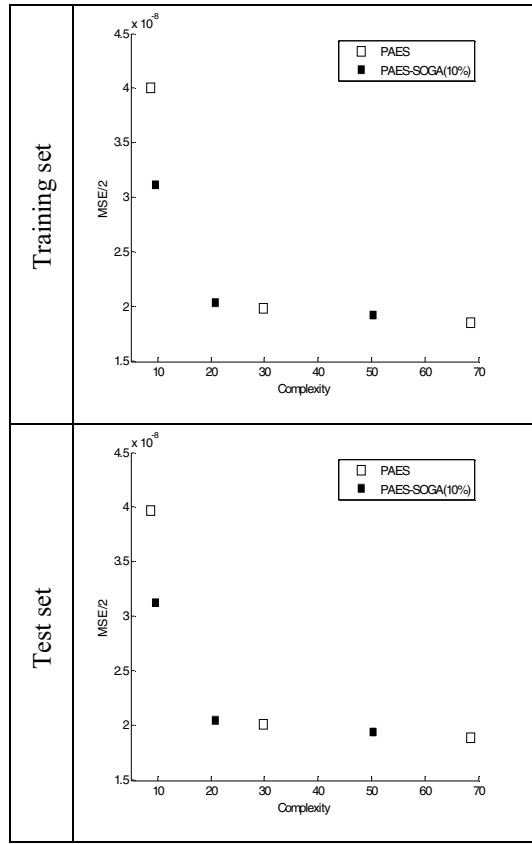


Figure 5. Average values of the FIRST, MEDIAN and LAST solutions for the AIL dataset

The results of the test confirm that choosing a specific solution in a Pareto front approximation generated by PAES is not different from choosing a corresponding solution in a Pareto front approximation obtained by the PAES-SOGA(10%), at least in terms of MSE/2. Also, typically, the solutions generated by the PAES-SOGA(10%) are characterized by a lower complexity.

TABLE IV. AVERAGE RESULTS ACHIEVED BY THE DIFFERENT MOEAS IN CORRESPONDENCE TO THE FIRST, MEDIAN AND LAST SOLUTIONS FOR THE ELV DATASET

		$\overline{MSE}_{TR} / 2$ (σ_{TR})	$W-t_{TR}$	$\overline{MSE}_{TS} / 2$ (σ_{TS})	$W-t_{TR}$	$Comp$ (σ_{Comp})	R (σ_R)
FIRST	PAES-SOGA(10%)	0.843E-5 (0.040E-5)	=	0.856E-5 (0.053E-5)	=	41.5 (14.9)	10.9 (2.5)
	PAES	0.824E-5 (0.049E-5)	*	0.846E-5 (0.053E-5)	*	48.1 (13.1)	11.4 (2.3)
MEDIAN	PAES-SOGA(10%)	0.899E-5 (0.087E-5)	=	0.915E-5 (0.090E-5)	=	20.5 (6.2)	9.0 (2.0)
	PAES	0.867E-5 (0.045E-5)	*	0.882E-5 (0.051E-5)	*	23.1 (5.4)	9.4 (1.8)
LAST	PAES-SOGA(10%)	1.233E-5 (0.110E-5)	=	1.253E-5 (0.121E-5)	=	5.3 (1.4)	5.0 (0.0)
	PAES	1.231E-5 (0.171E-5)	*	1.250E-5 (0.155E-5)	*	5.9 (3.9)	5.0 (0.0)

Finally, by analyzing the execution times of the two algorithms, we realized that using an RTS that contains the

10% of the overall training set allows us to save on average the 86% of the execution time.

TABLE V. AVERAGE RESULTS ACHIEVED BY THE DIFFERENT MOEAS IN CORRESPONDENCE TO THE FIRST, MEDIAN AND LAST SOLUTIONS FOR THE AIL DATASET

		$\overline{MSE}_{TR} / 2$ (σ_{TR})	$W-t_{TR}$	$\overline{MSE}_{TS} / 2$ (σ_{TS})	$W-t_{TR}$	$Comp$ (σ_{Comp})	R (σ_R)
FIRST	PAES-SOGA(10%)	1.92E-8 (0.10E-8)	=	1.94E-8 (0.11E-8)	=	50.3 (31.7)	7.9 (1.75)
	PAES	1.85E-8 (0.09E-8)	*	1.89E-8 (0.11E-8)	*	68.5 (49.5)	9.2 (2.4)
MEDIAN	PAES-SOGA(10%)	2.04E-8 (0.15E-8)	=	2.05E-8 (0.18E-8)	=	20.7 (13.8)	6.2 (0.9)
	PAES	1.99E-8 (0.18E-8)	*	2.02E-8 (0.19E-8)	*	29.7 (19.2)	6.9 (1.0)
LAST	PAES-SOGA(10%)	3.12E-8 (0.77)	*	3.13E-8 (0.82E-8)	*	9.6 (11.7)	5.0 (0.1)
	PAES	4.00E-8 (2.57E-8)	=	3.97E-8 (2.46E-8)	=	8.7 (9.8)	5.0 (0.0)

V. CONCLUSION

In this paper, we have proposed a novel approach to deal with high dimensional and large datasets in the framework of multi-objective evolutionary learning of Mamdani Fuzzy Rule-based Systems (MFRBSs) with different trade-offs between accuracy and interpretability.

The proposed algorithm is based on a co-evolutionary approach which allows us to concurrently perform an evolutionary training set selection (TSS) and a multi-objective evolutionary learning of the rule base (RB) and of the membership functions parameters. The RB is built by selecting first a reduced number of rules from a previously generated rule base, and then a reduced number of conditions for each selected rule. In this way, we are able to improve the convergence ability of the learning process in terms of speed and quality of the generated solutions. Furthermore, the evolutionary TSS allows us to save considerable execution time.

We experimented our method (PAES-SOGA) on two real world regression problems, characterized by a high number of both input variables and input/output patterns. The obtained results have been compared with the ones obtained by the same learning method applied using the overall dataset (PAES). By using statistical tests, we have shown that the Pareto front approximations are statistically equivalent in terms of the hypervolume indicator and three representative points. Furthermore, though the PAES-SOGA uses a reduced training set, the algorithm does not suffer from overfitting. Finally, we have observed that, the use of a training set containing the 10% of the overall training set allows us on average to save the 86% of the execution time needed when the overall training set is used.

REFERENCES

- [1] L.X Wang, and J.M. Mendel, "Generating fuzzy rules by learning from examples," IEEE Trans. Syst. Man. Cybern., vol. 22, n. 6, pp. 1414–1427, 1992.

- [2] O. Cordón, M.J. del Jesus, F. Herrera, and M. Lozano, "MOGUL: a methodology to obtain genetic fuzzy rule-based systems under the iterative rule learning approach," *Int. J. Intell. Syst.*, vol. 14, n. 11, pp. 1123–1153, 1999.
- [3] J. Casillas, O. Cordón, and F. Herrera, "COR: a methodology to improve ad hoc data-driven linguistic rule learning methods by inducing cooperation among rules," *IEEE Trans. Syst. Man. Cybern.*, vol. 32, pp. 526–537, 2002.
- [4] F. Herrera, "Genetic fuzzy systems: taxonomy, current research trends and prospects," *Evolutionary Intelligence*, vol. 1, pp. 27–46, 2008.
- [5] H. Ishibuchi, and Y. Nojima, "Multiobjective genetic fuzzy systems," *Computational Intelligence*, vol. 1, C.L. Mumford and L.C. Jain, Ed. Berlin: Springer, 2009, pp 131–173.
- [6] P. Ducange, B. Lazzerini, and F. Marcelloni, "Multi-objective Genetic Fuzzy Classifiers for Imbalanced and Cost-sensitive Datasets," *Soft Computing*, vol. 14, n. 10, pp. 713–728, 2010.
- [7] M. Cococcioni, P. Ducange, B. Lazzerini, and F. Marcelloni, "A Pareto-Based Multi-objective Evolutionary Approach to the Identification of Mamdani Fuzzy Systems," *Soft Computing*, vol. 11, n. 11, pp. 1013–1031, 2007.
- [8] H. Ishibuchi, and T. Yamamoto, "Fuzzy Rule Selection by Multi-Objective Genetic Local Search Algorithms and Rule Evaluation Measures in Data Mining," *Fuzzy Sets and Systems*, vol. 141, pp. 59–88, 2004.
- [9] A. Botta, B. Lazzerini, F. Marcelloni, and D. Stefanescu, "Context Adaptation of Fuzzy Systems Through a Multi-Objective Evolutionary Approach Based on a Novel Interpretability Index," *Soft Computing*, vol. 13, n. 5, pp. 437–449, 2009.
- [10] R. Alcalá, P. Ducange, F. Herrera, B. Lazzerini, and F. Marcelloni, "A Multi-objective Evolutionary Approach to Concurrently Learn Rule and Data Bases of Linguistic Fuzzy Rule-Based Systems," *IEEE Trans. Fuzzy. Syst.*, vol. 17, n. 5, pp 1106–1122, 2009.
- [11] M. Antonelli, P. Ducange, B. Lazzerini, and F. Marcelloni, "Multi-objective evolutionary learning of granularity, membership function parameters and rules of Mamdani fuzzy systems," *Evolutionary Intelligence*, vol. 2, n. 1-2, pp. 21–37, 2009.
- [12] M. Antonelli, P. Ducange, B. Lazzerini, and F. Marcelloni, "Learning Concurrently Partition Granularities and Rule Bases of Mamdani Fuzzy Systems in a Multi-objective Evolutionary Framework," *Int. J. Approx. Reason.*, vol. 50, n. 7, pp. 1066–1080, 2009.
- [13] P. Pulkkinen and H. Koivisto, "A Dynamically Constrained Multiobjective Genetic Fuzzy System for Regression Problems," *IEEE Trans. Fuzzy. Syst.*, vol. 18, n. 1, pp. 161–177, 2010.
- [14] R. Alcalá, M.J. Gacto, F. Herrera, and J. Alcalá-Fdez, "A Multi-objective Genetic Algorithm for Tuning and Rule Selection to Obtain Accurate and Compact Linguistic Fuzzy Rule-Based Systems," *Int. J. Uncertain. Fuzz.*, vol. 15, n. 5, 2007, pp. 539–557.
- [15] M.J. Gacto, R. Alcalá, and F. Herrera, "Adaptation and Application of Multi-Objective Evolutionary Algorithms for Rule Reduction and Parameter Tuning of Fuzzy Rule-Based Systems," *Soft Computing*, vol. 13, n. 5, pp. 419–436, 2009.
- [16] M. J. Gacto, R. Alcalá, and F. Herrera, "Integration of an Index to Preserve the Semantic Interpretability in the Multi-Objective Evolutionary Rule Selection and Tuning of Linguistic Fuzzy Systems," *IEEE Trans. Fuzzy. Syst.*, vol. 18, n.3, pp. 515–531, 2010.
- [17] J. Casillas, O. Cordón, M.J. del Jesus, and F. Herrera, "Genetic feature selection in a fuzzy rule-based classification system learning process," *Information Sciences*, vol. 136, pp.135–157, 2001.
- [18] O. Cordón, and A. Quirin, "Comparing Two Genetic Overproduce-and-choose Strategies for Fuzzy Rule-based Multiclassification Systems Generated by Bagging and Mutual Information-based Feature Selection", *International Journal of Hybrid Intelligent Systems*, vol. 7, pp.45–64, 2010
- [19] Y. Nojima and H. Ishibuchi, "Effects of data reduction on the generalization ability of parallel distributed genetic fuzzy rule selection," in: *Proc. of 9th International Conference on Intelligent Systems Design and Applications*, Pisa (Italy), 30 Nov. – 2 Dec., 2009, pp. 96–101.
- [20] M. Antonelli, P. Ducange, F. Marcelloni, "Exploiting a Coevolutionary Approach to Concurrently Select Training Instances and Learn Rule Bases of Mamdani Fuzzy Systems," in: *Proc. of the WCCI 2010 International Conference*, Barcelona (Spain), 18 – 23 Jul., 2010, pp. 1366–1372.
- [21] M.A. de Vega, J.M. Bardallo, F.A. Marquez, and A. Peregrin, "Parallel Distributed Two-level Evolutionary Multiobjective Methodology for Granularity Learning and Membership Functions Tuning in Linguistic Fuzzy Systems," in: *Proc. of 9th International Conference on Intelligent Systems Design and Applications*, Pisa (Italy), 30 Nov. –2 Dec., 2009, pp. 134–139
- [22] M. Cococcioni, B. Lazzerini, and F. Marcelloni, "On reducing computational overhead in multi-objective genetic Takagi–Sugeno fuzzy systems," *Applied Soft Computing*, vol. 11, n. 1, pp. 675–688, 2011.
- [23] J. R. Alcalá, M.J. Gacto, F. Herrera, "A Fast and Scalable Multi-Objective Genetic Fuzzy System for Linguistic Fuzzy Modeling in High-Dimensional Regression Problems," *IEEE Transactions on Fuzzy Systems* doi: 10.1109/TFUZZ.2011.2131657, in press (2011)
- [24] M. Antonelli, P. Ducange, B. Lazzerini, F. Marcelloni, "Multi-objective Evolutionary Generation of Mamdani Fuzzy Rule-Based Systems based on Rule and Condition Selection," in: *Proc. of 5th IEEE International Workshop on Genetic and Evolutionary Fuzzy Systems*, Paris (France), 11 – 15 Apr., 2011, in press.
- [25] J. M. Alonso, L. Magdalena, and G. González-Rodríguez, "Looking for a good fuzzy system interpretability index: an experimental approach," *Int. J. Approx. Reason.*, vol. 51, n. 1, pp. 115–134, 2009.
- [26] S. M. Zhou, and J. Q. Gan, "Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modelling," *Fuzzy Sets and Systems*, vol. 159, pp. 3091–3131, 2008.
- [27] E. Zitzler, L. Thiele, M. Laumanns, C.M. Fonseca, and V.G. de Fonseca, "Performance assessment of multiobjective optimizers: An analysis and review," *IEEE Transactions on Evolutionary Computation*, vol. 7, n. 2, pp. 117–132, 2003.
- [28] S. Bleuler, M. Laumanns, L. Thiele, E. Zitzler, "PISA –platform and programming language independent interface for search algorithms," in: C.M. Fonseca et al. (Eds.), *Evolutionary Multi-Criterion Optimization (EMO 2003)*, LNCS, vol. 2632, pp. 494–508, Springer-Verlag, Heidelberg, 2003.