An efficient adaptive fuzzy inference system for complex and high dimensional regression problems in linguistic fuzzy modelling

Antonio A. Márquez, Francisco A. Márquez, Ana M. Roldán, Antonio Peregrín*

University of Huelva, Dept. of Information Technologies, University of Huelva, Huelva 21819, Spain

A R T I C L E I N F O

Article history:

Keywords
Linguistic fuzzy modelling
Multi-objective genetic fuzzy systems
Adaptive Inference Systems
High-dimensional regression problems

A B S T R A C T

The use of adaptive connectors as conjunction operators in adaptive fuzzy inference systems is one of the methodologies, also compatible with others, to improve the accuracy of fuzzy rule-based systems by means of local adaptation of the inference process to each rule of the rule base. However, when dealing with such currently challenging issues as high-dimensional regression problems, adapting their parameters becomes difficult due to the exponential rule explosion.

In this paper, we propose to address the problem by using a new adaptive conjunction operator. This operator provides considerable advantages in efficiency while maintaining the accuracy. Moreover, it is completed with a multi-objective evolutionary algorithm as a search method due to its efficiency in achieving different balances between complexity and accuracy in the learned fuzzy systems.

An in-depth experimental study is performed to show the advantages of the proposal presented, using 17 regression problems of different size and complexity, using different rule bases, analysing the multi-objective algorithms and Pareto fronts obtained and performing statistical analyses. It confirms its effectiveness in terms of efficiency, but also in terms of accuracy and complexity of the obtained models.

1. Introduction

Fuzzy rule-based system (FRBS) design using genetic algorithms leads the Genetic Fuzzy Systems area [23, 33] and constitutes a broad research area with many different branches, methodologies and applications [32]. Some of the most widely known approaches are based on improvements to the Knowledge Base, which in fact usually have greater impact on the accuracy.

Nevertheless, there are other Genetic Fuzzy Systems designed to deal with other important elements of FRBSs, such as those based on the setup of the Inference System and Defuzzification Interface. The most interesting point is that they are complementary to those based on Knowledge Base improvements, and both together can reach higher levels of quality.

In the trade-off between interpretability and accuracy in fuzzy modelling, Adaptive Inference Systems and defuzzification methods have taken greater importance [5, 44]. Focusing on the Adaptive Inference System, the most common approach is to use adaptive conjunction operators. Usually, one or more parameters are employed in their expressions to modify their behaviour, allowing us to customize the way each rule infers.

The use of multi-objective evolutionary algorithms (MOEA) is currently applied to improve the aforementioned trade-off between interpretability and accuracy of linguistic fuzzy systems [1, 2, 4, 7, 11, 18, 26, 27, 35, 36, 46]. Some of them get the complete Pareto front (the set of non-dominated solutions with different trade-offs) by selecting or learning the set of rules which best represents the example data, i.e., improving the system accuracy and decreasing the FRBS complexity. Recently, in [40, 41, 42] we also proposed evolutionary multi-objective learning models, achieving cooperation between the Rule Base (RB) and adaptive fuzzy operators of the Inference System to obtain simpler, more compact and still accurate linguistic fuzzy models.

In this context, one of the most important challenges nowadays is the design of models that deal with high-dimensional and large scale regression datasets. The different techniques to perform the learning or improve the accuracy or

* Corresponding author.
E-mail address: {amarquez (A. A. Márquez), alfredo.marquez (F.A. Márquez), amroldan (A. M. Roldán), peregrin (A.Peregrín)}@dti.uhu.es.

interpretability in the sense of complexity of linguistic FRBSs entail two problems: the efficiency of the process and the complexity of the models obtained. Some methods widely known and employed cannot be used with high dimensional data because they need extraordinarily large computational resources: time and memory. Usually, when the number of variables and/or data patterns becomes high, most of the algorithms employed in the design of linguistic FRBSs suffer from exponential rule explosion [14, 38], and this high amountof rules makes the model obtained difficult to manage. Some authors address this problem by proposing three types of solutions: the first is data reduction, which includes instance selection, feature selection and linguistic partition reduction [3, 8, 49] to reduce the number of instances and rules; the second is the use of the fitness approximation approach to reduce the execution time by computing an approximation of the fitness function rather than by evaluating the complete fitness function [12, 13, 39]; and the other method is by applying parallel or distributed methodologies [9, 19, 47]. Recently, interesting hybrid methods between parallel implementations and data reduction have also been considered in the literature when dealing with high dimensional problems [37].

We must emphasise that in this context of high-dimensional and large scale regressions datasets, the use of MOEAs to obtain models with the aforementioned different trade-offs between accuracy and interpretability in the sense of complexity is particularly concerned with complexity, because when dealing with a very high number of rules, interpretability is considered in second place.

In this framework, we propose a new adaptive conjunction operator suitable for use with high-dimensional problems, which is more efficient (faster) and also has an associated learning method for the parameters that lets us achieve more accurate and simpler FRBSs. The learning method we propose is a MOEA which is able to generate a set of FRBSs with different optimal trade-offs between accuracy and complexity.

To show the advantages of the proposed method, it is tested on 17 different problems with a number of variables ranging from 6 to 85 and a number of samples ranging from 337 to 20,640. In addition, we compared our proposal against three of the most usual adaptive conjunction operators from the literature. The results obtained reveal the effectiveness of the proposed method, particularly in terms of scalability, but also in terms of capability of generalization of the obtained models.

This contribution is organized as follows: Section 2 reviews the Adaptive Inference Systems. In Section 3 we present the new adaptive conjunction operator and its multi-objective evolutionary learning algorithm associated. Section 4 is devoted to the experimental study, with a description, results analysis and a reference to the paper's associated website (http://www.uhu.es/gisimd/papers/ACO-HDP), which contains complementary material to this study. Finally, Section 5 gives some concluding remarks.

2. Preliminaries: Adaptive Inference System

In this section, the foundation of the Adaptive Inference System is reviewed, setting out from concepts and followed by the adaptive models.

2.1. Inference Components

As is well known, linguistic FRBSs for system modelling use IF - THEN rules of the following type:

$$R_i : \text{If } X_{i1} \text{ is } A_{i1} \text{ and ... and } X_{in} \text{ is } A_{in} \text{ then } Y \text{ is } B_i \tag{1}$$

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

The expression of the Compositional Rule of Inference in fuzzy modelling with punctual fuzzification is as follows:

$$\mu_{B'}(y) = I\left(C\left(\mu_{A1}(x_1), \ldots, \mu_{An}(x_n)\right), \mu_B(y)\right) \tag{2}$$

where $\mu_{B'}$ is the membership function of the inferred consequent, I is the rule connective or implication operator, C is the conjunction operator or antecedent connection, μ_{Ai} are the values of the *matching degree* of each input of the system with the membership functions of the rule antecedents, and μ_B is the consequent of the rule.

Therefore, the Inference System performs the two following tasks:

1) First, it computes $C(\mu_{A1}(x_1), \ldots, \mu_{An}(x_n))$, which is the aforementioned matching degreeof each rule. Usually, the conjunction operator $C(.)$ is modelled with a t-norm.

2) Second, it infers using the fuzzy rule connective $I(.)$, the previously computed matching degree and the consequent of the rule. Fuzzy rule connectives can be classified into different families, of which implication functions [50] and t-norms [31] are the best known. T-norms are the most and widely used in practical fuzzy modelling and control systems.

Frequently, t-norms are used as conjunction operator and inference operator [17, 31, 48]. The expressions of the most used classic t-norms are:

$$T_{Minimum}(x, y) = Min(x, y) \tag{3}$$

$$T_{Hamacher}(x, y) = \frac{x \cdot y}{(x + y - x \cdot y)} \tag{4}$$

$$T_{Algebraic}(x, y) = x \cdot y \tag{5}$$

$$T_{Einstein}(x, y) = \frac{x \cdot y}{(1 + (1 - x) \cdot (1 - y))} \tag{6}$$

$$T_{Bounded}(x, y) = Max(0, x + y - 1) \tag{7}$$

$$T_{Drastic}(x, y) = \begin{cases} x, & \text{if } y = 1 \\ y, & \text{if } x = 1 \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

2.2. Classical Adaptive Inference Systems

The aforementioned two components, conjunction and rule connective, are suitable for parameterization to get an Adaptive Inference System. If we name the parameter of the adaptive conjunction α, the resulting adaptive component will be $C(\alpha)$. Similarly, if β is used to name the parameter of the inference operator, the resultant adaptive component will hence be $I(\beta)$.

Conventionally, two models of Adaptive Inference System can be considered depending on the amount of parameters they use: a single parameter to tune the behaviour of the connector globally, or individual parameters for every rule, having a local tuning mechanism for the behaviour of the Inference System for every fuzzy rule.

If we talk about conjunction, the single parameter model lets us adapt the behaviour of the operator globally between the classic t-norms. However, the benefits of this model will not yield remarkable improvements in accuracy. The reason is the low influence of the conjunction operator on the design of linguistic fuzzy systems [16] with a similar behaviour, with the use of different t-norms considering the same operator for all rules.

On the contrary, the model that uses individual parameters for each rule has a mechanism to adapt the behaviour of the conjunction individually for every rule. This model shows the highest accuracy in [5] because of its higher number of degrees of freedom.

Rule connective showed in our previous studies [5, 44] that the model based on the adaptive conjunction operator is a more valuable option than the one based on the adaptive rule connective in terms of improving the accuracy of linguistic Fuzzy Systems. Consequently, we selected the use of the adaptive conjunction in this study and specifically for each rule separately to parameterize the Inference System. Table 1 exemplifies three classic adaptive t-norms [45] with their corresponding α parameter. Table 2 shows the relation between the six classic t-norms and the values of the α parameter of the adaptive t-norms.

In the studies performed in [5, 44] the Dubois t-norm with a separate connector for every rule showed the highest accuracy compared with Frank and Dombi t-norms. In addition, it was seen that it is also more efficiently computed. We also established that the adaptive conjunction supports the cooperation within the rules of the Knowledge Base.

In [52] an interesting study was carried out, seeking better performance than with traditional minimum or product t-norms for the antecedent connections. The authors suggested the use of adaptive t-norms to look for better performance than traditional minimum or product t-norms. They studied the use of adaptive connectors that are extended from t-norms and t-conorms in order to cover the range between them, including *compensatory and* and S-OWA operators and many others. Fig. 1 shows the ranges covered by them. Other recent studies can be found in[10].

The use of MOEAs to adapt the parameters of the adaptive t-norm was introduced in [42] to generate a set of models, in a single execution of the learning process, with different trade-offs between accuracy and complexity while looking for the best cooperation among the Adaptive Inference System and the RB.

Table 1
Adaptive t-norms.

Name (Domain)	Expression	
Dombi (α>0)	$T_{Dombi}(x,y,\alpha) = \dfrac{1}{1+\sqrt[\alpha]{\left(\frac{1-x}{x}\right)^{\alpha}+\left(\frac{1-y}{y}\right)^{\alpha}}}$	(9)
Frank (α>0)(α≠1)	$T_{Frank}(x,y,\alpha) = log_{\alpha}\left[1+\dfrac{(\alpha^{x}-1)(\alpha^{y}-1)}{\alpha-1}\right]$	(10)
Dubois (0≤α≤1)	$T_{Dubois}(x,y,\alpha) = \dfrac{x\cdot y}{Max(x,y,\alpha)}$	(11)

Table 2
Relation between classic and adaptive t-norms depending on the parameter.

	T_{Min}	T_{Ham}	T_{Alg}	T_{Ein}	T_{Bou}	T_{Dras}
T_{Dombi}	∞	1				$\to 0$
T_{Frank}	$\to 0$		$\to 1$		∞	
T_{Dubois}	0		1			

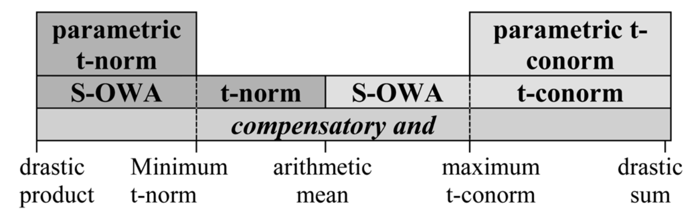


Fig. 1. Ranges covered by adaptive t-norms, adaptive t-conorms, *compensatory and*, and S-OWA operators.

3. ACO$_{HDP}$: A proposal for an adaptive conjunction operator for high dimensional problems

The set of continuous domain parameters of the adaptive conjunction are commonly learned using real coding evolutionary algorithms. They are a good choice to find the corresponding values for each rule of the RB, as shown in our previous studies [5, 41, 42, 44], but when dealing with large and complex datasets, the number of rules can be very high and consequently so can the number of parameters, making it harder to adjust due to the huge search space tackled.

Furthermore, when we use adaptive t-norms such as those of Dombi or Frank (expressions 9 and 10 in Table 1), the fitness computation of the evolutionary algorithm considerably slows the search process.

Both drawbacks result in low quality solutions and poor efficiency. To solve this problem, in this paper we propose the use of an Adaptive Conjunction Operator for High Dimensional Problems (ACO_{HDP}), which is a parameterized t-norm that uses discrete values for its parameter that allow us to select among different classic t-norms, reducing the search space and allowing to use a *look-up* table. Next, we describe the t-norm selector, the MOEA proposed and the look-up table.

3.1. T-norm selector

This mechanism selects the most appropriate classic t-norm for each rule and therefore the search space is greatly reduced, greatly improving the speed of the learning process.

The lower number of degrees of freedom of the proposed discrete t-norm selector against a continuous parameters scheme should be offset by the higher quality results achievable due to a major convergence of the evolutionary algorithm handling a reduced search space.

We propose to use several of the classic t-norms combined with some specific t-norms derived from the use of specific values for the parameter of an adaptive t-norm. Here, we present the different options considered:

- The classical t-norms are the Minimum, Hamacher and Algebraic Product (expressions 3, 4 and 5). They were selected because they are the most widely used as conjunction operators, with their simplicity and good computing performance shown in [17].
- Using the Dubois t-norm (expression 10) with the following eight values for α, which are 0.1, 0.2, 0.3, 0.4, 0.6, 0.7, 0.8, 0.9, which are t-norms among the Minimum and Algebraic Products, i.e. those that have shown better performance in previous studies.

3.2. Multi-objective Evolutionary Algorithm

This section presents the MOEA proposed. We selected a MOEA based on the popular NSGA-II [20]. It is one of the most known and widely used MOEAs in the literature for solving multi-objective problems. The offspring population is generated from the current population through selection, crossover and mutation. The next generation is built from the current population and the offspring until it reaches the stop condition in this work, the number of evaluations. The NSGA-II algorithm has two features that make it one of the main and most important MOEAs: i) the assignment of fitness based on Pareto ranking and crowding operator, and ii) the procedure for updating each generation through elitism.

The components needed to implement this algorithm are explained in detail below.

1) Codification Scheme: In this paper, we use an integer coding scheme with size M (with M representing the number of rules of RB) in which the conjunction operator of the different rules are coded,

$$C = (O^1, \ldots, O^M)$$

Each gene O^i represents the parameter used by the rule i-th and takes values in the set {1,…,11}. Additionally, in some cases, it can take a value equal to 0 to determine that the corresponding rule is not used.
2) Objectives: we minimize the following:
    a. The number of rules, as a complexity measure.
    b. The Mean Square Error (MSE) which measures the accuracy of the system.

$$MSE\left(FM\right) = \frac{\frac{1}{2}\sum_{k=1}^{P}\left(y_k - FM(x_k)\right)^2}{P} \qquad (12)$$

where FM denotes the fuzzy model whose Inference System uses the adaptive conjunction operator ACO_{HDP}; the inference operator is the Minimum t-norm, and the defuzzification method is the centre of gravity weighted by the matching degree. This measure uses a set of system evaluation data formed by *P* pairs of numerical data $Z_k = (x_k, y_k)$, $k=1,..,P$, with x_k as the values of the input variables, and y_k as the corresponding values of the associated output variables.
3) Initial Gene Pool: The initial population will be comprised of two different subsets of individuals: An individual from the initial population has all the genes initially set to 1 to begin the evolutionary process with all the rules with the conjunction operator as the Minimum t-norm. The remaining individuals of the initial population are created randomly.
4) Crossover and Mutation Operators: The crossover operator employed is HUX [22]. This operator exchanges exactly half of the genes that are different in the parents, ensuring the maximum distance between children and their parents (exploration). This way, two individuals are obtained by combining the two offspring generated from parents. For each of them, the mutation operator changes a gene value at random with probability P_m. The mutation operator randomly determines other possibility in {0, .. , 11}.

3.3. Look-up table

The use of classic adaptive conjunction operators implies that during the evolutionary learning phase of the parameters, each chromosome evaluation needs to assess the matching degree for each rule. The time required for this evaluation is determined, among other factors, by the number of variables and number of rules used in the RB. Thus, this process can take a significant amount of time especially in high dimensional problems. Due to the number of evaluations required, it poses a problem that we can improve by using a *look-up table (LUT)*. During the learning phase of the parameters, the matching degree can be pre-computed and stored in a LUT, as there are only 11 different t-norms against a classical continuous adaptive t-norm with infinite possible values (so it cannot be pre-computed).

The LUT consists of a two dimensional table (examples and rules). Each element of this table is a structure with the corresponding matching degree for each of the 11 t-norms for each example of the learning dataset and rule of the RB. To avoid the unnecessary use of memory, the structure is not allocated into the memory if the example does not trigger that rule. Consequently, the corresponding computations of the matching degree are performed once at the beginning of the process.

The above implementation entails an important consequence: the execution time required by the proposed model also depends on the rules triggered by each instance.

4. Experimental study and analysis of results

To evaluate the usefulness of the proposed approach when dealing with high dimensional problems, we selected 17 problems with different number of variables and patterns/data samples. Table 3 summarizes the main features of the different problems considered and their link to the KEEL dataset repository [6], Torgo site and UCI Machine Learning Repository where they can be downloaded. Problems were selected from lower to higher complexity, covering a range from 6 to 85 input variables and from 337 to 20,640 examples (even though each of them is difficult in itself in terms of the modelling task). Anyway, the more complex problems are ELV, CA, AIL, and TIC because of the large number of variables and instances. To the best of our knowledge, these problems have never been solved using classical adaptive fuzzy Inference Systems. This is due to the long time needed to evaluate each candidate solution and to the minimum number of the evaluations needed to reach convergence of evolutionary algorithm. The large number of rules that these big problems produce make it particularly difficult to obtain their corresponding adaptive inference linguistic fuzzy models. These problems therefore pose an important challenge for the proposed algorithm.

Table 3
Datasets considered for the experimental study.

Problem	Abbr.	Instances	Variables	Repository
Delta_elv	DEL	9517	6	KEEL
Abalone	ABA	4177	8	KEEL
California	CAL	20640	8	KEEL
Concrete	CON	1030	8	KEEL
Kinematics Robot	KIN	8192	8	Torgo
Puma8nh	PUM	8192	8	Torgo
Stock	STP	950	9	KEEL
Weather Ankara	WAN	1609	9	KEEL
Wine-red	WINR	1599	11	UCI
Wine-white	WINW	4898	11	UCI
Forest Fires	FOR	517	12	KEEL
Mortgage	MOR	1049	15	KEEL
Baseball	BAS	337	16	KEEL
Elevators	ELV	16599	18	KEEL
Computer-Activity	CA	8192	21	KEEL
Ailerons	AIL	13750	40	KEEL
The Insurance Co.	TIC	9822	85	KEEL

Available at KEEL (http://sci2s.ugr.es/keel/datasets.php), Torgo (http://www.liaad.up.pt/~ltorgo/Regression/DataSets.html) and UCI (http://www.ics.uci.edu/~mlearn/MLSummary.html) repositories.

This section is organized as follows:

- First, we describe the experimental setup and introduce the information shown on the website accompanying the paper in Subsection 4.1.
- Second, Subsection 4.2 focuses on comparing the most accurate solutions of our proposal to three well-known classic adaptive fuzzy Inference Systems.
- Third, we show the computational costs of the different algorithms and discuss the scalability of the proposed approach. This can be found in Subsection 4.3.
- Finally, Subsection 4.4 contains the comparison between the Pareto fronts of our proposal with those of three well-known models using classical adaptive fuzzy Inference Systems, for each of the datasets. Hence, we plot the average Pareto fronts to provide reliable information about the shape and characteristics of the Pareto fronts obtained, allowing us to check and analyse the trend and the kind of correlation between the training and test errors.

4.1. *Experimental framework*

This section describes the experimental setup, including a brief description of the methods and the non-parametric statistical tests considered for the performed comparisons. Then, it introduces the contents of the paper's companion website, which includes additional material associated with the paper.

4.1.1. Experimental setup

To evaluate the effectiveness of the proposed method, designed specifically for fast learning and its applicability to high dimensional problems, three well-known classic adaptive conjunction operators were considered for comparison: the adaptive conjunction operators using the adaptive t-norms of Dubois, Dombi and Frank [45]. The Adaptive Inference Systems built with them were studied in depth in [44] and later in [42] by using a MOEA. Now we shall designate the linguistic fuzzy systems that use them as ACO_{DUB}, ACO_{DOM} and ACO_{FRA} respectively.

The learning procedure for the conjunction parameter of all of them, ACO_{HDP}, ACO_{DUB}, ACO_{DOM} and ACO_{FRA} is the same: we use a MOEA with a rule selection mechanism and two objectives: the model accuracy, which is better when higher, and number of rules, which is better when lower. The coding scheme for ACO_{HDP} was described previously, although in the case of the other models (ACO_{DUB}, ACO_{DOM} and ACO_{FRA}), we shall use a double coding chromosome scheme (C_C+C_S) where:

- C_C encodes the α_i parameters of the conjunction connector. They are m real coded parameters (genes), one for each rule, R_i of the linguistic RB. Each gene can take any value in the interval of classic adaptive t-norms parameters.

- C_S encodes the rule selection. It is a binary string of m genes, each one representing a candidate rule of the initial RB. Depending on whether a rule is selected or not, values '*1*' or

'*0*' are respectively assigned to the corresponding gene.

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

- C_C is initialized with the *m* genes initiated as follows:
  - When we use Dubois t-norm, we initiated them to *0* in order to make this t-norm equivalent to Minimum t-norm at the beginning of the searching process.
  - When we use Dombi t-norm, we initiated them to *1* to make this t-norm initially equivalent to Minimum t-norm.
  - When we use Frank t-norm, we initiated them to *0.5* to make this t-norm begins with the same value than Hamacher t-norm.
- C_S is initialized with the all rules selected, that is, with the *m* genes to *1*.

Regarding the crossover operator employed by the MOEAs, for the fuzzy operators part, C_C, is BLX-0.5 [22, 34] while the one used for the rule learning part, C_S, is HUX [22].

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

The population size used was 61 individuals. The evolutionary algorithms performed 200000 evaluations. The P_m both for the C_C and C_S parts were fixed to 0.3 for the models with ACO_{HDP}, while the rest of the models were fixed to 0.2 because this probability presents the best behaviour in these models.

Concerning the Knowledge Bases, linguistic partitions were considered consisting of 5 triangular shaped linguistic terms for all problems. The initial set of candidate rules of the RBs were obtained using the well-known ad-hoc data-driven algorithms of Wang and Mendel [51] (WM) and the first stage of MOGUL-IRL [15], which we name FS-MOGUL. No other methods from the literature were used, due to the problems they present when they perform on a high number of variables like the datasets used in this work.

Tables 4 and 5 shows the average number of rules (#R) and MSE for training and test datasets of the FRBSs using the WM and FS_MOGUL RBs. They are shown simply as a reference. It is not the aim of this paper to analyse the improvements of the adaptive conjunction operators, as it was analysed in depth in previous works cited in Section 2.

For all our experiments, we considered a 5-fold cross-validation model, i.e. 5 random partitions of data each with 20%, and the combination of four of them (80%) as training and the remaining one as test. For each of the 5 data partitions, the methods were run six times using different seeds for the random number generators, showing for each problem the averaged results of a total of 30 runs. For each dataset and for each trial we generate the approximated Pareto front and then, we consider the average results of 30 Pareto fronts. In order to compare the different approaches we focus on three different points of the aforesaid Pareto front: the most accurate (MAX ACC), the accurate/interpretable midpoint (MEDIUM ACC/INT) and the most interpretable (simple) point (MAX INT). For each dataset, we compute the mean values over the 30 trials of the MSE on the training and test sets and the number of rules. Our main aim in this approach is to make statistical comparisons between different approaches of the most accurate solutions and show the tendency of the rest of the Pareto fronts.

Table 4
Reference values of average number of rules and MSE of the FRBSs built with the RB obtained with the WM method. Values of MSE in this table must be multiplied by 10^{-6}, 10^{8}, 10^{-4}, 10^{2}, 10^{4}, 10^{-6}, 10^{-8} and 10^{-4} in the case of DEL, CAL, KIN, FOR, BAS, ELV, AIL or TIC respectively.

Name	#R	MSE_{TRA}	MSE_{TST}
DEL	679.8	1.624	1.684
ABA	199.0	3.341	3.477
CAL	623.8	38.323	38.712
CON	309.8	35.388	47.644
KIN	6423.2	47.688	120.523
PUM	6435.6	3.281	9.256
STP	265.4	1.453	1.487
WAN	456.8	4.878	6.128
WINR	714.2	0.229	0.251
WINW	1019.8	0.296	0.306
FOR	374.6	14.350	342.347
MOR	198.8	0.128	0.134
BAS	252.6	7.821	64.852
ELV	4322.0	11.591	12.271
CA	1539.4	8.449	12.440
AIL	6697.0	2.431	2.950
TIC	6597.8	73.428	904.269

Table 5
Reference values of average number of rules and MSE of the FRBSs built with the RB obtained with the FS_MOGUL method. Values of MSE in this table must be multiplied by 10^{-6}, 10^{8}, 10^{-4}, 10^{2}, 10^{4}, 10^{-6}, 10^{-8} and 10^{-4} in the case of DEL, CAL, KIN, FOR, BAS, ELV, AIL or TIC respectively.

Name	#R	MSE_{TRA}	MSE_{TST}
DEL	589.8	2.979	3.426
ABA	190.0	14.461	14.646
CAL	713.2	132.370	133.311
CON	645.2	74.397	99.676
KIN	10000.0	132.381	282.011
PUM	10000.0	11.247	23.925
STP	156.0	3.297	5.332
WAN	296.8	15.017	18.553
WINR	652.4	0.406	0.642
WINW	577.2	0.585	0.628
FOR	747.2	27.750	751.659
MOR	98.8	0.475	0.488
BAS	500.0	16.060	128.301
ELV	1747.6	61.887	64.067
CA	970.8	25.550	35.050
AIL	2722.4	6.408	19.013
TIC	10000.0	928.853	1939.588

We also adopted the Two Set Coverage [53] (CS) ratio as a tool to compare the Pareto fronts of different multi-objective approaches, also used in [43]. CS considers $X', X'' \subseteq X'$ as two sets of phenotype decision vectors and a' and a'' are two points belong to sets X' and X'', respectively. CS is defined as the mapping of the order pair (X', X'') to the interval [0, 1] per equation (13).

$$\mathrm{CS}(X', X'') = \frac{|\{a'' \epsilon X'';\ \exists\, a' \epsilon X' : a' \geq a''\}|}{|X''|} \quad (13)$$

If all points in *X'* dominate or are equal to all points in *X''*, then by definition CS=*1*. CS=*0* implies the opposite. In general, CS(*X'*,*X''*) and CS(*X''*,*X'*) both have to be considered due to the set intersection not being empty. The advantage of this metric is that it is easy to calculate and provides a relative comparison between MOEAs.

To compare the different MOEAs we perform different CS ratios for each partition and seed, then compare each of them. The algorithm to carry out this process will be explained in subsection 4.4**.**

To assess whether there are significant differences among the results, we use statistical analysis [21, 28, 29, 30] and in particular non-parametric tests, according to the recommendations made in [28]. We employ different approaches for multiple comparisons, including Friedman's test[25], to detect statistical differences among a group of results, and the Finner post-hoc test[24] to observe the difference in performance between the methods and the retention or rejection of the hypothesis with the level of significance fixed. To perform the tests, we used a level of confidence α = 0.05. In particular, these tests are based on computing the differences on sample means (typically, mean test errors obtained by a pair of different algorithms on different datasets). In the classification framework, these differences are well defined, as the errors are the same domain. In the regression framework, to make the differences comparable, we adopt a normalized difference DIFF, defined as:

$$DIFF = \frac{Mean\ (other\ ) - Mean\ (reference\ )}{Mean\ (other\ )} \quad (14)$$

where Mean(x) represents the MSE or #R obtained by the x algorithm. This difference expresses the improvement percentage of the reference algorithm on the other one.

4.1.2. Website accompanying paper:

To provide additional information to the paper content of the experimental study carried out, we have created a website (accessible at (http://www.uhu.es/gisimd/papers/ACO-HDP)) including the following information:

- The partitions of the datasets employed in the study of the paper. These partitions can be found in a table together with the main characteristics of the datasets.
- An Excel file with the complete tables of results of the experimental study. We include a file with the training and test sets results for all the algorithms so any interested researcher can use them to include their own results and extend the present comparison. We also show all Pareto fronts in the three different points (MAX ACC, MEDIUM ACC/INT and MAX INT) obtained with the 30 runs of different models both with the initial RBs of WM and FS-MOGUL. Figures with the average Pareto fronts for all the datasets used have been also added.
- A table where the readers can see the influence of the different factors (the initial #R, the number of instances in the training data set and the number of rules triggered per instance) in the time gained by ACO_{HDP}.

4.2. Results and analysis of the most accurate solutions

The results obtained by the FRBSs considered at the point of maximum accuracy are shown in Table 6 and 7. These tables are grouped in columns by FRBSs models based on the four different adaptive conjunctions compared (the new proposal and the three classic ones) and show the average of the results obtained by each model for all the datasets used. For each one, the first column shows the average #R and the average MSE for training and test datasets. Note that noinformation is shown in the cases of ACO_{DUB}, ACO_{DOM} and ACO_{FRA} for the most complex and largest four datasets of the study at the end of the tables, because the large number of variables and instances make them impossible to compute, despite the authors' efforts to optimize and improve the algorithms.

Table 8 shows the rankings obtained with the Friedman's test using both RBs. They tell us that there are significant differences among the results observed with all datasets when the p-Fried <0.05. Indeed, for MAX ACC there are significant differences in MSE_{TST} and #R. Particularly for MSE_{TST}, the best ranking is obtained by ACO_{HDP}. On the other hand, for the #R the best ranking is obtained by ACO_{DOM} when using WM RBs and by ACO_{FRA} when using FS-MOGUL RBs. However, we can point out that in both cases our proposal is very close to them.

Finally, Tables 9 and 10 show the results when using Finner's post-hoc procedure to compare the best ranking model in each case with the remaining ones. Algorithms are ordered by the p-value obtained compared to the control algorithm. Finner's test rejects the hypothesis of equality when the p-Finner is < 0.05. Indeed, the Finner test rejects the hypothesis of equality for MSE_{TST} when the control algorithm is ACO_{HDP} except for ACO_{DUB}. However, we highlight that ACO_{HDP} is still better than ACO_{DUB} in terms of *#R*.

Table 6

Average results of the four different adaptive conjunction FRBS models for the most accurate point regarding complexity and accuracy when using the initial RB obtained with the WM method. Results in this table for MSE must be multiplied by 10^{-6}, 10^{8}, 10^{-4}, 10^{2}, 10^{4},10^{-6}, 10^{-8} and 10^{-4} in the case of DEL, CAL, KIN, FOR, BAS, ELV, AIL or TIC respectively.

Datasets	ACO_{HDP}			ACO_{DUB}			ACO_{DOM}			ACO_{FRA}		
	#R	MSE_{TRA}	MSE_{TST}	#R	MSE_{TRA}	MSE_{TST}	#R	MSE_{TRA}	MSE_{TST}	#R	MSE_{TRA}	MSE_{TST}
DEL	256.13	1.0423	1.2177	294.10	1.0445	1.2223	242.80	1.0873	1.2405	240.93	1.1158	1.3514
ABA	75.07	2.3757	2.5950	82.57	2.3744	2.6052	73.63	2.4892	2.6621	79.13	2.3462	2.6505
CAL	215.13	21.0229	21.9481	239.70	21.0328	21.9599	172.70	24.6511	25.3395	232.47	20.6081	22.1077
CON	156.47	15.2526	31.2032	179.00	15.0353	31.1829	164.63	18.9389	35.0770	174.20	15.8395	36.5558
KIN	4968.40	30.5995	108.0379	5197.27	26.9888	109.2050	5057.03	33.1820	92.8559	5865.83	36.8488	114.9363
PUM	5050.47	2.0069	7.6021	5291.20	1.4989	8.6435	5088.40	2.4072	7.6352	5406.47	1.4594	8.9901
STP	95.77	0.4777	0.6004	113.13	0.4693	0.6003	89.40	0.6133	0.7815	94.83	0.5999	0.8328
WAN	212.14	1.6969	4.4078	230.87	1.6659	4.4439	188.07	2.3553	4.6864	204.60	2.4138	5.6817
WINR	195.33	0.1355	0.2189	219.83	0.1355	0.2183	163.40	0.1591	0.2241	244.57	0.1165	0.2358
WINW	334.27	0.2098	0.2570	367.90	0.2076	0.2567	261.67	0.2349	0.2673	428.27	0.1875	0.2638
FOR	179.63	9.6025	284.2226	179.10	9.5502	289.0661	174.27	9.6009	305.6473	168.47	9.3414	308.6922
MOR	75.73	0.0427	0.0586	81.90	0.0427	0.0593	68.37	0.0469	0.0630	53.43	0.1163	0.1203
BAS	129.47	4.9020	57.7261	136.23	4.8270	61.4822	127.70	5.4407	57.5143	132.33	5.2430	63.1273
ELV	2676.47	5.1144	6.6920	-	-	-	-	-	-	-	-	-
CA	707.63	3.1529	9.4692	-	-	-	-	-	-	-	-	-
AIL	4723.23	1.2918	2.1573	-	-	-	-	-	-	-	-	-
TIC	6314.40	69.8025	898.5079	-	-	-	-	-	-	-	-	-

Table 7

Average results of the four different adaptive conjunction FRBS models for the most accurate point regarding complexity and accuracy when using the initial RB obtained with the FS-MOGUL method. Results in this table for MSE must be multiplied by 10^{-6}, 10^{8}, 10^{-4}, 10^{2}, 10^{4},10^{-6}, 10^{-8} and 10^{-4} in the case of DEL, CAL, KIN, FOR, BAS, ELV, AIL or TIC respectively.

Datasets	ACO_{HDP}			ACO_{DUB}			ACO_{DOM}			ACO_{FRA}		
	#R	MSE_{TRA}	MSE_{TST}	#R	MSE_{TRA}	MSE_{TST}	#R	MSE_{TRA}	MSE_{TST}	#R	MSE_{TRA}	MSE_{TST}
DEL	228.30	1.0833	1.2311	260.83	1.0894	1.2377	228.17	1.1196	1.2437	181.13	1.1460	1.3214
ABA	74.50	2.3844	2.5992	80.73	2.3832	2.5743	68.93	2.4348	2.5903	72.07	2.3890	2.6604
CAL	229.20	24.0739	25.0128	248.87	24.0855	25.0122	200.90	26.9570	27.7096	242.57	25.0647	26.1682
CON	223.03	15.4176	32.4437	258.63	15.9592	32.9127	238.50	19.5425	36.2815	232.33	15.5709	33.8978
KIN	7767.70	52.6087	147.8134	8257.33	52.1064	148.2074	8247.87	55.2433	140.9773	8902.63	65.7062	163.4085
PUM	7654.63	3.8650	10.8437	8263.17	3.7888	10.8523	8264.67	4.1565	10.9331	8358.67	4.0480	11.8748
STP	60.80	0.8821	1.0596	62.90	0.8914	1.0688	55.77	1.0135	1.2139	49.50	1.0640	1.3996
WAN	111.70	3.3672	6.5059	124.53	3.5356	6.7360	104.73	3.7667	7.0111	89.03	4.9432	9.1283
WINR	225.40	0.1386	0.2313	261.30	0.1393	0.2325	221.53	0.1627	0.2338	211.30	0.1315	0.2488
WINW	226.70	0.2328	0.2694	269.90	0.2321	0.2687	222.57	0.2479	0.2746	217.47	0.2285	0.2812
FOR	228.67	2.2664	477.7946	211.90	2.2086	487.5386	226.37	3.8820	483.7357	186.27	2.0400	488.6680
MOR	50.20	0.1097	0.1348	47.67	0.1145	0.1395	40.43	0.1019	0.1245	31.20	0.3064	0.3680
BAS	172.60	5.1023	81.8764	184.73	5.1309	81.3165	176.73	5.6304	78.2598	142.20	5.4827	82.8408
ELV	782.77	8.0021	10.6360	-	-	-	-	-	-	-	-	-
CA	325.23	5.9463	12.9131	-	-	-	-	-	-	-	-	-
AIL	1501.57	1.6639	2.7844	-	-	-	-	-	-	-	-	-
TIC	6586.37	464.4229	969.9089	-	-	-	-	-	-	-	-	-

Table 8

Rankings obtained with Friedman's test on MSE_{TST} and #R measures with both RBs for MAX ACC point.

	FRBSs with WM RBs		FRBSs with FS-MOGUL RBs	
Algorithm	Ranking on MSEtst (p-value Fried: 0,000004)	Ranking on #R (p-value Fried: 0,000404)	Ranking on MSEtst (p-value Fried: 0,000014)	Ranking on #R (p-value Fried: 0,004474)
ACO_{HDP}	1.3529	1.9412	1.4706	2.2353
ACO_{DUB}	2.1765	3.4706	2.2353	3.4412
ACO_{DOM}	2.8824	1.8235	2.6471	2.3824
ACO_{FRA}	3.5882	2.7647	3.6471	1.9412

Table 9

Finner's test table with p-value= 0.05for the methods on MSE_{TST} and #R on the MAX ACC point when using WM RBs control: ACO_{HDP} for MSE_{TST}, ACO_{DOM} for #R.

WM RBs							
MSEtst				#R			
i	Algorithm	pFinner	Hypot	i	Algorithm	p-Finner	Hypot
1	ACO_{DUB}	0.062915	Accept.	1	ACO_{HDP}	0.790482	Accept.
2	ACO_{DOM}	0.000829	Reject.	2	ACO_{FRA}	0.049896	Reject.
3	ACO_{FRA}	0.000001	Reject.	3	ACO_{DUB}	0.000599	Reject.

Table 10
Finner's test table with p-value= 0.05 for the methods on MSE_{TST} and #R on the MAX ACC point when using FS-MOGUL RBs control: ACO_{HDP} for MSE_{TST}, ACO_{FRA} for #R.

FS-MOGUL RBs							
MSEtst				#R			
I	Algorithm	p-Finner	Hypot	i	Algorithm	p-Finner	Hypot
1	ACO_{DUB}	0.084177	Accept.	1	ACO_{HDP}	0.506555	Accept.
2	ACO_{DOM}	0.011808	Reject.	2	ACO_{DOM}	0.438139	Accept.
3	ACO_{FRA}	0.000003	Reject.	3	ACO_{DUB}	0.002115	Reject.

On the other hand, for #R the control algorithms are ACO_{DOM} and ACO_{FRA} when using WM and FS-MOGUL RBs respectively. In these cases Finner's test does not reject (accept) the hypothesis when it is compared with ACO_{HDP}. Indeed, in these cases there are not significant differences among the results and the #R could be considered similar.

Therefore, analysing the results from Tables 6 and 7, and the statistical evidence obtained on Tables 8, 9 and 10, we can highlight that in the MAX-ACC point:

- ACO_{HDP} achieves the best accuracy, with ACO_{DUB}being very close to it with no statistical significance. However, the important issue is that the proposed new approach shows much simpler models than ACO_{DUB}.
- In the case of #R there are not significant differences between our approach and the best (ACO_{FRA}), but our proposal achieves much higher accuracy.

4.3. Study of computational times and scalability

In this section, we analyse how the use of ACO_{HDP} considerably reduce the execution times compared to models with the classical adaptive conjunction operators. This times were obtained with a computer based on an Intel® Core™ 2 Quad Q9650 (12M cache, 3.00 GHz, 1333 MHz FSB) processor with 4 GB of memory and a Linux 64-bit operating system by using only one of the four cores.

Tables 11 and 12 show the running times of the studied methods for each of 30 runs for those models obtained with WM and FS-MOGUL RBs, respectively. Furthermore, we also show the time gain between our approach and the different models of comparison. We emphasise that these times were obtained without considering the time required to obtain the initial RB. Except for the four most complex datasets, the proposed method is between 6.0 and 18.8 times quicker. ACO_{HDP} allows a considerable amount of time saving. Focusing on the four most complex problems, ACO_{HDP} needs considerable time (particularly when using WM RBs which has a notably higher number of rules than FS-MOGUL ones), but obtains models that is impossible to reach without it.Thus, we can conclude that:

- The advantage of ACO_{HDP} regarding time gained is very significant.
- The influence of the number of rules and examples is very important. We can see significant differences between the results obtained when using the WM and FS-MOGUL RBs, due to the difference between the #R they produce initially.
- The time gain obtained by ACO_{HDP} depends on the combination of several factors: dataset complexity, the aforementioned initial #R, the number of instances in the training data set, the number of rules triggered per instance (see the website accompanying the paper http://www.uhu.es/gisimd/papers/ACO-HDP), and the possibility of reducing the #R during execution (we think that when the rule selection has difficulties to reduce the initial set of rules, ACO_{HDP} gains more time over the other models because the difficulty remains higher, which is particularly problematical for the traditional adaptive conjunctions models), etc.

Table 11
Average time of a run — hours, minutes and seconds (H:M:S) when using WM RBs.

Datasets	ACO_{HDP}	ACO_{DUB}	Time Gain	ACO_{DOM}	Time Gain	ACO_{FRA}	Time Gain
DEL	03:03:42	25:40:36	8.4	93:48:13	30.6	102:26:21	33.5
ABA	00:29:06	02:59:07	6.2	10:38:39	21.9	12:01:17	24.8
CAL	07:13:20	56:13:34	7.8	246:52:09	34.2	268:52:28	37.2
CON	00:07:17	00:56:59	7.8	02:26:26	20.1	03:11:06	26.2
KIN	12:19:00	131:40:32	10.7	331:41:08	26.9	361:13:10	29.3
PUM	11:35:49	135:11:52	11.7	334:51:08	28.9	371:43:43	32.1
STP	00:07:45	01:03:42	8.2	04:30:02	34.8	04:44:10	36.7
WAN	00:21:45	03:36:42	10.0	15:39:19	43.2	18:11:26	50.2
WINR	00:42:02	08:21:49	11.9	43:54:40	62.7	46:50:19	66.9
WINW	03:35:14	57:59:37	16.2	339:36:34	94.7	363:27:32	101.3
FOR	00:04:08	00:36:32	8.8	01:58:10	28.6	02:15:28	32.7
MOR	00:07:24	01:26:36	11.7	07:55:42	64.3	08:16:32	67.2
BAS	00:02:02	00:35:46	17.6	03:23:03	99.9	03:44:45	110.6
ELV	28:17:34	-	-	-	-	-	-
CA	07:34:59	-	-	-	-	-	-
AIL	39:20:49	-	-	-	-	-	-
TIC	15:27:27	-	-	-	-	-	-

Table 12
Average time of a run — hours, minutes and seconds (H:M:S) when using FS-MOGUL RBs.

Datasets	ACO_{HDP}	ACO_{DUB}	Time Gain	ACO_{DOM}	Time Gain	ACO_{FRA}	Time Gain
DEL	03:05:42	22:56:01	7.4	84:49:42	27.4	94:44:31	30.6
ABA	00:29:27	02:56:51	6.0	11:31:26	23.5	12:35:10	25.6
CAL	05:36:19	40:32:53	7.2	150:35:13	26.9	164:10:55	29.3
CON	00:14:12	01:50:15	7.8	04:38:20	19.6	06:11:42	26.2
KIN	18:14:59	211:19:54	11.6	480:14:50	26.3	534:27:57	29.3
PUM	18:23:55	197:17:38	10.7	471:32:51	25.6	515:19:24	28.0
STP	00:04:43	00:37:04	7.9	02:44:02	34.8	02:51:08	36.3
WAN	00:14:38	02:17:53	9.4	10:13:59	41.9	11:44:54	48.1
WINR	00:30:33	06:21:23	12.5	32:16:38	63.4	34:17:32	67.4
WINW	01:40:05	27:33:21	16.5	157:50:35	94.6	172:14:05	103.3
FOR	00:08:06	01:09:55	8.6	03:47:30	28.1	04:42:53	35.0
MOR	00:03:35	00:36:52	10.3	03:34:38	59.8	03:36:07	60.2
BAS	00:03:33	01:06:46	18.8	06:18:47	106.7	06:58:51	118.0
ELV	09:12:11	-	-	-	-	-	-
CA	03:13:09	-	-	-	-	-	-
AIL	12:11:09	-	-	-	-	-	-
TIC	20:26:54	-	-	-	-	-	-

4.4. *Analysis of the Pareto fronts*

In this section, we analyse the effectiveness of the proposed algorithm in the remaining solutions that the multi-objective algorithm obtains in the Pareto fronts. In order to study these Pareto fronts, we compare each of the thirty Pareto fronts obtained for each approach using the CS ratio. We used it in [43] due to it provide more reliable information than other classical criteria, also used by us in [42] and other authors, where the averages for point of MAX INT, MEDIUM ACC/INT and MAX ACC are compared. Nevertheless, we have included an Excel format file with these representative points for all models so that interested readers can access this information in the website accompanying the paper (http://www.uhu.es/gisimd/papers/ACO-HDP).

To perform the CS ratio, we applied the following criterion: we use all Pareto fronts of each one of the FRBSs models for comparison (those with ACO_{HDP}, ACO_{DUB}, ACO_{DOM} and ACO_{FRA}), i.e. thirty Pareto fronts for each one. Thus, we perform the CS ratio between our approach, ACO_{HDP}, and each of the other approaches in each one of the thirty Pareto fronts. Then, we show the average of thirty CS ratios. Thus, we can analyse the differences between the diverse solutions obtained by the ACO_{HDP} against ACO_{DUB}, ACO_{DOM} and ACO_{FRA} FRBSs models in terms of the Pareto fronts.

Tables 13 and 14 show these CS ratios. We compare the different approaches: our approach against the classic adaptive conjunction operators. To highlight the results with the best behaviour, they are highlighted in bold. Note also that in these two tables we have employed only the first 13 datasets, as no results were obtained from the classic adaptive conjunction operators FRBSs models in the remaining four.

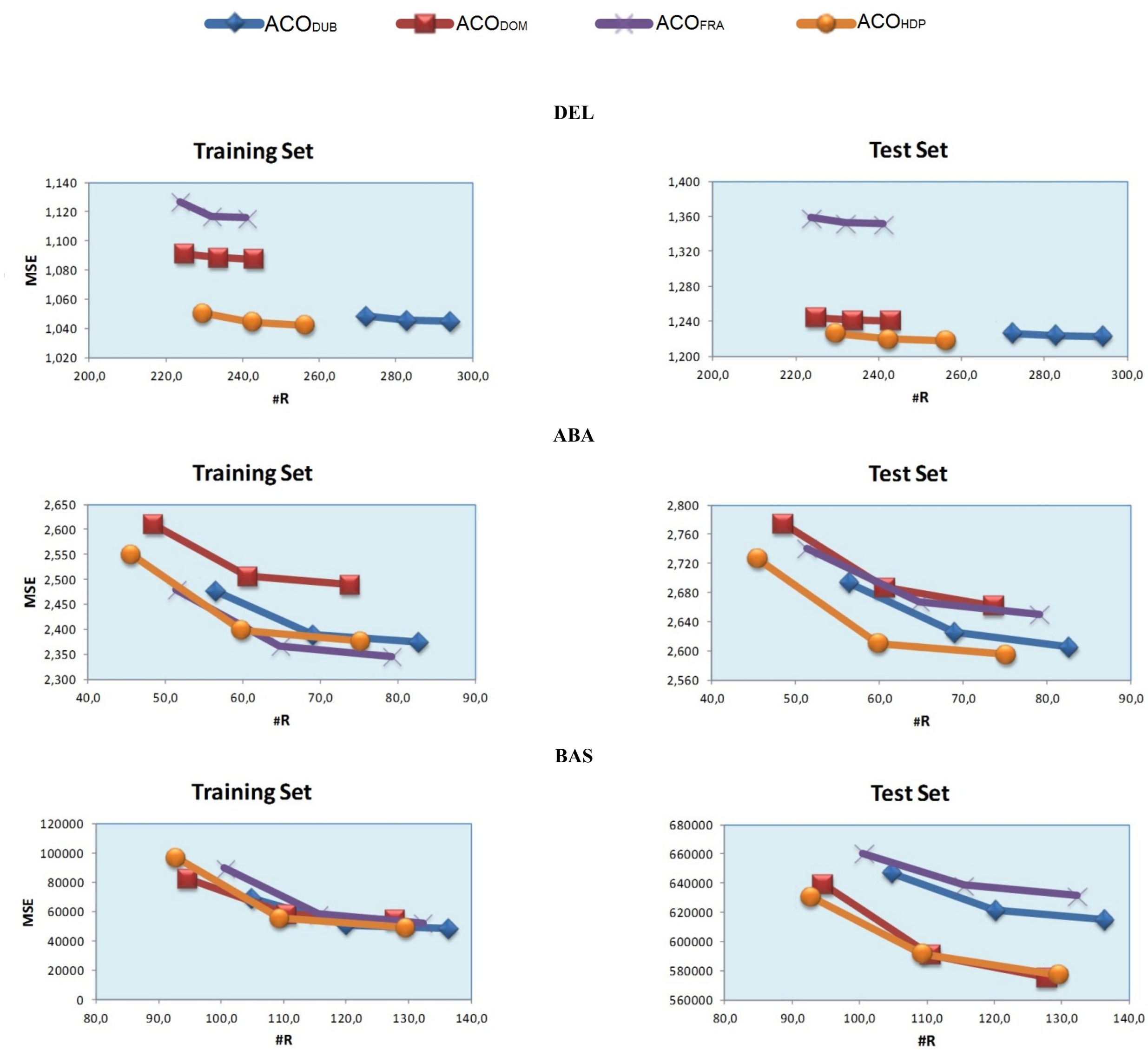
The results show that ACO_{HDP} achieves a higher CS ratio than the model based on the adaptive classic t-norms for almost all problems and regardless of the RB employed.

Table 13
CS ratios obtained by all Pareto fronts when using WM RBs

Datasets	ACO_{HDP} vs. ACO_{DUB}	ACO_{DUB} vs. ACO_{HDP}	ACO_{HDP}vs.ACO_{DOM}	ACO_{DOM}vs.ACO_{HDP}	ACO_{HDP}vs.ACO_{FRA}	ACO_{FRA}vs.ACO_{HDP}
DEL	**0.719**	0.000	**0.599**	0.123	**0.630**	0.000
ABA	**0.777**	0.229	**0.874**	0.114	**0.849**	0.155
CAL	**0.657**	0.005	**0.050**	0.000	**0.690**	0.026
CON	**0.677**	0.145	**0.873**	0.116	**0.936**	0.048
KIN	**0.693**	0.000	**0.000**	0.000	**0.971**	0.000
PUM	**0.861**	0.000	**0.700**	0.000	**1.000**	0.000
STP	**0.587**	0.112	**0.806**	0.029	**0.899**	0.027
WAN	**0.500**	0.024	**0.313**	0.165	**0.612**	0.068
WINR	**0.474**	0.062	0.008	**0.122**	**0.800**	0.000
WINW	**0.565**	0.000	**0.000**	**0.000**	**0.833**	0.000
FOR	**0.547**	0.273	**0.550**	0.342	**0.554**	0.302
MOR	**0.580**	0.311	**0.491**	0.457	**0.266**	0.000
BAS	**0.682**	0.224	**0.538**	0.432	**0.715**	0.254

Table 14
CS ratios obtained by all Pareto fronts when using FS-MOGUL RBs

Datasets	ACO_{HDP} vs. ACO_{DUB}	ACO_{DUB} vs. ACO_{HDP}	ACO_{HDP}vs.ACO_{DOM}	ACO_{DOM}vs.ACO_{HDP}	ACO_{HDP}vs.ACO_{FRA}	ACO_{FRA}vs.ACO_{HDP}
DEL	**0.794**	0.008	**0.790**	0.128	**0.031**	0.000
ABA	**0.586**	0.380	0.361	**0.629**	**0.770**	0.229
CAL	**0.540**	0.087	**0.157**	0.000	**0.291**	0.021
CON	**0.739**	0.000	**0.989**	0.004	**0.755**	0.151
KIN	**0.533**	0.000	**0.000**	0.000	**0.994**	0.000
PUM	**0.586**	0.000	**0.560**	0.000	**1.000**	0.000
STP	**0.652**	0.337	**0.814**	0.138	**0.754**	0.075
WAN	**0.687**	0.178	**0.552**	0.331	**0.262**	0.057
WINR	**0.640**	0.006	**0.387**	0.290	**0.465**	0.045
WINW	**0.515**	0.000	**0.739**	0.171	**0.781**	0.006
FOR	**0.071**	0.067	**0.391**	0.122	0.000	**0.098**
MOR	**0.827**	0.194	0.155	**0.833**	**0.666**	0.008
BAS	**0.451**	0.272	**0.423**	0.326	0.004	**0.593**



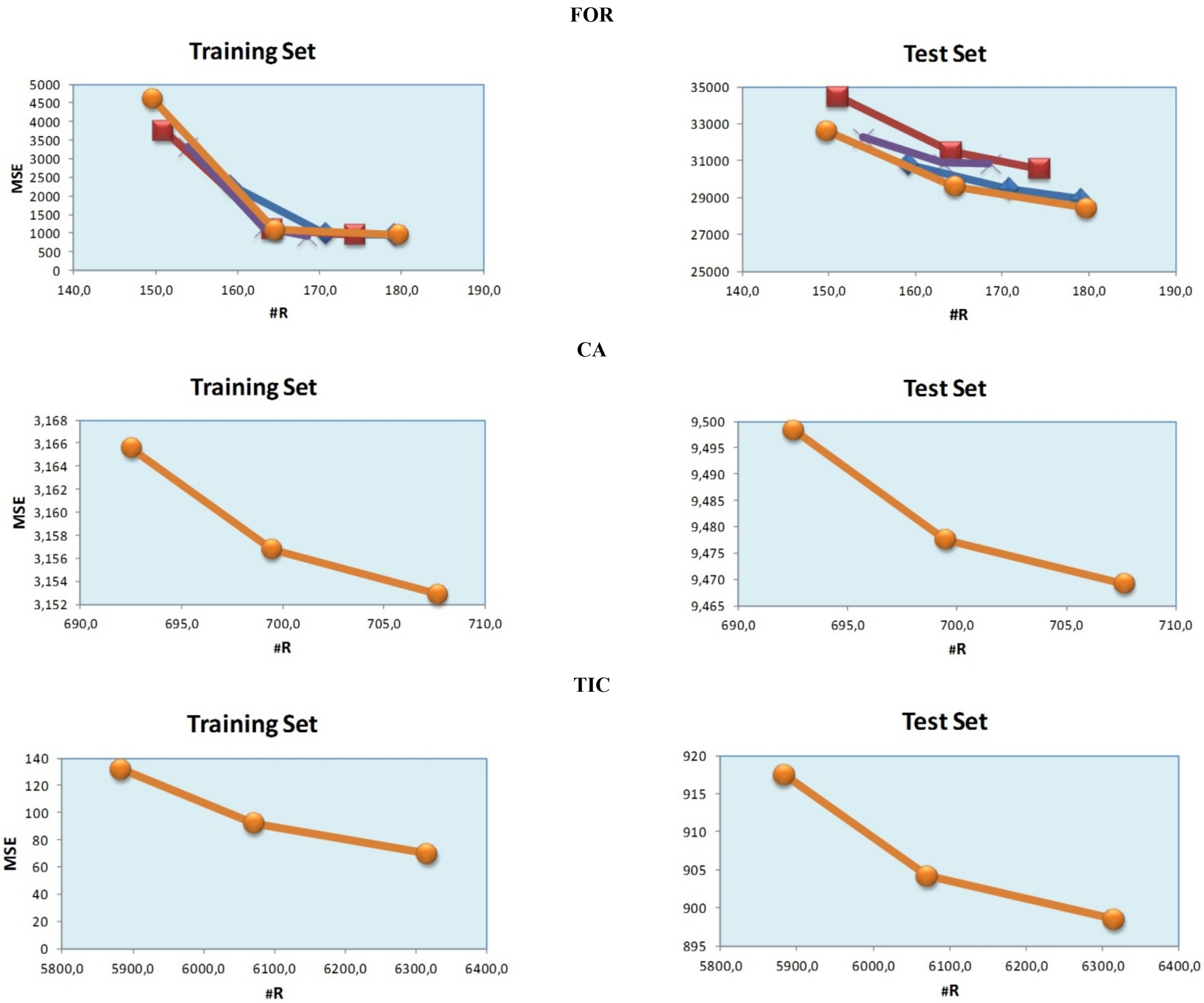


Fig. 2. Average Pareto fronts obtained by ACO_{HDP}, ACO_{DUB}, ACO_{DOM}, ACO_{FRA} for WM RB with the different datasets.

The average Pareto fronts obtained in a representative set of the studied datasets are shown in Fig. 2. This figure also includes the average solutions obtained for the classic adaptive conjunction operators approaches. To plot these figures, we used the three representative points described at the beginning of this section (MAX INT, MEDIUM ACC/INT, and MAX ACC), then connected these points to show the hypothetical Pareto fronts. We can see that in most of the datasets, many solutions present better results in test than those obtained by the other approaches (classic adaptive t-norms).

The results for the remaining datasets plots of the average Pareto fronts together with the global result Excel format files can be found on the paper's companion website at (http://www.uhu.es/gisimd/papers/ACO-HDP).

5. Conclusions

In this paper, we have proposed an efficient Multi-objective evolutionary Adaptive Inference Linguistic Fuzzy System for high-dimensional regression problems. Our proposal, based on an adaptive conjunction operator we name ACO_{HDP}, addresses high dimensional problems, which benefit from the adaptive conjunction operator proposed and also from the use of an effective implementation to reduce the necessary time spent tuning the parameters of the adaptive operator against three of the most classical ones used in the literature.

An in-depth experimental study carried out with 17 datasets of different complexities, using two different RBs and performing statistical analyses, confirms the effectiveness of the mechanism proposed, particularly in the following points:

- ACO_{HDP} improves in most of cases or at least equals the accuracy, maintaining a similar model complexity over the models that use the classic adaptive conjunctions.
- It is considerably more efficient so saves a lot of time, and can be applied to datasets impossible to model with the other adaptive inference operators.

To sum up, the proposed methodology based on the local adaptation of the Inference System of linguistic fuzzy models, which is not designed to compete with other mechanisms to improve the quality of linguistic fuzzy modelling (i.e. those focused on improving the Knowledge Base) since it is complementary, is particularly suitable for use when dealing with large and complex datasets.

ACKNOWLEDGMENT

Paper partially supported by the Spanish Ministry of Economy and Competitiveness under grant no. TIN2008-06681-C06-06 and ERDF Funds.

REFERENCES

[1] R. Alcalá, P. Ducange, F. Herrera, B. Lazzerini, F. Marcelloni, A multi-objective evolutionary approach to concurrently learn rule and data bases of linguistic fuzzy rule-based systems, IEEE Transactions on Fuzzy System 17 (5) (2009) 1106–1122.

[2] R. Alcalá, M. J. Gacto, F. Herrera, J. Alcalá-Fdez, A multi-objective genetic algorithm for tuning and rule selection to obtain accurate and compact linguistic fuzzy rule-based systems, International Journal of Uncertainty, Fuzziness Knowledge-Based System 15 (5) (2007) 539–557.

[3] R. Alcalá, M.J. Gacto, F. Herrera, A fast and scalable multiobjective genetic fuzzy system for linguistic fuzzy modeling in high-dimensional regression problems, IEEE Transaction on Fuzzy System 19 (4) (2011) 666–681.

[4] R. Alcalá, Y. Nojima, F. Herrera, H. Ishibuchi, Multiobjective genetic fuzzy rule selection of single granularity-based fuzzy classification rules and its interaction with the lateral tuning of membership functions, Soft Computing, Special Issue on Evolutionary fuzzy Systems, 15 (12) (2011) 2303–2318.

[5] J. Alcala-Fdez, F. Herrera, F.A. Márquez, A. Peregrín, Increasing fuzzy rules cooperation based on evolutionary adaptive inference systems, International Journal of Intelligent System 22 (9) (2007) 1035–1064.

[6] J. Alcala-Fdez, L. Sánchez, S. García, M. del Jesus, S. Ventura, J. Garrell, J. Otero, C. Romero, J. Bacardit, V. Rivas, J. Fernández, F. Herrera, Keel: a software tool to assess evolutionary algorithms for data mining problems, Soft Computing 13 (3) (2009) 307–318.

[7] M. Antonelli, P. Ducange, B. Lazzerini, F. Marcelloni, Learning concurrently partition granularities and rule bases of mamdani fuzzy systems in a multi-objective evolutionary framework, International Journal of Approximate Reasoning 50 (7) (2009) 1066–1080.

[8] M. Antonelli, P. Ducange, F. Marcelloni, Genetic training instance selection in multi-Objective evolutionary fuzzy system: A co-evolutionary approach, IEEE Transactions on Fuzzy System 20(2) (2012) 276–290.

[9] J. M. Bardallo, M. A. De Vega, F. A. Márquez, A. Peregrín, Parallel evolutionary multiobjective methodology for granularity and rule base learning in linguistic fuzzy Systems, in: Proceedings of the FUZZ-IEEE'09 IEEE International Conference on Fuzzy System, Jeju Island, Korea, 2009, pp. 1700–1705.

[10] I. Z. Batyrshin, I. J. Rudas, L. A. Villa, P. Cortés, On the monotone sum of basic t-norms in the construction of parametric families of digital conjunctors for fuzzy systems with reconfigurable logic, Knowledge-Based Systems, Special Issue on Advances in Fuzzy Knowledge Systems: Theory and Application, 38 (2013) 27–36.

[11] A. Botta, B. Lazzerini, F. Marcelloni, D. C. Stefanescu, Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index, Soft Computing 13 (5) (2008) 437–449.

[12] M. Cococcioni, B. Lazzerini, F. Marcelloni, Toward efficient multi-objective genetic Takagi-Sugeno fuzzy systems for high dimensional problems, Computational Intelligence, in Expensive Optimization Problems, Springer Series Studies in Evolutionary Learning and Optimization (ELO), Springer-Verlag , 2010, pp. 397–422.

[13] M. Cococcioni, B. Lazzerini, F. Marcelloni, On reducing computational overead in multi-objective genetic Takagi-Sugeno fuzzy systems, Applied Soft Computing 11 (1) (2011) 675–688.

[14] W. E. Combs, J. E. Andrews, Combinatorial rule explosion eliminated by a fuzzy rule configuration, IEEE Transactions on Fuzzy System, 6 (1) (2008) 1–11.

[15] O. Cordón, F. Herrera, A three-stage evolutionary process for learning descriptive and approximate fuzzy logic controller knowledge bases from examples, International Journal of Approximate Reasoning 17 (4) (1997) 369–407.

[16] O. Cordón, F. Herrera, A. Peregrín, Applicability of the fuzzy operators in the design of fuzzy logic controllers, Fuzzy Sets and Systems86 (1997) 15–41.

[17] O. Cordón, F. Herrera, A. Peregrín, Searching for basic properties obtaining robust implication operators in fuzzy control, Fuzzy Sets and Systems 111 (2000) 237–251.

[18] O. Cordon, A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems, International Journal of Approximate Reasoning 52 (2011) 894–913.

[19] M. A. De Vega, J. M. Bardallo, F. A. Márquez, A. Peregrín, Parallel distributed two-level evolutionary multiobjective methodology for granularity learning and membership functions tuning in linguistic fuzzy systems, in: Proceedings of the ISDA'09 International Conference on Intelligent System Design and Applications, Pisa, Italy, 2009, pp. 134–139.

[20] K. Deb, A. Pratap, S. Agarwal T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Transaction on Evolutionary Computation 6 (2) (2002) 182–197.

[21] J. Demšar, Statistical comparisons of classifiers over multiple data sets, Journal of Machine Learning Research 7 (2006) 1–30.

[22] L.J. Eshelman, The CHC adaptive search algorithm: how to safe search when engaging in nontraditional genetic recombination, in: G.J.E. Rawlings (Ed.), Foundations of genetic algorithms, 1991, pp. 265–283.

[23] M. Fazzolari, R. Alcalá, Y. Nojima, H. Ishibuchi, F. Herrera, A review of the application of Multi-Objective Evolutionary Systems: Current status and further directions, IEEE Transactions on Fuzzy Systems 21(1) (2013) 45–65.

[24] H. Finner, On a monotonicity problem in step-down multiple test procedures, Journal of the American Statistical Association 88 (1993) 920–923.

[25] M. Friedman, The use of ranks to avoid the assumption of normality implicit in the analysis of variance, Journal of the American Statistical Association 32 (1937) 674–701.

[26] M. J. Gacto, R. Alcalá, F. Herrera, Adaptation and application of multi-objective evolutionary algorithms for rule reduction and parameter tuning of fuzzy rule-based systems, Soft Computing 13 (5) (2009) 419–436.

[27] M. J. Gacto, R. Alcalá, 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 Transaction on Fuzzy System 18 (3) (2010) 515–531.

[28] S. García, A. Fernández , J. Luengo, F. Herrera, A study of statistical techniques and performance measures for genetics-based machine learning: Accuracy and interpretability, Soft Computing 13(10) (2009) 959–977.

[29] S. García, A. Fernández, J. Luengo, F. Herrera, Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: experimental analysis of power, Information Sciences 180 (10) (2010) 2044–2064.

[30] S. García, F. Herrera, An extension on statistical comparisons of classifiers over multiple data sets for all pairwise comparisons, Journal of Machine Learning Research 9 (2008) 2579–2596.

[31] M. Gupta, J. Qi, Theory of T-norms and fuzzy inference methods,Fuzzy Sets and Systems 40 (1991) 431–450.

[32] E. Hadavandi, H. Shavandi, A. Ghanbari, Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting, Knowledge-Based Systems, 23 (8) (2010) 800–808.

[33] F. Herrera. Genetic fuzzy systems: Taxonomy, current research trends and prospects. Evolutionary Intelligence 1 (2008) 27–46.

[34] F. Herrera, M. Lozano, A. Sánchez, A taxonomy for the crossover operator for real-coded genetic algorithms: An experimental study, Internation Journal of Intelligent System 18 (2003) 309–338.

[35] H. Ishibuchi, Y. Nojima, Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning, International Journal of Approximation Reasoning 44 (1) (2007) 4–31.

[36] H. Ishibuchi, Y. Nakashima, Y. Nojima, Performance evaluation of evolutionary multiobjective optimization algorithms for multiobjective fuzzy genetics-based machine learning, Soft Computing, Special Issue on Evolutionary Systems, 15 (12) (2011) 2415–2434

[37] H. Ishibuchi, S. Mihara, Y. Nojima, Parallel Distributed Hybrid Fuzzy GBML models with rule set migration and training data rotation, IEEE Transactions on Fuzzy Systems, 2013, in press.

[38] Y. Jin, Fuzzy modeling of high-dimensional systems: Complexity reduction and interpretability improvement, IEEE Transactions on Fuzzy System 8 (2) (2000) 212–221.

[39] Y. Jin, A comprehensive survey of fitness approximation in evolutionary computation, Soft Computing 9 (2005) 3–12.

[40] A.A. Márquez, F.A. Márquez, A. Peregrín, Cooperation between the inference system and the rule base by using multiobjective genetic algorithms, in: Proceedings of the International Conference of Hybrid Artificial Intelligence Systems, 2008, Burgos, Spain, pp. 739–746.

[41] A.A. Márquez, F.A. Márquez A. Peregrín, Rule base and inference system cooperative learning of mamdani fuzzy systems with multiobjective genetic algorithms, in: Proceedings of the IFSA World Congress (IFSA-EUSFLAT'09), Lisbon, Portugal, 2009, pp. 1045-1050.

[42] A.A. Márquez, F.A. Márquez, A. Peregríın, Rule base and adaptive fuzzy operators cooperative learning of mamdani fuzzy systems with multi-objective genetic algorithms, Evolutionary Intelligence, Special Issue on GFS: New Advances 2 (1-2) (2009) 39–51.

[43] A.A. Márquez, F.A. Márquez, A. Peregrín, A Mechanism to improve the interpretability of linguistic fuzzy systems with adaptive defuzzification based on the use of a multi-objective evolutionary algorithm, International Journal of Computational Intelligent Systems, Special Issue on Evolutionary Fuzzy Systems 5 (2) (2012) pp. 297–321.

[44] F.A. Márquez, A. Peregrín, F. Herrera, Cooperative evolutionary learning of fuzzy rules and parametric aggregation connectors for Mamdani linguistic fuzzy systems, IEEE Transactions on Fuzzy System 15 (6) (2007) 1162–1178.

[45] M. Mizumoto, Pictorial representations of fuzzy connectives, part I:Cases of T-norms, T-conorms and averaging operators, Fuzzy Sets and Systems 31 (1989) 217–242.

[46] P. Pulkkinen, J. Hytönen, H. Koivisto, Developing a bioaerosol detector using hybrid genetic fuzzy systems, Engineering Applications of Artificial Intelligence 21 (8) (2008) 1330–1346.

[47] I. Robles, R. Alcalá, J.M. Benítez, F. Herrera, Evolutionary parallel and gradually distributed lateral tuning of fuzzy rule-based systems, Evolutionary Intelligence, Special Issue on GFS: New Advances 2 (2009) 5–19.

[48] I. J. Rudas, E. Pap, J. Fodor, Information aggregation in intelligent systems: An application oriented approach, Knowledge-Based Systems, Special Issue on Advances in Fuzzy Knowledge Systems: Theory and Application, 38 (2013) 3–13.

[49] D. G. Stavrakoudis, G. N. Galidaki, I. Z. Gitas, J. B. Theocharis, Reducing the complexity of genetic fuzzy classifiers in highly-dimensional classification problems, International Journal of Computational Intelligence Systems, Special Issue on Evolutionary Fuzzy Systems 5 (2) (2012) 254–275.

[50] E. Trillas, L. Valverde, On implication and indistiguishability inthe setting of fuzzy logic, in Management Decision Support SystemsUsing Fuzzy Sets and Possibility Theory, J. Kacpryzk and R. Yager,Eds. New York: VerlagTüVRheinland, 1985, pp. 198–212.

[51] L. Wang, J. Mendel, Generating fuzzy rules by learning from examples, IEEE Transaction on Systems, Man and Cybernetics 22 (6) (1992) 1414–1427.

[52] H.Wu, J. M. Mendel, On choosing models for linguistic connectorwords for mamdani fuzzy logic systems, IEEE Transactions on Fuzzy Systems 4 (2004) 29–44.

[53] E. Zitzler, K. Deb, L. Thiele, Comparison of multiobjective evolutionary algorithms: Empirical results, Evolutionary Computation 8 (2) (2000) 173–195.