Genetic Training Instance Selection in Multiobjective Evolutionary Fuzzy Systems: A Coevolutionary Approach

Michela Antonelli, Pietro Ducange, *Member, IEEE*, and Francesco Marcelloni, *Member, IEEE*

***Abstract*—When dealing with datasets that are characterized by a large number of instances, multiobjective evolutionary learning (MOEL) of fuzzy rule-based systems (FRBSs) suffers from high computational costs, mainly because of the fitness evaluation. The use of a reduced set of representative instances in place of the overall training set (TS) would considerably lessen the computational effort. Even though a large number of papers have proposed instance selection approaches, mainly in classification problems, how this selection should be performed, especially in the context of regression, is still an open issue. In this paper, we tackle the instance selection problem in the framework of MOEL of FRBSs through a coevolutionary approach. In the execution of the MOEL, periodically, a single-objective genetic algorithm (SOGA) evolves a population of reduced TSs. The SOGA aims to maximize a purposely defined index which measures how much the Pareto fronts computed by using, respectively, the reduced TS and the overall TS are close to each other: The closer the fronts, the more the reduced TS is representative of the overall TS. During the execution of the MOEL, the rule base and the membership function parameters of the fuzzy sets are concurrently learned by maximizing the accuracy and minimizing the complexity. We tested our approach on 12 large datasets. We adopted reduced TSs composed of 5%, 10%, and 20% of the overall TS. Using nonparametric statistical tests, we verified that with 10% and 20% of the overall TS, the Pareto front approximations that are generated by our coevolutionary approach are comparable with the ones generated by applying the MOEL with the overall TS, although the coevolution allows us to save up to 86.36% of the execution time. In addition, the analysis of the behavior of three representative solutions on the test set highlights that the use of the reduced TSs does not affect the generalization capabilities of the generated FRBSs.**

***Index Terms*—Large datasets, multiobjective evolutionary fuzzy systems (MOEFS), regression problems, training set selection.**

I. Introduction

During the past few years, multiobjective evolutionary algorithms (MOEAs) have been so extensively used to learn or optimize the structural parameters of fuzzy rule-based systems (FRBSs) [1], [2] that the term multiobjective evolutionary fuzzy systems (MOEFSs) [2] has been coined to identify the hybridization of FRBSs with MOEAs. MOEAs have been widely employed to generate FRBSs with different tradeoffs between accuracy and interpretability [3]–[16]. From the interpretability point of view, Mamdani-type FRBSs (MFRBSs) have attracted a lot of attention, since they are composed of a completely linguistic rule base (RB). The meaning of the linguistic terms used in the rules is defined by the fuzzy sets that are contained in the database (DB) of the MFRBS. RB and DB form the knowledge base (KB) of the MFRBS.

In the first generation of MOEFSs, MOEAs have been used to perform the selection [3], [4] or the learning [5]–[7] of rules and the tuning of the DB [8], with a prefixed DB and RB, respectively. Recently, MOEAs have been employed to concurrently learn [9]–[13] or select [14]–[16] rules together with some elements of the DB, such as the membership function (MF) parameters and the granularities of the fuzzy partitions. As stated in [1] and confirmed by the experimental results discussed in several papers [9]–[11], one of the major drawbacks in applying MOEAs to the FRBS generation is the computation time required by the fitness evaluation, especially when dealing with datasets with a large number of instances. This hinders the full exploitation of the evolutionary process by obliging the designer to limit the number of iterations.

In the context of evolutionary learning of FRBSs, three main approaches have been proposed to tackle this drawback, namely parallel implementation of the evolutionary algorithms, fitness approximation approaches, and data reduction techniques. Parallel implementations are usually based on spatial structures, such as island and cellular models. For instance, in [17] and [18], the scalability of genetic fuzzy rule selection, in classification problems, is improved by dividing the training set (TS) into subsets and assigning each subset to a subpopulation allocated to each of the available CPUs of a multicore workstation. Each individual in a subpopulation is evaluated using only the assigned training subset. Parallel implementations of genetic algorithms have also been used for tuning the MFs that are associated with the linguistic terms of a predefined DB [19] and to learn concurrently the RB and the granularity of the MFs [20].

Fitness approximation reduces the execution time by computing an approximation of the fitness function rather than by evaluating the fitness function itself [21]. This technique has been widely investigated in the framework of classical evolutionary optimization, but only a few papers have discussed its use in MOEFSs. For instance, in [22] and [23], a simple approach to perform a fast, though approximate, identification of the consequent parameters has been employed in the multiobjective evolutionary learning (MOEL) of Takagi–Sugeno FRBSs.

Manuscript received January 26, 2011; revised June 21, 2011; accepted August 31, 2011. Date of publication October 26, 2011; date of current version April 4, 2012.

The authors are with the Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni, University of Pisa, Pisa I-56100, Italy (e-mail: michela.antonelli@iet.unipi.it; pietro.ducange@iet.unipi.it; francesco.marcelloni@iet.unipi.it).

Digital Object Identifier 10.1109/TFUZZ.2011.2173582

Data reduction speeds up the evaluation of the fitness function by using a reduced TS. In MOEFS, however, to the best of our knowledge, data reduction has been discussed only in [24], where the authors analyze the effects of three different types of pseudorandom data extraction on the generalization ability of fuzzy rule-based classifiers.

Instance selection (IS) is one of the most used techniques in the context of data reduction [25]. There are two main strategies in IS: prototype selection and TS selection. Prototype selection performs an instance removal from the TS to retain only those critical instances which allow the 1-NN classifier to achieve the maximum classification rate. In some experimental results, prototype selection has proved to decrease storage and classification time cost, as well as to increase classification accuracy. TS selection aims to select a reduced set of representative instances: unlike prototype selection, which is specifically targeted to the nearest neighbor classification, the representative instances that are extracted by TS selection are used in different machine learning algorithms for different applications, such as regression, classification, subgroup discovery, and clustering.

In the literature, there exist a number of surveys [25]–[27] which discuss the most important and interesting approaches to IS. In general, the majority of these approaches focus on prototype selection. Furthermore, the few approaches taking TS selection into consideration deal almost exclusively with classification problems. In the past years, evolutionary algorithms have been extensively used to perform both prototype [28]–[32] and TS selections [29], [33]–[36]. In most cases, evolutionary IS outperforms classical IS. In [29], evolutionary IS for both prototype and TS selections is discussed, using the 1-NN and the C4.5 algorithms, respectively, and a comparison between evolutionary and nonevolutionary IS approaches is performed. The comparison highlights that evolutionary IS outperforms nonevolutionary IS in terms of both instance reduction rate and classification accuracy. In [31] and [35], the evolutionary IS proposed in [29] has been improved by introducing the concept of data stratification. With regard to evolutionary TS selection, some contributions can be found in the field of neural networks [33], [34] and subgroup discovery [36].

In this paper, we focus on regression problems [37]–[41]. To the best of our knowledge, the work discussed in [42] is the only attempt of applying IS to regression problems. Unlike our approach, here, the authors exploit the concept of mutual information to decide which prototypes should be included in the reduced TS: First, they calculate, for each instance of the TS, the value of the mutual information when the instance is removed from the TS; then, this value is compared with the values of the mutual information of its neighbors so as to decide if the instance must be included or not in the reduced TS. The method is mainly used to filter noisy instances from the original TS. Indeed, it is not tested on large datasets but on very simple regression datasets with the objective of evaluating the capability of the approach in generating robust approximation models using noisy training inputs. Because of its high computational complexity, the method is not suitable for selecting the most representative instances of a large TS.

We propose a particular evolutionary IS approach in the framework of MOEL of a set of KBs of MFRBSs characterized by different tradeoffs between accuracy and RB complexity. We exploit a coevolutionary approach: The MOEL and a single-objective genetic algorithm (SOGA) aimed at selecting the reduced TS are cyclically executed one after the other. To experimentally prove the effectiveness of our evolutionary IS, we adopt the MOEL that we have proposed and successfully experimented in our previous works [6], [9]–[12]. To perform the IS, we split the TS into disjoint blocks: each block contains randomly chosen instances. A reduced TS consists of a combination of a prefixed number of blocks. The SOGA evolves integer chromosomes, where each gene codifies the index of a block: the length of the chromosome determines the size of the reduced TS. The SOGA is executed periodically for a prefixed low number of iterations by using, as fitness function, the average of the absolute values of the normalized differences between the accuracies that are computed with the overall TS and the reduced TSs, respectively, on all the MFRBSs of the Pareto front generated so far by the MOEL. The best individual of the SOGA is used to execute the MOEL for a prefixed number of iterations. Thus, MOEL and SOGA evolve by searching synergically for the Pareto front with the best tradeoffs between accuracy and complexity, and for the reduced TS which performs similarly to the overall TS, respectively.

We tested our approach on 12 datasets. In order to evaluate which percentage of the overall TS could be sufficient to achieve Pareto front approximations that are comparable with the ones obtained by using the overall TS, we used reduced TSs composed of 5%, 10%, and 20% of the overall TS. We applied nonparametric statistical tests for multiple comparisons on the mean values of epsilon dominance, hypervolume, and accuracy on the test set obtained on all the datasets. We show that there do not exist statistical differences between the Pareto front approximations that are generated, respectively, by the coevolutionary approach with 10% and 20% of the overall TS and by the MOEL with the overall TS, although the use of the reduced TSs allows us to save up to 86.36% of the execution time.

This paper is organized as follows. Section II briefly describes the MFRBSs and introduces some notation. In Section III, we describe the coevolutionary approach aimed at concurrently selecting training instances and performing MOEL of MFRBSs. Finally, Section IV shows the experimental results, and Section V draws some final conclusions.

II. Mamdani Fuzzy Rule-Based Systems

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

$$R_m\text{: } \mathbf{IF}\ X_1\ \mathbf{is}\ A_{1,j_{m,1}}\ \mathbf{AND} \ldots \mathbf{AND}\ X_F\ \mathbf{is}\ A_{F,j_{m,F}}$$
$$\mathbf{THEN}\ X_{F+1}\ \mathbf{is}\ A_{F+1,j_{m,F+1}} (m = 1, \ldots, M) \quad (1)$$

where $j_{m,f} \in [1, T_f]$, $f = 1, \ldots, F+1$, identifies the index of the fuzzy set (among the T_f fuzzy sets of partition P_f), which has been selected for X_f in rule R_m.

We adopt triangular fuzzy sets $A_{f,j}$ that are 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. Further, we assume that $a_{f,1} = b_{f,1}$, $b_{f,T_f} = c_{f,T_f}$, and for $j = 2 \ldots T_f - 1$, $b_{f,j} = c_{f,j-1}$ and $b_{f,j} = a_{f,j+1}$.

To take the "don't care" condition into account, a new fuzzy set $A_{f,0}$ is added to all the F input partitions P_f, $f = 1, \ldots, F$. This fuzzy set is characterized by an MF equal to 1 on the overall universe [4]. The terms $A_{f,0}$ allow us to generate rules which contain only a subset of the input variables. It follows that $j_{m,f} \in [0, T_f]$, $f = 1, \ldots, F$, and $j_{m,F+1} \in [1, T_{F+1}]$. Thus, the RB of an MFRBS can be completely described by a matrix $J \in \mathbb{N}^{M \times (F+1)}$

$$J = \begin{bmatrix} j_{1,1} & \cdots & j_{1,F} & j_{1,F+1} \\ \cdots & \cdots & \cdots & \cdots \\ j_{m,1} & \cdots & j_{m,F} & j_{m,F+1} \\ \cdots & \cdots & \cdots & \cdots \\ j_{M,1} & \cdots & j_{M,F} & j_{M,F+1} \end{bmatrix} \tag{2}$$

where the generic element (m, f) indicates that fuzzy set $A_{f,j_{m,f}}$ has been selected for variable X_f in rule R_m.

We adopt the product and the centre of gravity method as AND logical operator and defuzzification method, respectively. Since we search for compact RBs with a reduced number of rules and of conditions in the antecedents, it is possible that the number of distinct labels used for one variable in the RB is lower than its granularity. Thus, it might occur that some input activates no rule and, therefore, results to be "covered" by no rule. In these cases, we adopt the inference strategy proposed in [43], which determines an output for a noncovered input based on the two closest rules to the input. The distance between the point and the rules is calculated considering the cores of the labels used in the rules.

III. Coevolutionary Approach

In this section, we introduce our approach to concurrently perform an MOEL process together with the selection of a reduced TS. As stated in Section I, when dealing with large datasets, i.e., with datasets with a large number of instances, the computational cost of the fitness evaluation during the evolutionary learning results to be very high. Even though the proposed approach can be used in any kind of MOEL, we focus our attention on MOEFSs and, in particular, on learning concurrently the RBs and some parameters of the DBs of MFRBSs in the same multiobjective evolutionary process. We adopt the MOEL algorithms that have been introduced in our previous works in this field [6], [9]–[12].

Given a set of N input observations $\mathbf{x}_n = [x_{n,1}, \ldots, x_{n,F}]$, with $x_{n,f} \in \Re$, and the set of the corresponding outputs $x_{n,F+1} \in \Re$, $n = 1, \ldots, N$, we aim to generate a set of KBs of MFRBSs with different tradeoffs between accuracy and RB complexity. In particular, we aim to identify the parameters of the MFs together with the learning of the rules. Regarding the granularity of the linguistic partitions, we decided to use a low number of fuzzy sets (three fuzzy sets for each linguistic variable in our experiments), both to ensure a high partition integrity level and to reduce the search space. Furthermore, we verified that the results obtained show a comparable accuracy level with the ones we discussed in our previous works [9]–[12].

We propose a sort of cooperative coevolutionary approach to concurrently perform TS selection and MOEL of KBs. TS selection is carried out by using an SOGA, whereas MOEL of KBs is achieved by applying the same MOEA that we proposed in [6]. Since this MOEA is a modified version of the classical (2+2)PAES, we denoted it as (2+2)M-PAES, where M stands for modified. Recently, we have successfully used (2+2)M-PAES in a number of our works [9]–[12]. SOGA and (2+2)M-PAES are cyclically executed one after the other. The SOGA population evolves for a fixed number E_{SO} of fitness evaluations in order to search for a reduced TS, which is a good representative of the TS. Then, the reduced TS with the highest fitness value is used in the (2+2)M-PAES to perform the multiobjective learning of a set of KBs with different tradeoffs between accuracy and RB complexity. (2+2)M-PAES is executed for a fixed number E_{MO} of KB fitness evaluations. The resulting approximated Pareto front (set of nondominated KBs with different accuracy–complexity tradeoffs) is used in the execution of the SOGA to calculate the fitness function.

At the end of its execution, the coevolutionary approach returns a set of nondominated KBs in the complexity–accuracy plane. In Fig. 1, we show the execution flow of our algorithm by using the unified modeling language activity diagram notation. In the activity diagram, rectangles and rounded rectangles represent data objects and activities, respectively, and diamonds are used as decision and merge points. Even though several stop conditions can be considered, we terminate the execution of our algorithm when a prefixed number of KB fitness evaluations (fitness used in the MOEL) 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 describe in detail the multiobjective evolutionary framework for the KB learning and the SOGA for the TS selection.

A. Multiobjective Evolutionary Knowledge-Base Learning

In our previous works [6], [9]–[12], we have proposed a multiobjective evolutionary framework to generate a set of MFRBSs from numerical data with different tradeoffs between accuracy and interpretability. With regard to MFRBS interpretability, several definitions can be found in the literature [44]–[47]. In particular, MFRBS interpretability can be analyzed in terms of RB complexity and DB integrity [11], [12]. In this paper, the DB integrity is ensured by using uniform partitions with a low level of granularity. Thus, for the sake of simplicity, we measure the interpretability in terms of only RB complexity.

To perform the RB learning together with the MF parameters learning, we exploit a chromosome C^{KB} composed of two parts (i.e., $C_1^{\mathrm{KB}}, C_2^{\mathrm{KB}}$), which define the RB and the positions of the centroids of each fuzzy set in each linguistic variable, respectively. In particular, C_1^{KB} encodes the RB by considering that

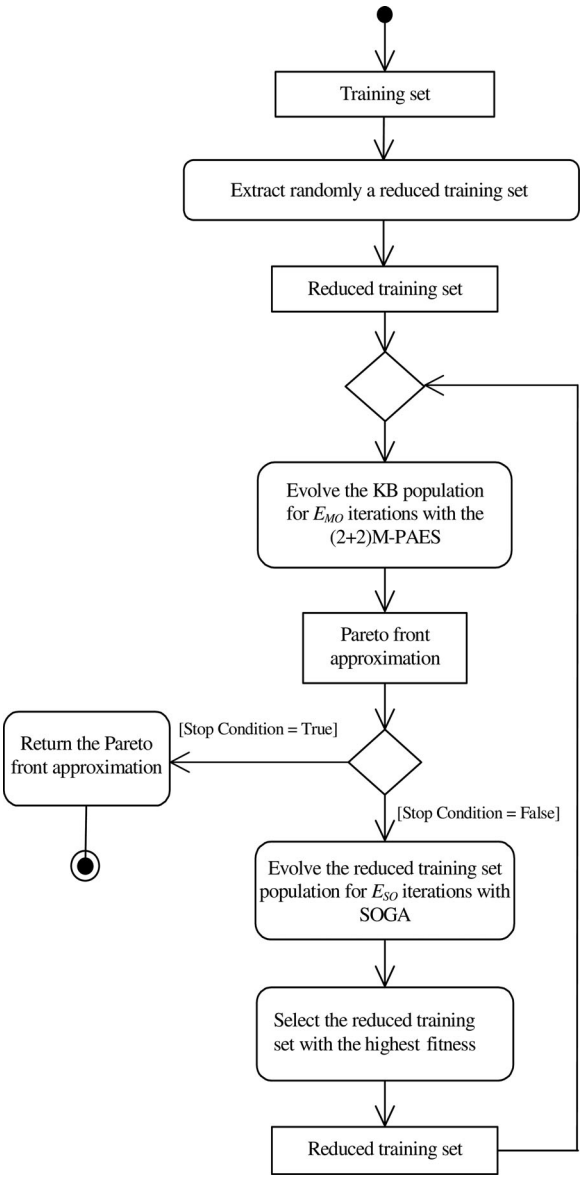

Fig. 1. Activity diagram of the coevolutionary approach.

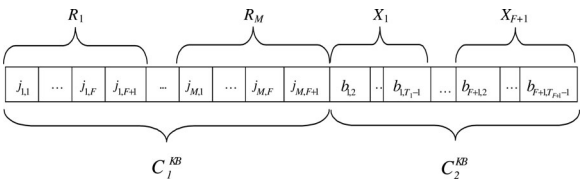

Fig. 2. Chromosome coding for the MOEL of KBs.

each variable X_f is partitioned by using T_f fuzzy sets. In Fig. 2, we show an example of chromosome coding for the MOEL of KBs.

As described in [6], $C_1^{\rm KB}$ is composed of $M \cdot (F+1)$ natural numbers, where M is the number of rules that are currently present in the RB. Each gene represents the index of the fuzzy set that has been selected for each variable of a specific rule.

With regard to $C_2^{\rm KB}$, since we adopt strong fuzzy partitions with, for $j = 2, \ldots, T_f - 1$, $b_{f,j} = c_{f,j-1}$, 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 fth variable (we normalize each variable in [0,1]). As $b_{f,1}$ 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. Thus, $C_2^{\rm KB}$ consists of $F+1$ vectors of $T_f - 2$ real numbers: The fth vector contains the $[b_{f,2}, \ldots, b_{f,T_f-1}]$ cores, which define the positions of the MFs for the linguistic variable X_f.

To ensure a good integrity level of the partitions, in terms of order, coverage, and distinguishability [44]–[47], $\forall j \in [2, T_f - 1]$, we force $b_{f,j}$ to vary in the definition interval $[b_{f,j} - \frac{b_{f,j} - b_{f,j-1}}{2}, b_{f,j} + \frac{b_{f,j+1} - b_{f,j}}{2}]$.

As MOEA, we adopt the (2+2)M-PAES that we have proposed in [6]. Unlike classical (2+2)PAES, which uses only mutation to generate new candidate solutions, (2+2)M-PAES exploits both crossover and mutation. Further, in (2+2)M-PAES, current solutions are randomly extracted at each iteration rather than maintained until they are not replaced by solutions with particular characteristics.

With regard to the crossover, we apply the one-point crossover to $C_1^{\rm KB}$ and the BLX-α crossover, with $\alpha = 0.5$, to $C_2^{\rm KB}$. Let s_1 and s_2 be two selected parent chromosomes. The common gene for $C_1^{\rm KB}$ is selected by extracting randomly a number in $[1, \rho_{\min} - 1]$, where $\rho_{\min}$ is the minimum 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 $C_1^{\rm KB}$, we can generate an MFRBS with one or more pairs of equal rules. In this case, we simply eliminate one of the rules from each pair. This allows us to reduce the total number of rules.

With regard to mutation, we apply two mutation operators for $C_1^{\rm KB}$. The first operator adds γ rules to the RB, where γ is randomly chosen in $[1, \gamma_{\max}]$. The upper bound $\gamma_{\max}$ is fixed by the user. If $\gamma + M > M_{\max}$, then $\gamma = M_{\max} - M$. $M_{\max}$ is fixed by the user and is used to constrain the search space. On the other hand, in real applications, we are not interested in generating MFRBSs with a large number of rules. For each rule R_m added to the chromosome, we generate a random number $v \in [1, F]$, which indicates the number of input variables that are used in the antecedent of the rule. Then, we generate v natural random numbers between 1 and F to determine the input variables which compose the antecedent part of the rule. Finally, for each selected input variable f, we generate a random natural number $j_{m,f}$ between 1 and T_f, which determines the fuzzy set $A_{f,j_{m,f}}$ to be used in the antecedent of rule R_m in the RB. To select the consequent fuzzy set $A_{F+1,j_{m,F+1}}$, a random number between 1 and T_{F+1} is generated. The second mutation operator randomly changes δ elements of the chromosome associated with the RB. The number δ is randomly generated in $[1, \delta_{\max}]$. The upper bound $\delta_{\max}$ is fixed by the user. For each element to be modified, a number is randomly generated in $[0, T_f]$. The mutation applied to $C_2^{\rm KB}$ first chooses randomly a variable $f \in [1, F+1]$, then extracts a random value $j \in [2, T_f - 1]$

and changes the value of $b_{f,j}$ to a random value in the interval $[b_{f,j} - \frac{b_{f,j}-b_{f,j-1}}{2}, b_{f,j} + \frac{b_{f,j+1}-b_{f,j}}{2}]$.

The results that are shown in our previous works [6], [9]–[12] have already pointed out the effectiveness of these crossover and mutation operators. 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 C_1^{KB}, mutation is applied with probability P_{m1}^{KB}, if the crossover has been executed; otherwise, mutation is always applied. Once application of mutation has been decided, the first and second mutation operators are applied with probability $P_{\mathrm{add}}^{\mathrm{KB}}$ and $1- P_{\mathrm{add}}^{\mathrm{KB}}$, respectively. Finally, the mutation operator is applied to C_2^{KB} with probability P_{m2}^{KB}.

(2+2)M-PAES determines an approximation of the optimal Pareto front by concurrently minimizing the MSE/2 and the complexity. The MSE/2 is calculated as follows:

$$\mathrm{MSE}/2 = \frac{1}{2 \cdot |S|} \sum_{l=1}^{|S|} \left(F(x^l) - y^l\right)^2 \tag{3}$$

where $|S|$ is the size of the dataset, $F(x^l)$ is the output obtained from the MFRBS when the lth input pattern is considered, and y^l is the desired output.

The 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 RBs characterized by a low number of rules and a low number of input variables really used in each rule.

B. Single-Objective Genetic Algorithm for Training Set Selection

When tackling the problem of selecting a reduced number of training instances by using evolutionary approaches, two main aspects have to be considered: the chromosome representation and the fitness function.

With regard to the chromosome representation, in [29], a binary string with N genes has been used, where N is the size of the TS. Each gene is associated with an instance: if the gene is equal to 1, then its associated instance is included in the reduced TS. Although this representation is very simple, it cannot adequately manage a large TS, since the size of the chromosome is equal to the size of the TS. To lessen this problem, in [35], authors have proposed to split the initial TS into disjoint strata and to perform the TS selection for each stratum. Similarly, we split the TS into B disjoint blocks: each block contains $I = N/B$ instances randomly extracted from the TS. Each reduced TS consists of K blocks. Thus, chromosome C^{TS} in the SOGA population codifies the K blocks that are included in the reduced TS: Each gene is an integer which varies from 1 to B and represents the index of the block selected to be included in the reduced TS.

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

$$DR_j = \frac{1}{N^{\mathrm{KB}}} \sum_{i=1}^{N^{\mathrm{KB}}} \frac{|\mathrm{MSE}_i/2 - \mathrm{MSE}_{i,j}/2|}{\max\left(\mathrm{MSE}_i/2, \mathrm{MSE}_{i,j}/2\right)}. \tag{4}$$

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

In order to perform the optimization of the TS selection 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 reduced TS. If this occurs, we change randomly the value of one of the equal genes 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% of individuals that 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%).

The values B and K are related to each other and have to be chosen accurately. If B is low (we assume that N is large in the following considerations), each block contains a high number of instances. Since the instances are inserted into the blocks randomly, it might be hard to find K blocks that are good representatives of the overall TS. On the other hand, if B is large, consequently, K is large as well. This implies that the search space of the SOGA is wide, and the SOGA could have difficulties to converge to an optimal solution. In our experiments, the values of B and K have been determined by fixing the number I of instances that are contained in a block and the desired percentage of instances of the overall TS contained in the reduced TS, respectively.

IV. Experimental Results

A. Experimental Setup

We tested our approach on 12 regression problem datasets: 11 datasets are extracted from Keel (available at http://sci2s.ugr.es/keel/datasets.php) and Torgo's (available at http://www.liaad.up.pt/∼ltorgo/Regression/DataSets.html) repositories, and one dataset is generated by sampling the nonlinear function that is discussed in [48]. As shown in Table I, the datasets are characterized by different numbers (from 6 to 21) of input

TABLE I
DATASETS USED IN THE EXPERIMENTS

Datasets	Number of Instances	Number of Input Variables	Repository
Analyzing Categorical Data (AN)	4052	7	Keel
Abalone (AB)	4177	9	Keel
Delta Ailerons (DA)	7129	6	Torgo
Computer activity (CA)	8192	21	Keel
Kinematics (KI)	8192	8	Torgo
Pumadyn (PM)	8192	8	Torgo
Delta Elevators (DE)	9517	6	Torgo
Non Linear Function (FO)	10000	6	-
Elevators (EL)	16559	18	Keel
California Housing (CH)	20460	8	Keel
House_16H (HO)	22784	16	Keel
MV Artificial Domain (MV)	40768	10	Keel

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
E_{SO}	Number of evaluations in the SOGA for each cycle	64
E_{MO}	Number of evaluations in the (2+2)M-PAES block for each cycle	640
T_f	Number of fuzzy sets in each variable X_f, f=1,…, F+1	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}	0.6
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}	0.55
P_{m2}^{KB}	Probability of applying the mutation operator to C_2^{KB}	0.2
$\gamma_{\max}$	Upper bound for γ in the first mutation operator applied to C_1^{KB}	5
$\delta_{\max}$	Upper bound for δ in the second mutation operator applied to C_1^{KB}	5
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

variables and different numbers (from 4052 to 40768) of input/output instances.

To assess the advantages of performing the multiobjective evolutionary KB learning of MFRBSs together with the TS selection, we compare the results that are achieved by our coevolutionary approach (denoted as PAES-SOGA in the following) with the ones obtained by applying the (2+2)M-PAES (denoted as PAES in the following) which uses the overall TS. In order to determine how many instances are necessary in the reduced TS to generate Pareto fronts comparable with the ones generated by using the overall TS, we performed experiments with the PAES-SOGA by using three different reduced TSs containing, respectively, 5%, 10%, and 20% of the overall TS. We denoted the PAES-SOGAs that employ the three different reduced TSs as PAES-SOGA(5%), PAES-SOGA(10%), and PAES-SOGA(20%), respectively.

Table II shows the values that are used in the experiments for the parameters of PAES and PAES-SOGA. These parameters have been obtained after a long experimentation. In particular, the values of $E_{\rm SO}$ and $E_{\rm MO}$, and the probabilities derive from an accurate analysis carried out to balance exploitation and exploration during the evolutionary process. As regards the values of T_f and $M_{\max}$, they might appear too low, especially when dealing with high-dimensional and highly nonlinear datasets. However, as discussed in our previous works [6], [9]–[12], when the search space is limited, the exploration of the evolutionary process is more accurate and can achieve solutions which, in general, are better than in the exploration of a very large search space. Further, we have to consider that during the evolutionary process, the parameter learning adapts the MFs to the specific TS, and the adaptation process is simple but quite powerful. Thus, to fix $T_f = 3$, $f = 1, \ldots, F+1$, and $M_{\max} = 30$ has a twofold benefit. On one side, the exploration results to be more accurate. On the other side, the resulting MFRBSs are characterized by a higher level of interpretability (lower complexity). Finally, we have to consider that our MOEL process manages "don't care" conditions and, therefore, can generate rules that are characterized by a number of conditions lower than the number of input variables. This means that a rule can cover a large amount of space. Since our learning process aims to reduce the complexity computed as total number of conditions used in the antecedent of rules, the final RB is usually composed of "general" rules with a low number of conditions (and therefore with a high number of "don't care" conditions). Thanks to this feature, we have verified that the RBs in the final Pareto fronts cover almost the totality of the patterns in each dataset.

As regards the PAES-SOGAs, we point out that the fitness evaluations of both the (2+2)M-PAES and the SOGA are computed by using the reduced TS. Only when we switch from the execution of the (2+2)M-PAES to the SOGA, we perform the evaluations of each individual in the Pareto front approximation by using the overall TS. Thus, we have that the PAES-SOGAs cyclically execute $E_{\rm MO} + E_{\rm SO}$ fitness evaluations with the reduced TS and $N^{\rm KB}$ (at most, AS) evaluations with the overall TS, where $N^{\rm KB}$ is the number of solutions contained in the (2+2)M-PAES archive when the switch is performed. Both PAES and PAES-SOGAs stop their execution when the total number of KB evaluations performed by the (2+2)M-PAES is equal to 300 000.

In the case of PAES-SOGAs, therefore, to determine the stopping criterion, we do not consider the number of evaluations performed by the SOGA using the reduced TS and the number of evaluations performed using the overall TS. As shown in Table II, the ratio between the number of evaluations of SOGA and the number of evaluations of the (2+2)M-PAES is 64/640, i.e., 1/10. This means that the number of evaluations performed by SOGA using the reduced TS is 30 000. Further, we recall that we adopt the overall TS only when we switch from (2+2)M-PAES to SOGA. Since we switch to SOGA after executing $E_{\rm MO} = 640$ evaluations of (2+2)M-PAES, we adopt the overall TS in the computation of only $\lceil \frac{300\,000}{640} \rceil \cdot N^{\rm KB} = 469 \cdot N^{\rm KB}$ evaluations. We have to also consider that $N^{\rm KB}$ varies during the evolutionary process: It is very small at the beginning and

grows with the number of iterations of the evolutionary process. However, we have observed that the final Pareto fronts for PAES-SOGA(5%), PAES-SOGA(10%), and PAES-SOGA(20%) contain on average 12, 16, and 22 solutions, respectively. This is the reason why we are able to save a lot of time with respect to PAES. Obviously, when we will discuss the execution times in Section IV-E, we will show the actual times needed by our approach to perform the 300 000 evaluations of the (2+2)M-PAES and the 30 000 evaluations of the SOGA using the reduced TS, and the $469 \cdot N^{\mathrm{KB}}$ evaluations using the overall TS.

For each dataset and for each algorithm, we carried out a fivefold cross validation and executed six trials for each fold with different seeds for the random function generator (30 trials in total). In the following sections, we first analyze the behavior of the three PAES-SOGAs and PAES in terms of unary indicators that are generally used in the literature to compare different MOEAs [49], [50]. The aim of this analysis is to show whether the PAES-SOGAs, though exploiting only a percentage of the overall TS, achieve Pareto front approximations comparable with the ones achieved by PAES. Second, the goodness of the solutions of the Pareto front approximations generated by PAES and PAES-SOGAs is assessed in terms of accuracy and complexity by using the procedure adopted in our previous papers [9], [12], as well as in [16]. Third, we show that the effectiveness of PAES-SOGA is just because of the synergy between (2+2)M-PAES and SOGA exploited by our coevolutionary approach. Finally, we present the execution times of PAES and PAES-SOGAs highlighting the percentage of time saved by using the reduced TSs.

B. Analysis of the Pareto Front Approximations

In order to analyze the effectiveness of the evolutionary process when the reduced TSs are used in place of the overall TS, we adopt two well-known unary quality indicators, namely the epsilon dominance and the hypervolume. These indicators represent a means to quantify the differences between Pareto front approximations by mapping each set of solutions to a real number. Independently of the TS (reduced TSs with 5%, 10%, and 20% of the overall TS and the overall TS itself) employed in the evolutionary process, to make the comparison sound, all the indicators are computed by using the MSE/2 calculated on the overall TS. Thus, for the PAES-SOGAs, we recompute the MSE/2 of the solutions contained in the final archive by using the overall TS in place of the reduced TS. The new values of MSE/2 might make some solution dominated by other solutions in the archive. Thus, before computing the indicators, we select the nondominated solutions to generate again a Pareto front approximation.

The epsilon dominance was proposed in [49] and makes direct use of the concept of Pareto dominance. Let us consider two Pareto front approximations A and B, where A dominates B. The epsilon dominance is a measure of the smallest distance one would need to translate every solution in B so that B dominates A. If C is chosen to be a reference Pareto front approximation, such that it dominates the A and B Pareto front approximations, then A and B can be directly compared with each other on the basis of the epsilon dominance with respect to C.

The hypervolume indicator, whose features and properties have been discussed in detail in [49], measures the hypervolume of the portion of the objective space that is weakly dominated by a Pareto front approximation. In order to compute this indicator, the objective space must be either bounded or a bounding reference point, that is (at least weakly) dominated by all the solutions, must be defined. The computations of the two indicators were performed using the performance assessment package provided in the PISA toolkit [51].

The Pareto front approximations that are generated by each approach, namely PAES-SOGA(5%), PAES-SOGA(10%), PAES-SOGA(20%), and PAES, are analyzed together. First, the maximum values of MSE/2 and complexity among the 120 approximations are computed in order to obtain the bounds for normalizing in [0,1] the objectives of each approximation. Then, the objectives are normalized. The hypervolume is calculated by using (1,1) as reference point. As a consequence of the normalization, the value of the two indicators is normalized in [0,1].

In order to assess if there exist statistical differences among the indicators and consequently among the Pareto front approximations generated by PAES and the PAES-SOGAs, we have performed a statistical analysis. As discussed in [52], we have applied nonparametric statistical tests for multiple comparisons by combining all the datasets: For each approach, we have generated a distribution consisting of the mean values of the epsilon dominance and of the hypervolume. We first have applied the Friedman test in order to compute a ranking among the distributions [53]. Then, we have applied the Iman and Davenport test in order to evaluate whether there exist statistically relevant differences among the mean values of the epsilon dominance and of the hypervolume computed for PAES and the PAES-SOGAs [54]. If there exist, then we apply a *post-hoc* procedure, namely the Holm test [55]. This test allows us to detect effective statistical differences between the control approach, i.e., the one with the lowest Friedman rank, and the remaining approaches.

Table III shows the results of the nonparametric statistical tests on both indicators: For each indicator and for each approach, we show the Iman and Davenport *p-value* and the Friedman rank. If the *p-value* is lower than the *level of significance* α (in the experiments, $\alpha = 0.05$), we can reject the null hypothesis and affirm that there exist statistical differences between the multiple distributions, associated with each approach, of the epsilon dominance and/or of the hypervolume indicators. Otherwise, no statistical difference exists among the distributions, and therefore, the four different algorithms evolve toward similar Pareto front approximations.

For both the indicators, the Iman and Davenport statistical hypothesis of equivalence is rejected. Thus, we have to apply the Holm *post-hoc* procedure considering the PAES as control algorithm (associated with the lowest rank and in bold in the Table). We observe that the statistical hypothesis of equivalence cannot be rejected for PAES-SOGA(10%) and PAES-SOGA(20%). On the other hand, the statistical hypothesis of equivalence is rejected for PAES-SOGA(5%).

The results of the statistical tests point out that the 5% of the overall TS is not sufficient to allow the MOEA to evolve toward Pareto front approximations comparable with the ones

TABLE III
RESULTS OF THE NONPARAMETRIC STATISTICAL TESTS ON EPSILON DOMINANCE AND HYPERVOLUME

Epsilon Dominance

	Algorithm	*Friedman Rank*	*Iman and Davenport p-value*		*Hypothesis*
	PAES-SOGA(5%)	3.666			
	PAES-SOGA(10%)	2.666	9.39E-04		Rejected
	PAES-SOGA(20%)	1.833			
	PAES	1.833			
	Holm Post-hoc Procedure				
i	*Algorithm*	*z-value*	*p-value*	*alpha/i*	*Hypothesis*
3	PAES-SOGA(5%)	3.479	0.0005	0.017	Rejected
2	PAES-SOGA(10%)	1.581	0.1138	0.025	Not Rejected
1	PAES-SOGA(20%)	0.000	1.0000	0.05	Not Rejected

Hypervolume

	Algorithm	*Friedman Rank*	*Iman and Davenport p-value*		*Hypothesis*
	PAES-SOGA(5%)	3.500			
	PAES-SOGA(10%)	2.667	6.42E-03		Rejected
	PAES-SOGA(20%)	1.917			
	PAES	1.917			
	Holm Post-hoc Procedure				
i	*Algorithm*	*z-value*	*p-value*	*alpha/i*	*Hypothesis*
3	PAES-SOGA(5%)	3.004	0.0026	0.017	Rejected
2	PAES-SOGA(10%)	1.423	0.1547	0.025	Not Rejected
1	PAES-SOGA(20%)	0.000	1.000	0.05	Not Rejected

obtained by using the other percentages and the overall TS itself. On the other hand, the tests highlight that, when adopting 10% and 20% of the overall TS and the overall TS itself, no statistical difference exists in terms of epsilon dominance and hypervolume of the obtained Pareto front approximations.

In conclusion, we can assert that the evolutionary process, in terms of both exploration and exploitation, is not affected by using the reduced TSs rather than the overall TS. From Table III, we can conclude that just 10% of the overall TS allows the MOEA to evolve toward Pareto front approximations which are statistically equivalent to the ones obtained by using the overall TS. As we will point out in Section IV-E, obviously, the use of PAES-SOGA(10%) in place of PAES allows us to save 86.36% of the execution time.

C. Analysis of the Solutions Generated by PAES-SOGAs and PAES in Terms of Accuracy and Complexity

In order to analyze the MFRBSs generated using PAES-SOGAs and PAES in terms of accuracy and complexity, we have applied the procedure proposed in some of our previous papers [9], [12] and also used in [16]. For the PAES-SOGAs, we recompute the MSE/2 of the solutions contained in the final archive by using the overall TS in place of the reduced TS. The new values of MSE/2 might make some solution dominated by other solutions in the archive. Thus, before applying the procedure, we select the nondominated solutions to generate again a Pareto front approximation. The procedure is based on the analysis of three representative solutions of the Pareto front approximations, namely the most accurate (denoted as FIRST), the least accurate (denoted as LAST), and the median (denoted as MEDIAN) between the FIRST and the LAST solutions. In practice, for each of the 30 trials, we compute the Pareto front approximations of each algorithm and order the solutions in each approximation to increase MSE/2 values. Then, for each approximation, we select the first (the most accurate), the median, and the last (the least accurate) solutions (this is the reason why we denote the solutions as FIRST, MEDIAN, and LAST). Finally, for the FIRST, MEDIAN, and LAST solutions, we compute the mean values over the 30 trials of the MSE/2 on the training and test sets, and of the complexity. To provide a glimpse of the three representative solutions, in Figs. 3 and 4, we plot the mean values of the MSE/2 and the complexity for the FIRST, MEDIAN, and LAST solutions for all the 12 datasets on both the training and test sets.

First of all, we observe in Figs. 3 and 4 that for each dataset, both the PAES-SOGAs and PAES do not suffer from the overfitting problem. Indeed, the mean values of the MSE/2 computed on the test set are approximately equal to the mean values of the MSE/2 computed on the TS. Further, from a careful analysis of the plots, we can deduce conclusions similar to the ones achieved by using the epsilon dominance and hypervolume indicators. Indeed, as regards the values of MSE/2, the corresponding FIRST, MEDIAN, and LAST points for the four MOEAs are very close to each other on both the training and the test sets (except sometimes for PAES-SOGA(5%) in the FIRST and MEDIAN solutions). On the other hand, we can observe some difference in terms of complexity. Indeed, the solutions that are generated by PAES are characterized by higher values of complexity than corresponding solutions generated by the other three approaches. Since PAES uses the overall TS and, therefore, a larger number of instances than the other approaches, the evolutionary process tends to generate a larger number of rules with a larger number of conditions to cover all the instances and consequently to obtain a low MSE/2. Further, we have to consider that, after recomputing the MSE/2 of the solutions contained in the final archive by using the overall TS in place of the reduced TS, we remove all the solutions that are not nondominated anymore. Often, these solutions lie in the high-complexity region.

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-SOGAs and PAES, we have applied nonparametric statistical tests for multiple comparisons by combining all the datasets: For each approach and for each of the three representative solutions, we have generated a distribution consisting of the mean values of the MSE/2. As stated in Section IV-B, we first have applied the Friedman test in order to compute a ranking among the distributions. Then, we have applied the Iman and Davenport test in order to evaluate whether there exist statistically relevant differences among the mean values of the MSE/2 computed for the four algorithms. If there exist, then we apply a *post-hoc* procedure, namely the Holm test. For the sake of brevity, Table IV shows only the results that are obtained on the test set. On the other hand, just these results are relevant to evaluate the generalization capabilities of the MFRBSs that are generated by the different approaches.

For the FIRST and MEDIAN solutions, since the Iman and Davenport test rejects the statistical hypothesis of equivalence, the Holm *post-hoc* procedure is applied by considering

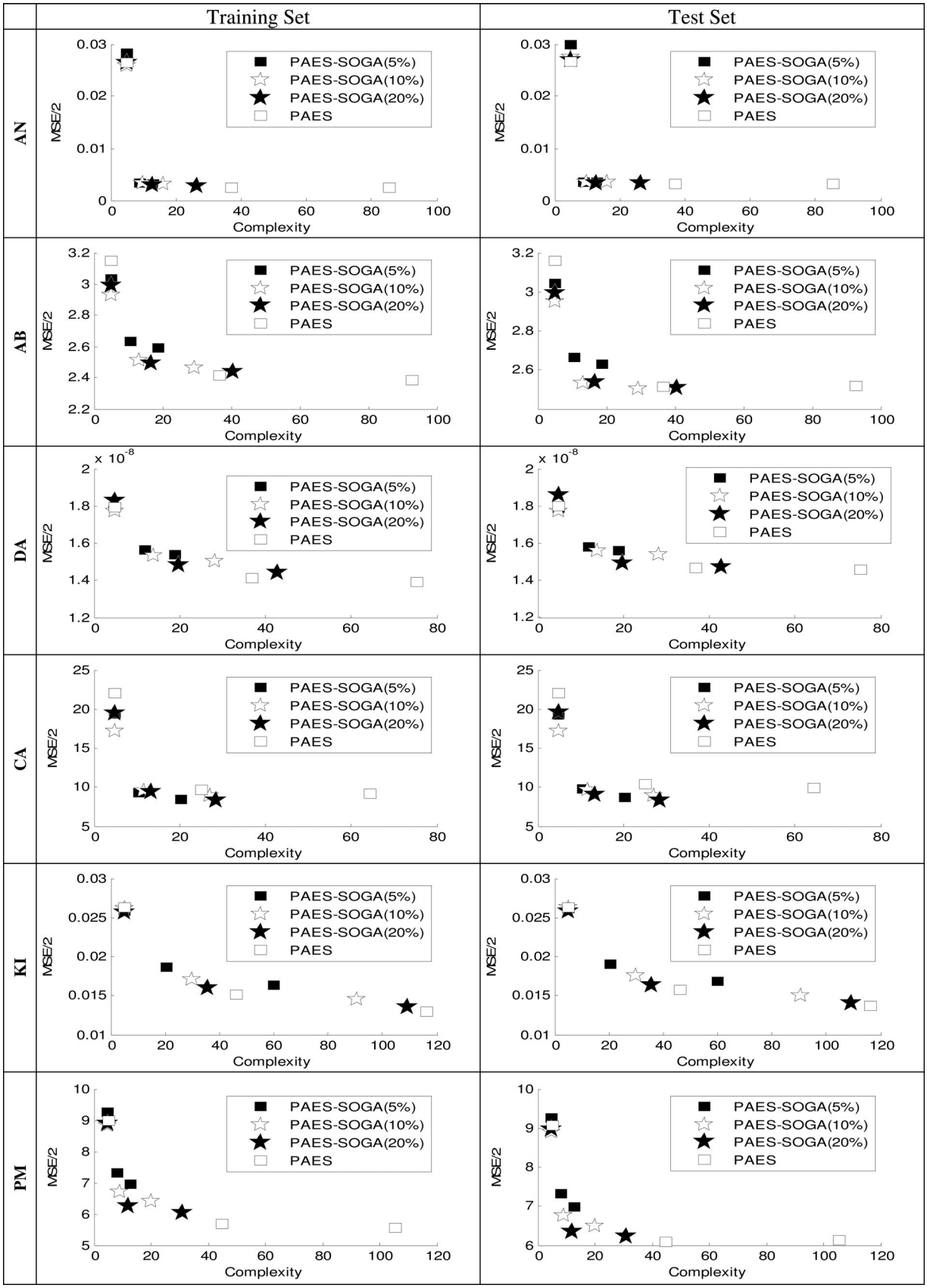

Fig. 3. Plots of the FIRST, MEDIAN, and LAST solutions onto the complexity-MSE/2 plane (first six datasets).

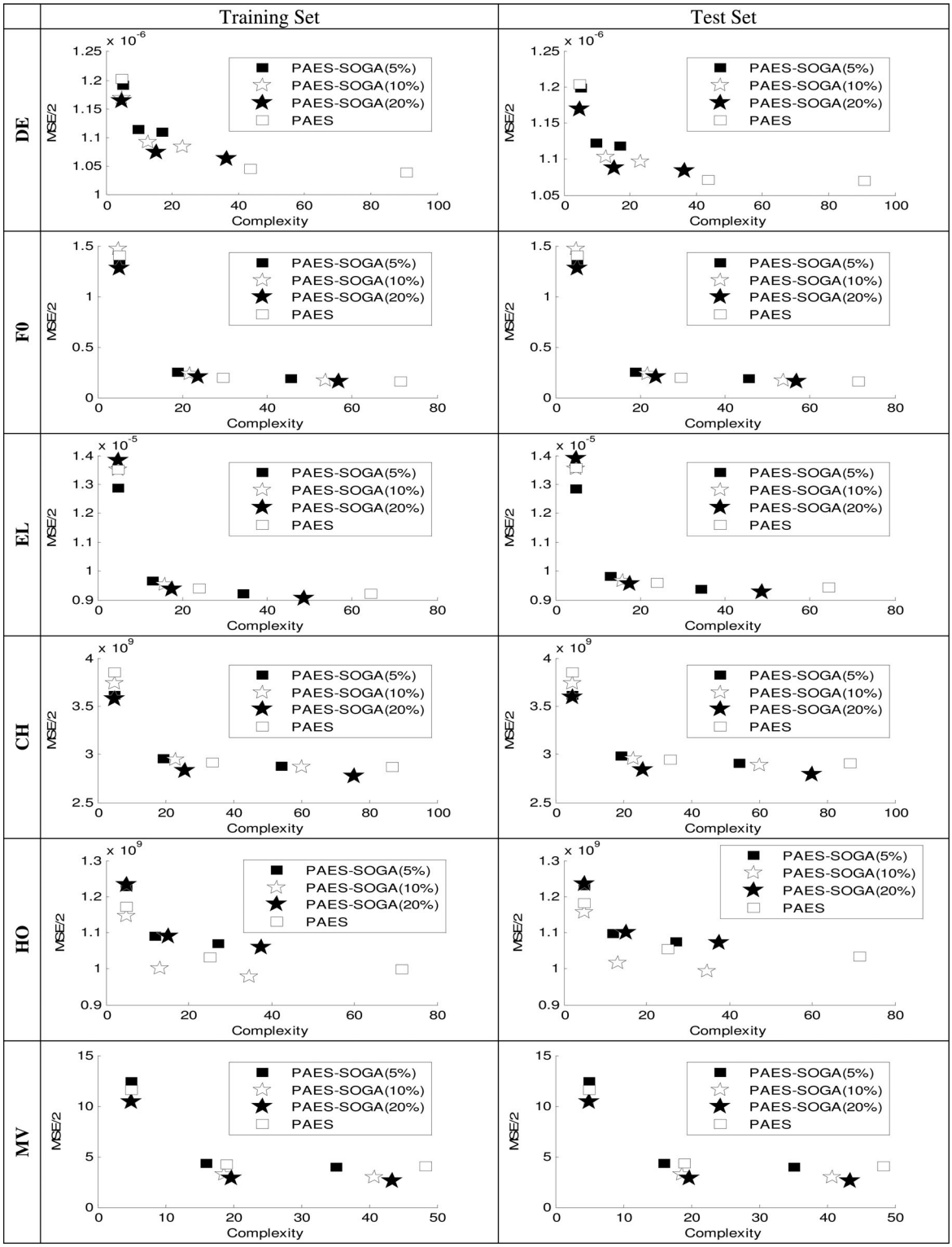

Fig. 4. Plots of the FIRST, MEDIAN, and LAST solutions onto the complexity-MSE/2 plane (last six datasets).

PAES-SOGA(20%) and PAES as control algorithms for the FIRST and MEDIAN solutions, respectively. For these solutions, the statistical hypothesis of equivalence is rejected only for PAES-SOGA(5%). Thus, we can conclude that PAES-SOGA(10%), PAES-SOGA(20%), and PAES result to be statistically equivalent in terms of MSE/2. With regard to the LAST solution, the Iman and Davenport test does not reject the statistical hypothesis of equivalence.

From the statistical comparison, since PAES-SOGA(10%) is always statistically equivalent to PAES and exploits only the 10% of the TS, it can be considered as the best approach, at least for the analyzed datasets.

TABLE IV
RESULTS OF THE NONPARAMETRIC STATISTICAL TESTS ON THE FIRST, MEDIAN, AND LAST SOLUTIONS OF THE TEST SETS, WHEN COMPARING PAES-SOGAS AND PAES

FIRST

Algorithm	*Friedman Rank*	*Iman and Davenport p-value*	*Hypothesis*
PAES-SOGA(5%)	3.666	5.40E-04	Rejected
PAES-SOGA(10%)	2.4166		
PAES-SOGA(20%)	1.750		
PAES	2.166		

Holm Post-hoc Procedure

i	*Algorithm*	*z-value*	*p-value*	*alpha/i*	*Hypothesis*
3	PAES-SOGA(5%)	3.636	2.76E-04	0.017	Rejected
2	PAES-SOGA(10%)	1.264	0.2059	0.025	Not Rejected
1	PAES	0.790	0.4291	0.05	Not Rejected

MEDIAN

Algorithm	*Friedman Rank*	*Iman and Davenport p-value*	*Hypothesis*
PAES-SOGA(5%)	3.833	4.07E-06	Rejected
PAES-SOGA(10%)	2.583		
PAES-SOGA(20%)	1.9166		
PAES	1.6666		

Holm Post-hoc Procedure

i	*Algorithm*	*z-value*	*p-value*	*alpha/i*	*Hypothesis*
3	PAES-SOGA(5%)	4.110	3.94E-05	0.017	Rejected
2	PAES-SOGA(10%)	1.739	0.0810	0.025	Not Rejected
1	PAES-SOGA(20%)	0.474	0.6352	0.05	Not Rejected

LAST

Algorithm	*Friedman Rank*	*Iman and Davenport p-value*	*Hypothesis*
PAES-SOGA(5%)	2.666	1.96E-01	Not Rejected
PAES-SOGA(10%)	2.0833		
PAES-SOGA(20%)	2.1666		
PAES	3.0833		

TABLE V
RESULTS OF THE NONPARAMETRIC STATISTICAL TESTS ON THE FIRST SOLUTIONS WHEN COMPARING PAES-SOGAS WITH DIFFERENT APPROACHES

Algorithm	*Friedman Rank*	*Iman and Davenport p-value*	*Hypothesis*
PAES-RFS(5%)	3.667	1.56E-04	Rejected
PAES-RSE(5%)	2.250		
PAES-STS(5%)	2.500		
PAES-SOGA(5%)	1.583		

Holm Post-hoc Procedure

i	*Algorithm*	*z-value*	*p-value*	*alpha/i*	*Hypothesis*
3	PAES-RFS(5%)	3.952	7.72E-05	0.017	Rejected
2	PAES-STS(5%)	1.739	0.0819	0.025	Not Rejected
1	PAES-RSE(5%)	1.264	0.2059	0.05	Not Rejected

Algorithm	*Friedman Rank*	*Iman and Davenport p-value*	*Hypothesis*
PAES-RFS(10%)	3.166	7.26E-04	Rejected
PAES-RSE(10%)	3.083		
PAES-STS(10%)	2.333		
PAES-SOGA(10%)	1.416		

Holm Post-hoc Procedure

i	*Algorithm*	*z-value*	*p-value*	*alpha/i*	*Hypothesis*
3	PAES-RFS(10%)	3.320	8.99E-04	0.017	Rejected
2	PAES-RSE(10%)	3.162	0.0015	0.025	Rejected
1	PAES-STS(10%)	1.733	0.0819	0.05	Not Rejected

Algorithm	*Friedman Rank*	*Iman and Davenport p-value*	*Hypothesis*
PAES-RFS(20%)	3.166	7.26E-04	Rejected
PAES-RSE(20%)	2.916		
PAES-STS(20%)	2.583		
PAES-SOGA(20%)	1.333		

Holm Post-hoc Procedure

i	*Algorithm*	*z-value*	*p-value*	*alpha/i*	
3	PAES-RFS(20%)	3.478	5.04E-04	0.017	Rejected
2	PAES-RSE(20%)	3.004	0.0026	0.025	Rejected
1	PAES-STS(20%)	2.371	0.0177	0.05	Not Rejected

D. Analysis on the Effectiveness and Usefulness of the Coevolutionary Approach

In the previous sections, we have discussed how PAES-SOGA(10%) and PAES-SOGA(20%) are statistically equivalent to PAES in terms of the generated Pareto front approximations and accuracy on the test set. Here, we would like to show that this equivalence is just because of the synergy between (2+2)M-PAES and SOGA exploited by our coevolutionary approach.

To this aim, we have performed three different experiments. For each experiment, we have carried out the same fivefold cross validation discussed in Section IV-A. In the first experiment, we desired to evaluate whether a randomly chosen reduced TS could allow (2+2)M-PAES to achieve the same performance as the corresponding PAES-SOGAs. Thus, for each trial, we have randomly chosen three reduced TSs consisting of 5%, 10%, and 20% of the overall TS, respectively. Then, we have executed the (2+2)M-PAES on the three reduced TSs for 300 000 evaluations. In the second experiment, we intended to asses whether the use of the SOGA in periodically selecting the most representative reduced TS was really effective. To this aim, we have executed the (2+2)M-PAES for 300 000 evaluations by periodically updating the reduced TS with the same frequency used in the execution of the SOGA. Unlike PAES-SOGAs that replace the current reduced TS with the reduced TS identified by the SOGA, here, the reduced TS is randomly generated from the overall TS. In the third experiment, we aimed to evaluate whether the reduced TS selected in the last execution of the SOGA was actually a good representative of the overall TS. Thus, for each analyzed percentage (5%, 10%, and 20%) of the TS, we have used the final reduced TS selected for each trial by the PAES-SOGAs as TS to execute 300 000 evaluations of (2+2)M-PAES. We have denoted the approaches that are used in the three experiments as PAES-RFS, PAES-RES, and PAES-STS, respectively, where RFS, RSE, and STS stand for random fixed selection, random selection during evolution, and selected TS.

Tables V–VII show the results of the same nonparametric statistical tests that are applied in the previous sections for the FIRST, MEDIAN, and LAST solutions, respectively. We observe that the Iman and Davenport test always rejects the null hypothesis.

For the FIRST and MEDIAN solutions, we observe that, as expected, PAES-SOGA is the control algorithm for each reduction percentage. In particular, for 10% and 20% of the TS, the Holm *post-hoc* procedure always rejects the statistical hypothesis of equivalence for PAES-RFS and PAES-RES, i.e., PAES-SOGA statistically outperforms both the approaches. On the other hand, PAES-STS result to be statistically equivalent to PAES-SOGA. This proves that the synergy between

TABLE VI
RESULTS OF THE NONPARAMETRIC STATISTICAL TESTS ON THE MEDIAN SOLUTIONS WHEN COMPARING PAES-SOGAS WITH DIFFERENT APPROACHES

Algorithm	*Friedman Rank*	*Iman and Davenport p-value*	*Hypothesis*
PAES-RFS(5%)	3.583	2.55E-03	Rejected
PAES-RSE(5%)	2.416		
PAES-STS(5%)	2.166		
PAES-SOGA(5%)	1.833		

Holm Post-hoc Procedure

i	*Algorithm*	*z-value*	*p-value*	*alpha/i*	*Hypothesis*
3	PAES-RFS(5%)	3.320	8.99E-04	0.017	Rejected
2	PAES-STS(5%)	1.106	0.2682	0.025	Not Rejected
1	PAES-RSE(5%)	0.632	0.5270	0.05	Not Rejected

Algorithm	*Friedman Rank*	*Iman and Davenport p-value*	*Hypothesis*
PAES-RFS(10%)	3.083	2.55E-03	Rejected
PAES-RSE(10%)	3.166		
PAES-STS(10%)	2.166		
PAES-SOGA(10%)	1.583		

Holm Post-hoc Procedure

i	*Algorithm*	*z-value*	*p-value*	*alpha/i*	*Hypothesis*
3	PAES-RSE(10%)	3.004	0.0026	0.017	Rejected
2	PAES-RFS(10%)	2.846	0.0044	0.025	Rejected
1	PAES-STS(10%)	1.106	0.2683	0.05	Not Rejected

Algorithm	*Friedman Rank*	*Iman and Davenport p-value*	*Hypothesis*
PAES-RFS(20%)	2.958	5.50E-04	Rejected
PAES-RSE(20%)	3.416		
PAES-STS(20%)	1.958		
PAES-SOGA(20%)	1.666		

Holm Post-hoc Procedure

i	*Algorithm*	*z-value*	*p-value*	*alpha/i*	
3	PAES-RSE(20%)	3.320	8.99E-04	0.017	Rejected
2	PAES-RFS(20%)	2.450	0.0142	0.025	Rejected
1	PAES-STS(20%)	0.553	0.5799	0.05	Not Rejected

TABLE VII
RESULTS OF THE NONPARAMETRIC STATISTICAL TESTS ON THE LAST SOLUTIONS WHEN COMPARING PAES-SOGAS WITH DIFFERENT APPROACHES

Algorithm	*Friedman Rank*	*Iman and Davenport p-value*	*Hypothesis*
PAES-RFS(5%)	3.916	8.04E-09	Rejected
PAES-RSE(5%)	1.416		
PAES-STS(5%)	2.083		
PAES-SOGA(5%)	2.5833		

Holm Post-hoc Procedure

i	*Algorithm*	*z-value*	*p-value*	*alpha/i*	*Hypothesis*
3	PAES-RFS(5%)	4.743	2.10E-06	0.017	Rejected
2	PAES-SOGA(5%)	2.213	0.0268	0.025	Not Rejected
1	PAES-STS(5%)	1.264	0.2059	0.05	Not Rejected

Algorithm	*Friedman Rank*	*Iman and Davenport p-value*	*Hypothesis*
PAES-RFS(10%)	3.375	3.16E-03	Rejected
PAES-RSE(10%)	1.541		
PAES-STS(10%)	2.583		
PAES-SOGA(10%)	2.500		

Holm Post-hoc Procedure

i	*Algorithm*	*z-value*	*p-value*	*alpha/i*	*Hypothesis*
3	PAES-RFS(10%)	3.478	5.04E-04	0.017	Rejected
2	PAES-STS(10%)	1.976	0.0481	0.025	Not Rejected
1	PAES-SOGA(10%)	1.818	0.0690	0.05	Not Rejected

Algorithm	*Friedman Rank*	*Iman and Davenport p-value*	*Hypothesis*
PAES-RFS(20%)	3.333	2.50E-02	Rejected
PAES-RSE(20%)	1.791		
PAES-STS(20%)	2.375		
PAES-SOGA(20%)	2.500		

Holm Post-hoc Procedure

i	*Algorithm*	*z-value*	*p-value*	*alpha/i*	
3	PAES-RFS(20%)	2.925	0.0034	0.017	Rejected
2	PAES-SOGA(20%)	1.343	0.1789	0.025	Not Rejected
1	PAES-STS(20%)	1.106	0.2683	0.05	Not Rejected

(2+2)M-PAES and SOGA is successful in selecting representative reduced TSs which allow evolving toward effective Pareto front approximations. For 5% of the TS, although PAES-SOGA results to be the algorithm with the lowest Friedman rank, the Holm *post-hoc* procedure rejects the statistical hypothesis of equivalence only for PAES-RFS. On the other hand, in Table IV, we have already shown that PAES-SOGA(5%) cannot achieve results that are statistically comparable with PAES, PAES-SOGA(10%), and PAES-SOGA(20%).

Probably 5% of samples is a percentage that is too small to be representative of the overall TS. Thus, the SOGA is not able to find the reduced TS which allows us to achieve performance comparable with the overall TS. In the case of PAES-RFS, the nonrepresentativeness of the reduced TS leads the (2+2)M-PAES to overfit. In the case of PAES-SOGA, this nonrepresentativeness leads PAES-SOGA to have a behavior statistically similar to PAES-RES. Anyway, the choice of the reduced TS performed by PAES-SOGA is effective, since PAES-STS does not suffer from overfitting like PAES-RFS.

For the LAST solution, the Holm *post-hoc* procedure rejects the statistical hypothesis of equivalence only for PAES-RFS, thus confirming that the random choice of the reduced TS at the beginning of the evolutionary process is not a successful approach. On the other hand, PAES-RES, PAES-STS, and PAES-SOGA result to be statistically equivalent. This result can be explained by considering that the LAST solutions are characterized by very high MSE/2 and very low complexity. Thus, an accurate selection of the reduced TS cannot considerably improve the accuracy. On the other hand, periodically changing the reduced TS, also randomly as in PAES-RES, can avoid overfitting problems.

From the three experiments, we can deduce two main conclusions. First, the periodical-guided update of the reduced TS permits us to achieve a good level of generalization, without experiencing overfitting problems. Second, the reduced TS that is selected by our approach is a good choice. Indeed, when in PAES-STS we use the reduced TS in the (2+2)M-PAES, the generated solutions do not suffer from overfitting problems. Further, they are statistically equivalent to the ones generated by PAES-SOGA.

E. Training Set Reduction and Computational Times

In this section, we analyze how the use of the reduced TS allows us to considerably reduce the execution times. Table VIII shows, for each dataset and for each algorithm, the average execution time in seconds and in percentage (between parentheses) with respect to the time needed by PAES to execute the same number (i.e., 300 000) of evaluations. The algorithms have been executed on a PC equipped with an Intel core i5 750 at 2.67 GHz, 4-GB RAM, and an Ubuntu operating system. We can observe that the PAES-SOGAs allow us to save a considerable

TABLE VIII
AVERAGE VALUES EXPRESSED IN SECONDS AND IN PERCENTAGES WITH RESPECT TO THE TIME NEEDED BY PAES FOR EXECUTING THE SAME NUMBER (300 000) OF EVALUATIONS

Dataset	PAES-SOGA(5%)	PAES-SOGA(10%)	PAES-SOGA(20%)	PAES
AN	121.7 (6.45%)	238 (12.62%)	458.3 (24.29%)	1886.4 (100%)
AB	173.7 (7.52%)	333.4 (14.43%)	632.7 (27.38%)	2310.5 (100%)
DA	245.0 (6.34%)	437.7 (11.33%)	1007.9 (26.08%)	3864.5 (100%)
CA	664.3 (5.83%)	1196.6 (10.50%)	2249.6 (19.74%)	11396.4 (100%)
KI	512.0 (9.64%)	997.4 (18.79%)	1957.3 (36.86%)	5309.4 (100%)
PM	516.0 (9.72%)	1001.3 (18.85%)	1968.7 (37.07%)	5311.3 (100%)
DE	497.6 (9.55%)	865.1 (16.60%)	1498.5 (28.75%)	5212.1 (100%)
FO	510.3 (9.56%)	876.4 (16.42%)	1517.8 (28.44%)	5337.6 (100%)
EL	1262.7 (7.39%)	2239.2 (13.11%)	4160.9 (24.36%)	17083 (100%)
CH	1055.3 (9.99%)	1971.1 (18.66%)	3472.8 (32.88%)	10560.5 (100%)
HO	1442.4 (7.19%)	2468.2 (12.30%)	4703 (23.43%)	20070.2 (100%)
MV	1707.0 (9.00%)	3065.8 (16.17%)	5576.8 (29.41%)	18961.8 (100%)
Mean	676.6 (7.44%)	1219.2 (13.64%)	2286.0 (26.12%)	8747.6 (100%)

amount of time. In particular, the use of reduced TSs containing 5%, 10%, and 20% of the overall TS allow us to save on average 92.56%, 86.36%, and 73.88% of the execution time needed by using the overall TS, respectively. We note that the percentages (7.44%, 13.64%, and 26.12% for the reduced TSs composed, respectively, of 5%, 10%, and 20% of the overall TS) of execution time are higher than the percentages of instances used in the reduced TSs. This is because of the time spent to switch from the execution of the (2+2)M-PAES to the execution of the SOGA and *vice versa* and to execute the SOGA.

V. CONCLUSION

In this paper, we have proposed a novel approach to perform an evolutionary TS selection in the framework of an MOEL of MFRBSs. In particular, the IS and the MOEL are carried out, respectively, by an SOGA and by the (2+2)M-PAES, which is a modified version of the well-known multiobjective evolutionary algorithm PAES.

Cyclically, the single objective genetic algorithm evolves a population of reduced TSs with the objective of maximizing a purposely defined index, which measures how much a reduced TS is representative of the overall TS in the context of the MOEL. The MOEL is used to generate a set of MFRBSs that are characterized by different tradeoffs between system accuracy and RB complexity. During the execution of the (2+2)M-PAES, the RB and the parameters of the MFs are learnt concurrently, and the fitness of each individual is computed by using the selected reduced TS, thus saving considerable execution time. The approach is, therefore, particularly suitable to deal with large datasets consisting of a high number of instances.

We have discussed the results of the application of our approach to 12 regression problems with a large number of instances. We have used three different reduced TSs composed of, respectively, 5%, 10%, and 20% of the overall TS. By using nonparametric statistical tests for multiple comparisons on all the datasets together, we have shown that there do not exist statistical differences in terms of epsilon dominance, hypervolume, and accuracy on the test set between the Pareto front approximations that are generated, respectively, by the coevolutionary approach with 10% and 20% of the overall TS and by the MOEL with the overall TS, although the use of the reduced TSs allows us to save up to 86.36% of the execution time.

REFERENCES

[1] F. Herrera, "Genetic fuzzy systems: Taxonomy, current research trends and prospects," *Evol. Intell.*, vol. 1, pp. 27–46, 2008.
[2] H. Ishibuchi and Y. Nojima, "Multiobjective genetic fuzzy systems," in *Computational Intelligence*, vol. 1, C. L. Mumford and L. C. Jain, Eds. Berlin, Germany: Springer-Verlag, 2009, pp. 131–173.
[3] H. Ishibuchi, T. Murata, and I. B. Turksen, "Single-objective and two objective genetic algorithms for selecting linguistic rules for pattern classification problems," *Fuzzy Sets Syst.*, vol. 89, no. 2, pp. 135–150, 1997.
[4] H. Ishibuchi and T. Yamamoto, "Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining," *Fuzzy Sets Syst.*, vol. 141, pp. 59–88, 2004.
[5] H. Ishibuchi, T. Nakashima, and T. Murata, "Three-objective genetics-based machine learning for linguistic rule extraction," *Inf. Sci.*, vol. 136, no. 1–4, pp. 109–133, 2001.
[6] M. Cococcioni, P. Ducange, B. Lazzerini, and F. Marcelloni, "A Pareto-based multi-objective evolutionary approach to the identification of Mamdani fuzzy systems," *Soft Comput.*, vol. 11, no. 11, pp. 1013–1031, 2007.
[7] P. Ducange, B. Lazzerini, and F. Marcelloni, "Multi-objective genetic fuzzy classifiers for imbalanced and cost-sensitive datasets," *Soft Comput.*, vol. 14, no. 10, pp. 713–728, 2010.
[8] 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 Comput.*, vol. 13, no. 5, pp. 437–449, 2009.
[9] 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, no. 5, pp. 1106–1122, Oct. 2009.
[10] M. Antonelli, P. Ducange, B. Lazzerini, and F. Marcelloni, "Multi-objective evolutionary learning of granularity, membership function parameters and rules of Mamdani fuzzy systems," *Evol. Intell.*, vol. 2, no. 1–2, pp. 21–37, 2009.

[11] 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, no. 7, pp. 1066–1080, 2009.

[12] M. Antonelli, P. Ducange, B. Lazzerini, and F. Marcelloni, "Learning knowledge bases of multi-objective evolutionary fuzzy systems by simultaneously optimizing accuracy, complexity and partition integrity," *Soft Comput.*, vol. 15, no. 12, pp. 2335–2354, 2011.

[13] P. Pulkkinen and H. Koivisto, "A dynamically constrained multiobjective genetic fuzzy system for regression problems," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 1, pp. 161–177, Feb. 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. Uncertainty Fuzziness Knowl.-Based Syst.*, vol. 15, no. 5, pp. 539–557, 2007.

[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 Comput.*, vol. 13, no. 5, pp. 419–436, 2009.

[16] M. J. Gacto, R. Alcala, 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, no. 3, pp. 515–531, Jun. 2010.

[17] Y. Nojima, H. Ishibuchi, and I. Kuwajima, "Parallel distributed genetic fuzzy rule selection," *Soft Comput.*, vol. 13, pp. 511–519, 2008.

[18] Y. Nojima, H. Ishibuchi, and S. Mihara, "Use of very small training data subsets in parallel distributed genetic fuzzy rule selection," presented at the 4th Int. Workshop Genet. Evol. Fuzzy Syst., Mieres, Spain, Mar. 2010.

[19] I. Robles, R. Alcalá, J. M. Benítez, and F. Herrera, "Evolutionary parallel and gradually distributed lateral tuning of fuzzy rule-based systems," *Evol. Intell.*, vol. 2, no. 1–2, pp. 5–19, 2009.

[20] 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. 9th Int. Conf. Intell. Syst. Design Appl.*, Pisa, Italy, Nov. 30–Dec. 2, 2009, pp. 134–139.

[21] Y. Jin, "A comprehensive survey of fitness approximation in evolutionary computation," *Soft Comput.*, vol. 9, pp. 3–12, 2005.

[22] M. Cococcioni, B. Lazzerini, and F. Marcelloni, "On reducing computational overhead in multi-objective genetic Takagi–Sugeno fuzzy systems," *Appl. Soft Comput.*, vol. 11, no. 1, pp. 675–688, 2011.

[23] M. Cococcioni, B. Lazzerini, and F. Marcelloni, "Towards efficient multiobjective genetic Takagi-Sugeno fuzzy systems for high dimensional problems," in *Computational Intelligence in Expensive Optimization Problems* (Springer Series Studies in Evolutionary Learning and Optimization). New York: Springer-Verlag, 2010, pp. 397–422.

[24] Y. Nojima and H. Ishibuchi, "Effects of data reduction on the generalization ability of parallel distributed genetic fuzzy rule selection," in *Proc. 9th Int. Conf. Intell. Syst. Design Appl.*, Pisa, Italy, Nov. 30–Dec. 2, 2009, pp. 96–101.

[25] H. Liu and H. Motoda, "On issues of instance selection," *Data Mining Knowl. Discov.*, vol. 6, pp. 115–130, 2002.

[26] N. Jankowski and M. Grochowski, "Comparison of instances selection algorithms I. Algorithms survey," in *Artificial Intelligence and Soft Computing*, vol. LNCS-3070. Berlin, Germany: Springer, 2004, pp. 598–603.

[27] J. Derrac, S. García, and F. Herrera, "A survey on evolutionary instance selection and generation," *Int. J. Appl. Metaheuristic Comput.*, vol. 1, no. 1, pp. 60–92, 2010.

[28] L. I. Kuncheva, "Editing for the k-nearest neighbors rule by a genetic algorithm," *Pattern Recognit. Lett.*, vol. 16, no. 8, pp. 809–814, 1995.

[29] J. R. Cano, F. Herrera, and M. Lozano, "Using evolutionary algorithms as instance selection for data reduction in KDD: An experimental study," *IEEE Trans. Evol. Comput.*, vol. 7, no. 6, pp. 561–575, Dec. 2003.

[30] J. R. Cano, F. Herrera, and M. Lozano, "Instance selection using evolutionary algorithms: An experimental study," in *Advanced Techniques in Knowledge Discovery and Data Mining*. New York: Springer-Verlag, 2005, pp. 127–152.

[31] J. R. Cano, F. Herrera, and M. Lozano, "Stratification for scaling up evolutionary prototype selection," *Pattern Recognit. Lett.*, vol. 26, pp. 953–963, 2005.

[32] S. Garcia and F. Herrera, "Evolutionary undersampling for classification with imbalanced dataset: Proposals and taxonomy," *Evol. Comput.*, vol. 17, no. 3, pp. 275–306, 2009.

[33] C. R. Reeves and D. R. Bush, "Using genetic algorithms for training data selection in RBF networks," in *Instance Selection and Construction for Data Mining*, H. Liu and M. Hiroshi, Eds. Norwell, MA: Kluwer, 2001, pp. 339–356.

[34] K. Kim, "Artificial neural networks with evolutionary instance selection for financial forecasting," *Expert Syst. Appl.*, vol. 30, pp. 519–526, 2006.

[35] J. R. Cano, F. Herrera, and M. Lozano, "Evolutionary stratified training set selection for extracting classification rules with trade-off precision-interpretability," *Data Knowl. Eng.*, vol. 60, pp. 90–108, 2007.

[36] J. R. Cano, S. Garcia, and F. Herrera, "Subgroup discovery in large size datasets preprocessed using stratified instance selection for increasing the presence of minority classes," *Pattern Recognit. Lett.*, vol. 29, pp. 2156–2164, 2008.

[37] C. F. Juang, R. B. Huang, and W. Y. Cheng, "An interval type-2 fuzzy-neural network with support-vector regression for noisy regression problems," *IEEE Trans. Fuzzy. Syst.*, vol. 18, no. 4, pp. 686–699, Aug. 2010.

[38] R. R. Yager and G. Beliakov, "OWA operators in regression problems," *IEEE Trans. Fuzzy. Syst*, vol. 18, no. 1, pp. 106–113, Feb. 2010.

[39] C. F. Juang and C. D. Hsieh, "A locally recurrent fuzzy neural network with support vector regression for dynamic-system modeling," *IEEE Trans. Fuzzy. Syst.*, vol. 18, no. 2, pp. 261–273, Apr. 2010.

[40] M. H Mashinchi, M. A Orgun, M. Mashinchi, and W. Pedrycz, "A tabu–harmony search-based approach to fuzzy linear regression," *IEEE Trans. Fuzzy. Syst*, vol. 19, no. 3, pp. 432–448, Jun. 2011.

[41] Z. Deng, K. S. Choi, F. L. Chung, and S. Wang, "Scalable TSK fuzzy modeling for very large datasets using minimal-enclosing-ball approximation," *IEEE Trans. Fuzzy. Syst.*, vol. 19, no. 2, pp. 210–226, Apr. 2011.

[42] A. Guillen, L. J. Herrera, G. Rubio, H. Pomares, A. Lendasse, and I. Rojas, "New method for instance or prototype selection using mutual information in time series prediction," *Neurocomputing*, vol. 73, pp. 2030–2038, 2010.

[43] R. Alcalá, J. Alcalá-Fdez, F. Herrera, and J. Otero, "Genetic learning of accurate and compact fuzzy rule based systems based on the 2-tuples linguistic representation," *Int. J. Approx. Reason.*, vol. 44, pp. 45–64, 2007.

[44] J. V. de Oliveira, "Semantic constraints for membership function optimization," *IEEE Trans. Syst., Man., Cybern. A*, vol. 29, no. 1, pp. 128–138, Jan. 1999.

[45] S. M. Zhou and J. Q. Gan, "Low-level interpretability and high-level interpretability: A unified view of data-driven interpretable fuzzy system modeling," *Fuzzy Sets Syst.*, vol. 159, pp. 3091–3131, 2008.

[46] C. Mencar and A. M. Fanelli, "Interpretability constraints for fuzzy information granulation," *Inf. Sci.*, vol. 178, pp. 4585–4618, 2008.

[47] J. M. Alonso, L. Magdalena, and G. Gonzalez-Rodriguez, "Looking for a good fuzzy system interpretability index: An experimental approach," *Int. J. Approx. Reason*, vol. 51, pp. 115–134, 2009.

[48] L. Gál, R. Lovassy, and L. T. Kóczy, "Function approximation performance of fuzzy neural networks based on frequently used, fuzzy operations and a pair of new trigonometric norms," in *Proc. IEEE World Congr. Comput. Intell.*, Barcelona, Spain, 2010, pp. 1514–1521.

[49] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. de Fonseca, "Performance assessment of multiobjective optimizers: An analysis and review," *IEEE Trans. Evol. Comput.*, vol. 7, no. 2, pp. 117–132, Apr. 2003.

[50] J. D. Knowles, L. Thiele, and E. Zitzler, "A tutorial on the performance assessment of stochastic multiobjective optimizers," in *Comput. Eng. Netw. Lab., ETH Zurich*, Zurich, Switzerland, TIK-Rep. 214, 2006.

[51] S. Bleuler, M. Laumanns, L. Thiele, and E. Zitzler, "PISA—Platform and programming language independent interface for search algorithms," in *Evolutionary Multi-Criterion Optimization*, M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, Eds. vol. LNCS-2632. Heidelberg, Germany: Springer-Verlag, 2003, pp. 494–508.

[52] J. Derrac, S. García, D. Molina, and F. Herrera, "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms," *Swarm Evol. Comput.*, vol. 1, no. 1, pp. 3–18, 2011.

[53] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *J. Amer. Statist. Assoc.*, vol. 32, pp. 675–701, 1937.

[54] R. L. Iman and J. H. Davenport, "Approximations of the critical region of the Friedman statistic," *Comm. Statist. Part A Theory Methods*, vol. 9, pp. 571–595, 1980.

[55] S. Holm, "A simple sequentially rejective multiple test procedure," *Scand. J. Statist.*, vol. 6, pp. 65–70, 1979.

Michela Antonelli received the M.Sc. degree in computer engineering and the Ph.D. degree in information engineering from the University of Pisa, Pisa, Italy, in 2003 and 2007, respectively.

She is currently a Research Fellow with the Department of Information Engineering, University of Pisa. During her Ph.D., she developed a computer-aided diagnosis system that is able to automatically detect and diagnose lung nodules in computed tomography scans. She has coauthored more than 20 papers in international journals and conference proceedings. She serves as reviewer for a number of international journals and conferences. Her research interests include the field of multiobjective evolutionary fuzzy systems with the aim of designing both accurate and interpretable fuzzy rule-based systems.

Pietro Ducange (M'09) received the M.Sc. degree in computer engineering and the Ph.D. degree in information engineering from the University of Pisa, Pisa, Italy, in 2005 and 2009, respectively.

He is currently a Research Fellow with the Department of Information Engineering, University of Pisa. He has coauthored more than 20 papers in international journals and conference proceedings. His current research interests mainly focus on designing fuzzy rule-based systems with different tradeoffs between accuracy and interpretability by using multiobjective evolutionary algorithms.

Dr. Ducange currently serves the following international journals as a member of the Reviewer Board: IEEE TRANSACTIONS ON FUZZY SYSTEMS, *Information Sciences*, *International Journal of Approximate Reasoning*, and *Soft Computing*. He is involved in several research and development projects mainly founded by the Tuscany Region.

Francesco Marcelloni (M'07) received the Laurea degree in electronics engineering and the Ph.D. degree in computer engineering from the University of Pisa, Pisa, Italy, in 1991 and 1996, respectively.

He is currently an Associate Professor with the Faculty of Engineering, University of Pisa. He cofounded the Computational Intelligence Group at the Department of Information Engineering, University of Pisa, in 2002. He is also the Founder and Head of the Competence Centre on Mobile Value Added Services. He has co-edited two volumes, two journal special issues, and is (co-)author of a book and more than 160 papers in international journals, books, and conference proceedings. His research interests include multiobjective evolutionary algorithms, genetic fuzzy systems, fuzzy clustering algorithms, pattern recognition, signal analysis, neural networks, mobile information systems, and data compression and aggregation in wireless sensor networks.

Dr. Marcelloni has been a Technical Program Committee Co-Chair of the International Conference on Intelligent Systems Design and Applications (ISDA) 2009 and ISDA 2011, General Co-Chair of ISDA 2010, and Co-Organizer of several workshops and special sessions. He is an Associate Editor of *Information Sciences*, the *Journal of Computational Intelligence Research*, the *Journal of Swarm Intelligence and Evolutionary Computation*, and the *Journal of Sensor Networks and Data Communications*. He has been the main investigator of several projects supported by the European Commission, the Italian Ministry, the Tuscany region, and many private companies.