2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Embedding Evolutionary Multiobjective Optimization into Fuzzy Linguistic Combination Method for Fuzzy Rule-Based Classifier Ensembles

Krzysztof Trawiński, Oscar Cordón and Arnaud Quirin

***Abstract*— In a preceding contribution, we proposed a novel combination method by means of a fuzzy linguistic rule-based classification system. The fuzzy linguistic combination method was based on a genetic fuzzy system in order to learn its parameters from data. By doing so the resulting classifier ensemble was able to show a hierarchical structure and the operation of the latter component was transparent to the user. In addition, for the specific case of fuzzy classifier ensembles, the new approach allowed fuzzy classifiers to deal with high dimensional classification problems avoiding the curse of dimensionality. However, this approach strongly depended on one parameter defining the complexity of the final classifier ensemble and in consequence affecting the final accuracy. To avoid this tedious problem, we propose to automatically derive this parameter. For this purpose, we use the most common evolutionary multiobjective algorithm, namely NSGA-II, in order to optimize two criteria, complexity and accuracy. We carry out comprehensive experiments considering 20 UCI datasets with different dimensionality, showing the good performance of the proposed approach.**

I. Introduction

Classifier ensembles (CEs), also called multiclassifiers, are machine learning tools capable to obtain better performance than a single classifier when dealing with complex classification problems, especially when the number of dimensions or the size of the data are really large [1]. The most common base classifiers are decision trees [2], neural networks [3], and more recently fuzzy classifiers [4], [5], [6], [7].

CE design is essentially based on two stages [8]: the learning of the component classifiers and the design of the combination mechanism for the individual decisions provided by them into the global CE output. The overall accuracy of the CE relies on the performance and the proper integration of these two tasks.

The research area of combination methods is very active. It considers both the direct combination of the results provided by all the initial set of component classifiers to compute the final output (*classifier fusion*) and the selection of the best single classifier or classifier subset which will be taken into account to provide a decision for each specific input pattern (static/dynamic classifier selection [9] and overproduce-and-choose strategies [10]). Besides, hybrid strategies between the two groups have also been introduced [1].

While the weighted majority vote could be considered as the most extended classifier fusion combination method [11], many other proposals have been developed in the specialized literature, including several successful procedures based on the use of fuzzy set theory and, specifically, of fuzzy aggregation operators [12], [13].

In a recent study [14], we introduced a framework to derive a fuzzy rule-based classification system (FRBCS) playing the role of the CE combination method. This fuzzy linguistic combination method presented an interpretable structure as it was based on the use of a single disjunctive fuzzy classification rule per problem class as well as on the classical single-winner fuzzy reasoning method [15], [16]. The antecedent variables corresponded to the component (fuzzy) classifiers (thus its number is bounded by the existing number) and each of them had a weight associated representing the certainty degree of each ensemble member in the classification of each class. A specific genetic algorithm (GA) to design such FRBCS-based combination method (FRBCS-CM) was proposed with the ability of selecting features and linguistic terms in the antecedent parts of the rules. In such way, it performed both classifier fusion and classifier selection at class level.

The resulting system is a genetic fuzzy system (GFS) [17], [18] dealing with the interpretability-accuracy trade-off in a proper way [19]. In that contribution the FRBCS-CM was be applied on fuzzy rule-based classifier ensembles (FRBCEs) generated from the bagging methodology we proposed in [6]. Therefore, the resulting FRBCE showed a clear hierarchical structure composed of two levels of FRBCSs allowing it to deal with high dimensional problems.

However, this approach has one significant drawback. The proposed GFS strongly relies on one parameter (provided by the user *a priori*) defining the desired complexity level for the final CE, which in consequence affects the final accuracy. To avoid this tedious issue, we propose to automatically derive this parameter in the current contribution.

To do so, we use the most common evolutionary multiobjective (EMO) algorithm [20], namely NSGA-II [21], as an engine to design FRBCS-CM. It will simultaneously optimize two different criteria: complexity (eliminating its manual definition *a priori*) and accuracy. In addition, it will generate several FRBCE designs with different accuracy-

Krzysztof Trawiński and Oscar Cordón are with European Centre for Soft Computing, 33600 Mieres, Spain (email: {krzysztof.trawinski, oscar.cordon}@softcomputing.es); Oscar Cordón is also with Department of Computer Science and Artificial Intelligence (DECSAI) and Research Center on Information and Communication Technologies (CITIC-UGR), University of Granada, 18071 Granada, Spain (email: ocordon@decsai.ugr.es); Arnaud Quirin is with Galician Research and Development Center in Advanced Telecommunications (GRADIANT), Communications Area, Edif. CITEXVI, local 14, University of Vigo, 36310 Vigo, Spain (email: aquiring@gradiant.org)

978-1-4799-2072-3/14/$31.00 ©2014 IEEE

complexity trade-offs in a single run as a consequence of the nature of the EMO algorithms. A comprehensive study will be conducted on twenty datasets with different dimensionality from the UCI machine learning repository to test the performance of the derived FRBCEs. We will compare our proposal with the original FRBCE, as well as with the GFS-based FRBCS-CM proposed in [14], showing the performance advantage of our proposal.

This paper is set up as follows. In the next section, the preliminaries required for a good understanding of our work (CE combination methods, fuzzy CE combination methods, and our approach for designing FRBCEs considering bagging) are reviewed. Section III describes the proposed FRBCS-CM framework, structure and the proposal of NSGA-II to design it. The experiments developed and their analysis are shown in Sec. IV. Finally, Sec. V presents some concluding remarks and future research lines.

II. Preliminaries

This section explores the current literature related to CE combination methods and reviews our generation method for FRBCEs.

A. Classifier Ensemble Combination Methods

There are two main approaches in the literature for the combination of the outputs provided by a previously generated set of individual classifiers into a single CE output [22]: *classifier fusion* and *classifier selection*.

Classifier fusion relies on the assumption that all ensemble members make independent errors. Thus, combining the decisions of the ensemble members may lead to increasing the overall performance of the system. Majority voting, sum, product, maximum and minimum are examples of functions used to combine their decisions [23]. The most extended one is the weighted majority voting, which allows to weight the contribution of each individual classifier to the final decision according to its "classification competence" using coefficients of importance [11].

Alternatively, classifier selection is based on the fact that not all the individual classifiers but only a subset of them will influence on the final decision for each input pattern. Two categories of classifier selection techniques exist: static and dynamic [9], [22]. In the first case, regions of competence are defined during the training phase, while in the second case, they are defined during the classification phase taking into account the characteristics of the sample to be classified. There is also another family of static classifier selection methods based on the assumption that the candidate classifiers in the ensemble could be redundant. These methods are grouped under the name of overproduce-and-choose strategy (OCS) [10] and they are based on the fact that a large set of candidate classifiers is generated and then selected to extract the best performing subset (removing duplicates and poor-performing candidate classifiers), which composes the final CE used to classify the whole test set. In addition, hybrid methods between the latter families have been proposed, such as the GA-based dynamic OCS procedure introduced in [24].

The FRBCS-CM used in the current contribution belongs to the static OCS group and it is able to either completely remove a whole candidate classifier or to reduce its contribution to only some specific classes with a specific weight measuring our confidence in the individual classifier for that specific class (as done in other existing classifier selection methods such as [25], [26]). All the latter is performed using a human-interpretable structure generated by means of a EMO-based FRBCS using NSGA-II.

B. Classifier Ensemble Fuzzy Combination Methods

Fuzzy set theory has been extensively and successfully considered for classifier fusion. The use of fuzzy connectives to combine the outputs of the component classifiers of an ensemble was first proposed in [27]. Since then, many different fuzzy aggregation operators have been considered in the specialized literature [12], [13], [28]. In [13] the accuracy of some of them was compared to that of seven of the usual crisp (i.e., non-fuzzy) aggregation operators when considered as combination operators for Boosting classifier ensembles. The conclusions drawn from that experimentation were that fuzzy combination methods clearly outperformed non-fuzzy ones.

Besides, some other works have extended the classifier fusion scope and have proposed some techniques which show some similarities with our proposal. On the one hand, Bulacio et al. [29] introduced a hybrid classifier selection-fusion strategy, considering Sugeno's fuzzy integral as combination method and a greedy heuristic for the ensemble member selection. On the other hand, Lu and Yamaoka [30] introduced a fuzzy combination method specifically designed for a hybrid ensemble of three classifiers which showed the novel characteristic of allowing the user to incorporate human expert knowledge on the bias of the component classifiers. This is done by means of an additional refinement module based on a FRBS comprised by Mamdani-type fuzzy rules. Lu and Yamaoka's fuzzy combination method makes use of a fuzzy reasoning process where the following components are considered: a linguistic partition for the ensemble members' outputs, a fuzzy aggregation of their membership degrees and a defuzzification method to modify them, and a new (crisp) aggregation for each class in order to take the final CE decision corresponding to the largest aggregated class membership value.

As said, the latter procedure can be complemented by expert-defined fuzzy rules to adjust the importance of the decisions taken for each class according to the nature of the component classifiers. Hence, the FRBS is used as a refinement module for the fuzzy combination method decisions. Nevertheless, this strategy shows several problems such as its specificity to the consideration of a simple three-classifier ensemble, its highly complex structure composed of two different nature fuzzy reasoning modules, the need of manually defining the fuzzy rules in the refinement module

[1] and the impossibility to perform classifier selection (which of course is not required in the simple ensemble structure considered).

C. Bagging Fuzzy Classifier Ensembles

We take the methodology for the base fuzzy classifier generation that we described in [6] as a base. We incorporated FURIA [31], [32] into a CE framework based on classical CE design approaches [2], [33], [34] in order to generate FRBCEs. We concluded that pure bagging without additional feature selection obtained the best performance when combined with FURIA-based fuzzy classifiers. Thus, we consider the use of bagging with the entire feature set to generate the initial FURIA-based FRBCEs.

In order to build these FRBCEs, a normalized dataset is split into two parts, a training set and a test set. The training set is submitted to an instance selection procedure in order to provide the K individual training sets (the so-called *bags*) to train the K FURIA-based FRBCSs. In every case, the bags are generated with the same size as the original training set, as commonly done.

The fuzzy classification rules R_j^k considered show a class C_j^k and a certainty degree CF_j^k in the consequent: If x_1^k is A_{j1}^k and $\ldots$ and x_n^k is A_{jn}^k then Class C_j^k with CF_j^k, $j = 1, 2, \ldots, N$, $k = 1, 2, \ldots, K$. The voting-based fuzzy reasoning method is used to take the decision of the individual classifier [35], [15].

After performing the training stage on all the bags in parallel, we get an initial whole FRBCE, which is validated using the training and the test errors as well as a measure of complexity based on the total number of rules in the FRBCSs. The standard majority voting approach is applied as the classifier fusion method [1], [23]: the ensemble class prediction will directly be the most voted class in the component classifiers output set. In the case of a tie, the output class is chosen at random.

For a more detailed description on the methodology, the interested reader is referred to the provided references.

III. AN EVOLUTIONARY MULTIOBJECTIVE OPTIMIZATION APPROACH TO DESIGN A FUZZY LINGUISTIC COMBINATION METHOD FOR BAGGING FRBCES

The next subsections will respectively provide a detailed description of the FRBCS-CM structure introduced in [14] and of the composition of the EMO algorithm based on NSGA-II designed to derive its fuzzy knowledge base.

A. Fuzzy linguistic combination

As said in Section II-C, the FRBCSs considered in the ensemble will be based on fuzzy classification rules with a class and a certainty degree in the consequent. Let R_j^k be the j-th rule of the k-th member of an ensemble of K components,

$$\text{if } x \text{ is } A_j^k \text{ then Class } C_j^k \text{ with } CF_j^k,$$

where $C_j^k \in \{1, \ldots, n_c\}$ and n_c is the number of classes.

We will use the expression $\mathcal{G}^k = \{R_1^k, \ldots, R_{N_k}^k\}$ to denote the list of fuzzy rules comprising the k-th ensemble member. Let us partition each one of these lists into so many sublists $\mathcal{G}_c^k$ as classes. $\mathcal{G}_c^k$ contains the rules of $\mathcal{G}^k$ whose consequent is the class c.

Let us also define $R^k(x)$ to be the intermediate output of the k-th member of the ensemble, which is the fuzzy subset of the set of class labels computed as follows:

$$R^k(x)(c) = \bigvee_{\{j \mid C_j^k = c\}} CF_j^k \cdot A_j^k(x). \tag{1}$$

Each component FRBCS maps an input value x to so many degrees of membership as the number of classes in the problem. The highest of these memberships determines the classification of the pattern. That is to say, the k-th FRBCS classifies an object x as being of class $\text{FRBCS}^k(x) = \arg\max_{c \in \{1,\ldots,n_c\}} R^k(x)(c)$. Observe also that $R^k(x)(c)$ is the result of applying the fuzzy reasoning mechanism to the knowledge base defined by the sublist $\mathcal{G}_c^k$.

The simplest linguistic combination of the component FRBCSs consists of stacking a selection of some of the rules R_j^k into a single large rule base. Let us define a binary matrix $[b_{ck}] \in \{0,1\}^{n_c \times K}$, and let us agree that, if b_{ck} is zero, then $\mathcal{G}_c^k$ is removed from the ensemble and $R^k(x)(c) = 0$. This selection is equivalent to the hierarchical FRBCS comprising n_c expressions of the form:

if (member$_1$ says that class is c) or $\ldots$ or (member$_K$ says that class is c) then class is c,

where the asserts "(member$_k$ says that class is c)" have a degree of certainty b_{ck} determined by the rules in the sublist $\mathcal{G}_c^k$, and those asserts for which b_{ck} is zero are omitted. The fuzzy output of this selected ensemble is

$$R^I(x)(c) = \bigvee_{\{(j,k) \mid C_j^k = c\}} b_{ck} \cdot CF_j^k \cdot A_j^k(x). \tag{2}$$

We can define more powerful linguistic selections which extend this basic fuzzy reasoning schema. In this paper we will use a *sparse* matrix of weights $[w_{ck}] \in [0,1]^{n_c \times K}$ and operate as follows:

$$R^{II}(x)(c) = \bigvee_{\{(j,k) \mid C_j^k = c\}} w_{ck} \cdot CF_j^k \cdot A_j^k(x). \tag{3}$$

Thus, the selected ensemble can be seen as a hierarchical knowledge base with n_c fuzzy classification rules with weights in the antecedent part

if (member$_1(w_{c1})$ says that class is c) or $\ldots$ (member$_K(w_{cK})$ says that class is c) then class is c,

where the asserts "member$_k(w_{ck})$ says that class is c" have a certainty determined by the rules in the sublist $\mathcal{G}_c^k$, after multiplying their confidence degrees by the same factor w_{ck}:

[1]It could be feasible when using a very small number of component classifiers –only three– but not with dealing with a more usual larger number. In fact, the FRBSs considered in their experimentation are only composed of a single rule with three inputs as well as the authors mention they were not able to incorporate expert knowledge to the Bayesian component classifier

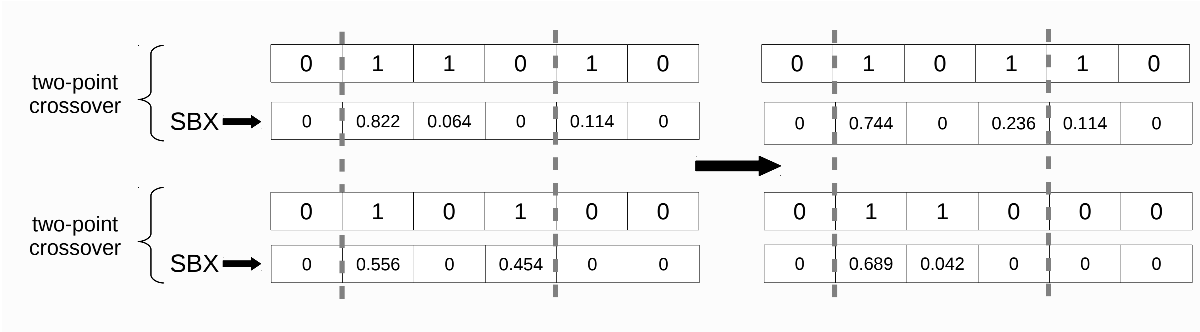

Fig. 1. Coding scheme and crossover operation: an individual is a sparse matrix, which is represented by a list of indexes and a list of values.

$$\text{if } x \text{ is } A_j^k \text{ then Class } C_j^k \text{ with } w_{C_j^k k} \cdot CF_j^k.$$

Again, those rules where $w_{C_j^k k} = 0$ are omitted.

In this case, any of these hierarchical rule bases we have introduced is univocally determined by a matrix $[w_{ck}]$. Therefore, the search of the best selection involves finding the best matrix $[w_{ck}]$, according to two criteria, accuracy and complexity, and guided by the NSGA-II algorithm that will be detailed next. Notice that, this search involves a selection process, because $[w_{ck}]$ is a sparse matrix.

B. Main Components of our NSGA-II Approach

NSGA-II [21] is based on a Pareto dominance depth approach, where the population is divided into several fronts and the depth of each front shows to which front an individual belongs to. A pseudo-dominance rank being assigned to each individual, which is equal to the front number, is a metric used for the selection of an individual.

We propose a hierarchical coding scheme where an individual is composed of binary vector at the first level corresponding to the binary matrix $[b_{ck}]$ and the vector of real numbers at the second associated with the values of the sparse matrix $[w_{ck}]$ (see Figure 1). At the first level it is decided whether the fuzzy rules of the given class of the particular FURIA base classifier are activated or not, while at the second level the weight is assigned to the given fuzzy rules. A binary digit and a corresponding real number are assigned to each gene, i.e. to a subset of rules for a given class by one of FURIA base classifiers. When the binary value is equal to 1, it means that the given fuzzy rules of the corresponding classifier are activated and a real value ($\neq 0$) is provided, while when 0 is obtained by binary value, then the given fuzzy rules are discarded and the real value is also equal to 0.

The initial population is composed of randomly generated individuals. To introduce a high amount of diversity, binary tournament is used as selection mechanism. That means that two individuals are randomly picked from the current population and the best one is selected. The two winners are crossed over to obtain a single offspring that directly substitutes the loser. We have considered the classical two-point crossover at the first level (binary vector) and the SBX crossover at the second (real vector). The standard bit-flip (at the first level) and the uniform (at the second level) mutation operators are used. Both crossover and mutation operators are applied with different pre-specified probabilities.

C. The Evaluation Criteria Used for Two-objective NSGA-II

In this subsection we describe the two considered optimization criteria. We will utilize measures of two different kinds combined into a two-objective fitness function:

- **Accuracy**. The training error (TE), which is a common accuracy measure, is used. We compute the error of each ensemble for a large number of bootstrapped resamples of the training set, and use a quantile of the distribution of these errors. This is intended to avoid overfitting when there are outliers in the training set, and also to detect the most robust selections, which are expected to generalize better.
- **Complexity**. The complexity of the FRBCS-CM is defined as the number of non-zero values. Specifically, it is the number of active terms w_{ck} different than zero in the sparse matrix:

$$Complx = |w_{C_j^k k} \neq 0| \tag{4}$$

IV. Experiments and Analysis of Results

This section is devoted to validate our new EMO-based fuzzy linguistic combination method proposal. While the first subsection introduces the experimental setup considered, the next ones shows the results obtained in the experiments developed and their analysis.

A. Experimental setup

To evaluate the performance of the FRBCS-CM with NSGA-II in the ensembles generated, twenty popular data sets from the UCI machine learning repository have been selected (see Table I). In all of them, every attribute is continuous. As can be seen, the number of features ranges from a small value (5) to a large one (64), while the number of examples does so from 208 to 19,020. We divided them into two groups with Low dimensionality (with $< 15attr.$) and with High dimensionality (with $\geq 15attr.$) as it can be seen in Table I.

TABLE I

DATA SETS CONSIDERED

Data set	#examples	#attr.	#classes
Low dimensional:			
abalone	4178	7	28
breast	700	9	2
glass	214	9	7
heart	270	13	2
magic	19020	10	2
pblocks	5474	10	5
phoneme	5404	5	2
pima	768	8	2
wine	178	13	3
yeast	1484	8	10
High dimensional:			
ionosphere	352	34	2
letter	20000	16	26
optdigits	5620	64	10
pendigits	10992	16	10
sat	6436	36	6
segment	2310	19	7
sonar	208	60	2
spambase	4602	57	2
texture	5500	40	11
vehicle	846	18	4
waveform	5000	40	3

In order to compare the accuracy of the considered classifiers, we used Dietterich's 5×2-fold cross-validation (5×2-cv), which is considered to be superior to paired k-fold cross validation in classification problems [36]. The Friedman test and the Iman-Davenport are used for assessing the statistical significance of the differences between algorithms, while the Holm test is carried out in case of $1 \times n$ comparison [37], [38], [39].

The bagging FRBCEs generated are initially comprised by 50 classifiers. The NSGA-II for the FRBCS-CM derivation works with a population of 100 individuals and runs during 1000 generations. The crossover probability considered is 0.6 and the standard mutation probability is 0.1. A different run is developed with each of the variants proposed for each initial FRBCE, thus resulting in 10 different runs per dataset as a consequence of the 5×2-cv procedure. All the experiments have been run on an Intel quadri-core i5-2400 3.1 GHz processor with 4 GBytes of memory, under the Linux operating system.

Let us call $\overline{P_i^j}$ the non-dominated solution set returned by NSGA-II using the variant of fitness function i in the j-th run for a specific problem instance; $P_i = \overline{P_i^1} \bigcup \overline{P_i^2} \bigcup \ldots \bigcup \overline{P_i^{10}}$, the union of the solution sets returned by the ten runs obtained from 5x2-cv of algorithm i, and finally $\overline{P_i}$ the set of all non-dominated solutions in the P_i set[2] (aggregated Pareto fronts). As a complement to the analysis of the numerical results obtained we will provide graphical representations of some of those aggregated Pareto fronts. When graphically represented, these plots offer a valuable visual information, not measurable, but sometimes more useful than numerical values.

B. Experiments developed

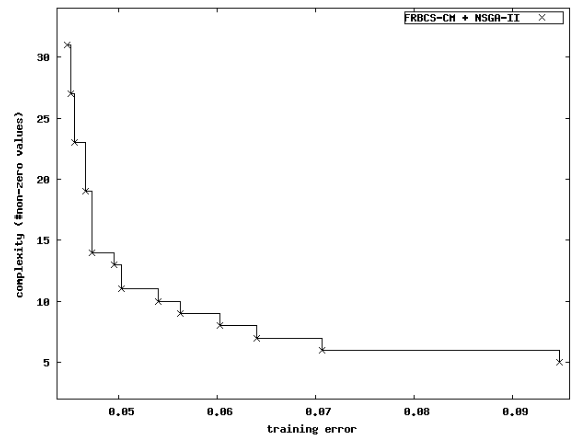

Fig. 2. Graphical representations of the Pareto front approximations obtained from the EMO approach for phoneme dataset. Objective 1 stands for training error and objective 2 for complexity in terms of the number of non-zero values.

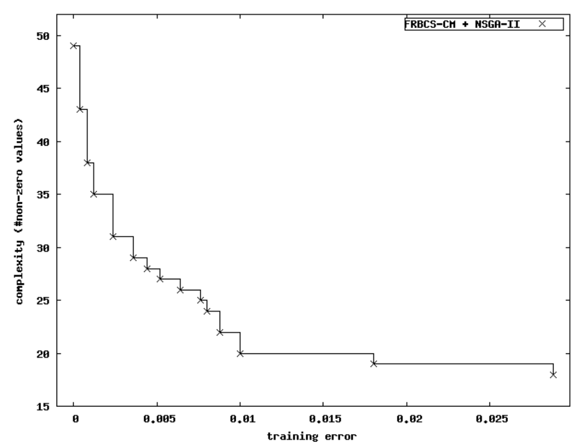

Fig. 3. Graphical representations of the Pareto front approximations obtained from the EMO approach for waveform dataset. Objective 1 stands for training error and objective 2 for complexity in terms of the number of non-zero values.

1) Analysis of the original Pareto front approximations: First of all, in order to give a flavor of the results obtained, we show a visual representation of the aggregated Pareto front approximation for two selected datasets. Figures 2 and 3 represent a visualization of the front obtained for the texture and waveform datasets.

An important conclusion that can be drawn is that our approach works properly as it allows the evolutionary multiobjective method to derive a representative number of solutions in the Pareto set approximations. Furthermore, the proposed NSGA-II generates FRBCS-CM designs spreading widely over the Pareto search space. They reach both edges

[2]Notice that, the pseudo-optimal Pareto front is the fusion of the $\overline{P_i}$ sets generated by every variant of the EMO-based FRBCS-CM in all the runs developed.

TABLE II

COMPARISON OF THE AVERAGED PERFORMANCE OF THE FOUR SINGLE SOLUTIONS SELECTED FROM THE OBTAINED PARETO SETS

		Best train				Best complx				Best tradeoff			
Dataset	Car	Tra	Tst	Cmpl	Red %	Tra	Tst	Cmpl	Red %	Tra	Tst	Cmpl	Red %
Low dim.:													
abalone	13.5	0.4814	0.7560	621.8	55.6	0.5236	0.7524	513.7	63.3	0.4923	0.7538	546.9	60.9
breast	1.8	0.0000	0.0458	5.6	94.4	0.0026	0.0412	3.8	96.2	0.0026	0.0412	3.8	96.2
glass	4.3	0.0028	0.2860	90.9	74.0	0.0439	0.2822	75.7	78.4	0.0196	0.2776	79.3	77.3
heart	1.7	0.0015	0.1763	6.1	93.9	0.0096	0.1778	4.5	95.5	0.0096	0.1778	4.5	95.5
magic	6.9	0.0841	0.1309	22.8	77.2	0.0891	0.1304	6.1	93.9	0.0856	0.1309	11.1	88.9
pblocks	6.8	0.0057	0.0273	75.7	69.7	0.0086	0.0266	46.3	81.5	0.0067	0.0262	54.5	78.2
phoneme	8.3	0.0531	0.1239	24.9	75.1	0.0603	0.1241	6.6	93.4	0.0551	0.1230	12.0	88.0
pima	5.5	0.0193	0.2422	16.5	83.5	0.0378	0.2435	5.6	94.4	0.0255	0.2432	8.5	91.5
wine	1.0	0.0000	0.0405	13.8	90.8	0.0000	0.0405	13.8	90.8	0.0000	0.0405	13.8	90.8
yeast	9.8	0.1364	0.4019	205.8	58.8	0.1733	0.4014	139.4	72.1	0.1441	0.4009	157.6	68.5
Avg. Low	6.0	0.0784	0.2231	108.4	77.3	0.0949	0.2220	81.6	85.9	0.0841	0.2215	89.2	83.6
High dim.:													
ionosphere	1.5	0.0000	0.1396	5.5	94.5	0.0028	0.1396	4.2	95.8	0.0011	0.1401	4.6	95.4
optdigits1	2.7	0.0000	0.0315	133.6	73.3	0.0008	0.0320	127.3	74.5	0.0005	0.0322	129.1	74.2
pendigits	4.0	0.0002	0.0143	151.0	69.8	0.0008	0.0148	130.1	74.0	0.0005	0.0138	135.7	72.9
sat	8.5	0.0028	0.0992	104.6	65.1	0.0083	0.1007	64.6	78.5	0.0040	0.1000	73.7	75.4
segment	3.5	0.0004	0.0303	89.2	74.5	0.0036	0.0328	75.8	78.3	0.0015	0.0304	79.7	77.2
sonar	1.2	0.0000	0.2183	3.4	96.6	0.0019	0.2192	3.2	96.8	0.0019	0.2192	3.2	96.8
spambase	6.1	0.0104	0.0555	21.1	78.9	0.0138	0.0549	5.5	94.5	0.0114	0.0557	9.8	90.2
texture	3.0	0.0001	0.0295	159.6	71.0	0.0013	0.0295	148.9	72.9	0.0006	0.0296	150.0	72.7
vehicle	3.8	0.0002	0.2655	38.7	80.7	0.0104	0.2667	29	85.5	0.0036	0.2603	31.9	84.1
waveform	4.2	0.0001	0.1498	28.1	81.3	0.0026	0.1479	16.5	89.0	0.0008	0.1496	19.9	86.7
Avg. High	3.9	0.0014	0.1033	73.5	78.6	0.0046	0.1038	60.5	84.0	0.0026	0.1031	63.8	82.6
Avg.	4.9	0.0399	0.1632	90.9	77.9	0.0498	0.1629	71.0	85.0	0.0434	0.1623	76.5	83.1

acquiring high performance for the two learning goals: accuracy (training error) and complexity (# of non-zero values).

In order to make a stronger conclusion, particular solutions containing FRBCE designs with different accuracy-complexity tradeoffs are extracted from the Pareto front approximations and analyzed in detail in the next section.

2) Single Solutions Extracted from the Obtained Pareto Front Approximations: Our objective is to analyze the final performance of our proposal by imitating the procedure expected to be followed by a human designer in order to select a desired FURIA-based fuzzy CE structure from those available in the obtained accuracy-complexity non-dominated fronts.

From each Pareto front approximation, we have selected three different solutions, the one having the best value in each of the two objectives that have been optimized, training error and complexity, as well as the one with the best accuracy-complexity trade-off value. The trade-off solution is selected as follows: 1000 random weights $w1 \in [0, 1]$ are computed for each solution and the two learning goals are normalized to $[0, 1]$. The average value of the aggregation function of two learning goals (training error and complexity) $LG1$ and $LG2$ is taken as: $(w1 * LG1 + (1 - w1) * LG2)$, and the solution with the lowest aggregated value is selected. The other more advanced, however highly computational solution is the one presented in [40]. For each solution we present the values of four different measures, training error (*Tra*), complexity in terms of non-zero values (*Cmpl*), test error (*Tst*), and complexity reduction (*Red %*). The average and standard deviation values for each of the four different solutions in the 20 problems are collected in Table II, together with the cardinality of each Pareto set approximation.

In the light of this table, it can be noticed that the three single solutions obtain good performance. The solutions based on the best criterion (Best train and Best complx) obtain low values in their optimization objective criteria, but they also do so considering the second conflicting criteria. For instance, Best train obtains low average training error (0.0399), while maintaining high complexity reduction (77.9%). The best tradeoff (tra-cmpl) solution obtains intermediate values on average for both objectives (training error equal to 0.434 and complexity reduction equal to 83.1%). Considering the test error, all the three approaches obtain similar results. The best tradeoff solution obtains slightly better performance.

As our approach involves the joint optimization of two different conflicting objectives, in our opinion their mixture is the best accuracy-complexity combination for the selection of the final solution. Thus, we will use the solutions with the best trade-off (tra-cmpl) value for the final comparison.

3) Comparison of EMO-based FRBCS-CM with FRBCS-CM and full ensemble with MV: In this subsection we present a final benchmarking of the performance of FRBCS-CM with NSGA-II. We compare it with the full original ensemble using MV and with FRBCS-CM from [14]. This approach is based on a GA, which uses a particular coding scheme, storing only non-zero values. Thus, the complexity of the final FRBCE is defined *a priori*. In order to make a fair comparison, we define its complexity to be similar to the one obtained by our proposal. That is to say, we will use 15 % of non-zero values (equivalent to an 85 % of reduction).

Table III reports the test error of the FRBCEs on the 20 datasets. The proposed FRBCS-CM based on NSGA-II outperforms the other approaches in 14 out of 20 cases (9 out of 10 for high dimensional datasets), obtaining the lowest average test error.

The average rankings of each FRBCE obtained through the Friedman test are shown in Table IV. The Iman-Davenport test indicates significant differences between the algorithms as the p-value is equal to $8.986e$-6.

TABLE III
COMPARISON OF FRBCS-CM WITH NSGA-II WITH THE CLASSICAL METHODS

	MV	FRBCS-CM + NSGA-II	FRBCS-CM 15%
Dataset	Test err.	Test err.	Test err.
Low dim.:			
abalone	**0.7458**	0.7538	0.7559
breast	**0.0409**	0.0412	0.0446
glass	0.2822	**0.2776**	0.3103
heart	0.1822	**0.1778**	0.1815
magic	0.1346	**0.1309**	0.1333
pblocks	0.0288	**0.0262**	0.0278
phoneme	0.1332	**0.1230**	0.1270
pima	**0.2385**	0.2432	0.2503
wine	**0.0393**	0.0405	0.0416
yeast	**0.4008**	0.4009	0.4128
Avg. Low	0.2227	**0.2215**	0.2285
High dim.:			
ionosphere	0.1459	**0.1401**	0.1475
optdigits1	0.0329	**0.0322**	0.0374
pendigits	0.0156	**0.0138**	0.0161
sat	0.1021	**0.1000**	0.1045
segment	0.0336	**0.0304**	0.0316
sonar	0.2269	**0.2192**	0.2202
spambase	0.0587	**0.0557**	0.0560
texture	0.0307	**0.0296**	0.0329
vehicle	0.2726	**0.2603**	0.2747
waveform	**0.1492**	0.1496	0.1553
Avg. High	0.1068	**0.1031**	0.1076
Avg.	0.1647	**0.1623**	0.1681

TABLE IV
AVERAGE RANKINGS OF THE FRIEDMAN'S TEST

Algorithm	Ranking
FRBCS-CM + NSGA-II	1.300
MV	2.050
FRBCS-CM 15%	2.650

TABLE V
HOLM TEST FOR THE COMPARISON OF FRBCS-CM + NSGA-II WITH THE OTHER APPROACHES.

Comparison	p-value
FRBCS-CM + NSGA-II vs MV	**+(0.018)**
FRBCS-CM + NSGA-II vs FRBCS-CM 15%	**+(3.925e-005)**

The adjusted p-values of the Holm test comparing FRBCS-CM + NSGA-II (the control algorithm) with the rest of the FRBCE design approaches are presented in Table V (the results showing a significant difference are presented in bold font). It reveals significant differences in favor of our approach when comparing with all the other FRBCEs.

Concluding, the proposed approach, FRBCS-CM with NSGA-II generates very good results in terms of the test accuracy of the final FRBCEs, which is the lowest than the full ensemble with MV showing on average complexity 6 times higher and FRBCS-CM 15% (showing a similar complexity level) previously proposed. That is confirmed by the statistical tests performed, which indicated significant differences.

V. CONCLUSIONS AND FUTURE WORKS

We have incorporated a EMO algorithm based on NSGA-II in order to automatically derive a CE fuzzy combination method based on the use of a FRBCS. This hybridization between the fuzzy linguistic combination method and the NSGA-II algorithm introduced some interesting characteristics. It eliminates a tedious problem of selecting the complexity of the system *a priori*. Additionally, it offers several FRBCE designs with different accuracy-complexity trade-offs in a single run. We carried out exhaustive experiments using 20 high dimensional datasets from the UCI repository. It turned out that our proposal provided very promising results.

Although these experiments clearly showed the good outcomes of this proposal, our next steps will concentrate on testing different alternative operator mechanisms of the NSGA-II algorithm. Furthermore, we would also like to evaluate its behavior with other classical CE approaches considering the standard machine learning classifiers.

ACKNOWLEDGMENT

This work has been supported by the Spanish *Ministerio de Economía y Competitividad* under the ABSYNTHE (TIN2011-29824-C02-01) and the SOCOVIFI2 (TIN2012-38525-C02-01) projects.

REFERENCES

[1] L. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Wiley, 2004.

[2] T. Ho, "The random subspace method for constructing decision forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, 1998.

[3] D. Optiz and R. Maclin, "Popular ensemble methods: An empirical study," *Journal of Artificial Intelligence Research*, vol. 11, pp. 169–198, 1999.

[4] J. Canul-Reich, L. Shoemaker, and L. O. Hall, "Ensembles of fuzzy classifiers," in *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, London, 2007, pp. 1–6.

[5] W. Pedrycz and K. Kwak, "Boosting of granular models," *Fuzzy Sets and Systems*, vol. 157, no. 22, pp. 2934–2953, 2006.

[6] K. Trawiński, O. Cordón, and A. Quirin, "On designing fuzzy rule-based multiclassification systems by combining FURIA with bagging and feature selection," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 19, no. 4, pp. 589–633, 2011.

[7] K. Trawiński, O. Cordón, A. Quirin, and L. Sánchez, "Multiobjective genetic classifier selection for random oracles fuzzy rule-based classifier ensembles: How beneficial is the additional diversity?" *Knowledge-Based Systems*, vol. 54, pp. 3–21, 2013.

[8] B. Dasarathy and B. Sheela, "A composite classifier system design: Concepts and methodology," *Proceedings of IEEE*, vol. 67, no. 5, pp. 708–713, 1979.

[9] G. Giacinto and F. Roli, "Dynamic classifier selection based on multiple classifier behaviour," *Pattern Recognition*, vol. 34, no. 9, pp. 1879–1881, 2001.

[10] D. Partridge and W. Yates, "Engineering multiversion neural-net systems," *Neural Computation*, vol. 8, no. 4, pp. 869–893, 1996.

[11] L. Lam and C. Suen, "Application of majority voting to pattern recognition: An analysis of its behavior and performance," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 27, pp. 553–568, 1997.

[12] A. Verikas, A. Lipnickas, K. Malmqvist, M. Bacauskiene, and A. Gelzinis, "Soft combination of neural classifiers: A comparative study," *Pattern Recognition Letters*, vol. 20, no. 4, pp. 429–444, 1999.

[13] L. Kuncheva, ""Fuzzy" versus "nonfuzzy" in combining classifiers designed by boosting," *IEEE Transactions on Fuzzy Systems*, vol. 11, no. 6, pp. 729–741, 2003.

[14] K. Trawiński, O. Cordón, L. Sánchez, and A. Quirin, "A genetic fuzzy linguistic combination method for fuzzy rule-based multiclassifiers," *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 5, pp. 950–965, 2013.

[15] O. Cordón, M. del Jesus, and F. Herrera, "A proposal on reasoning methods in fuzzy rule-based classification systems," *International Journal of Approximate Reasoning*, vol. 20, pp. 21–45, 1999.

[16] H. Ishibuchi, T. Nakashima, and M. Nii, *Classification and Modeling With Linguistic Information Granules*. Springer, 2005.

[17] O. Cordón, F. Herrera, F. Hoffmann, and L. Magdalena, *Genetic Fuzzy Systems. Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*. World Scientific, 2001.

[18] O. Cordón, F. Gomide, F. Herrera, F. Hoffmann, and L. Magdalena, "Ten years of genetic fuzzy systems: Current framework and new trends," *Fuzzy Sets and Systems*, vol. 141, no. 1, pp. 5–31, 2004.

[19] J. Casillas, F. Herrera, R. Pérez, M. del Jesus, and P. Villar, "Special issue on genetic fuzzy systems and the interpretability-accuracy trade-off," *International Journal of Approximate Reasoning*, vol. 44, no. 1, January 2007.

[20] C. Coello, G. Lamont, and D. V. Veldhuizen, *Evolutionary Algorithms for Solving Multi-Objective Problems, 2nd Edition*. Springer, 2007.

[21] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, pp. 182–197, 2002.

[22] K. Woods, W. Kegelmeyer, and K. Bowyer, "Combination of multiple classifiers using local accuracy estimates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 4, pp. 405–410, 1997.

[23] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–238, 1998.

[24] E. Dos Santos, R. Sabourin, and P. Maupin, "A dynamic overproduce-and-choose strategy for the selection of classifier ensembles," *Pattern Recognition*, vol. 41, no. 10, pp. 2993–3009, 2008.

[25] B. Gabrys and D. Ruta, "Genetic algorithms in classifier fusion," *Applied Soft Computing*, vol. 6, no. 4, pp. 337–347, 2006.

[26] N. Dimililer, E. Varoglu, and H. Altincay, "Classifier subset selection for biomedical named entity recognition," *Applied Intelligence*, vol. 31, pp. 267–282, 2009.

[27] S.-B. Cho and J. Kim, "Multiple network fusion using fuzzy logic," *IEEE Transactions on Neural Networks*, vol. 6, no. 2, pp. 497–501, 1995.

[28] M. Abreu and A. Canuto, "An experimental study on the importance of the choice of the ensemble members in fuzzy combination methods," in *Seventh International Conference on Intelligent Systems Design and Applications (ISDA)*, Rio de Janeiro, 2003, pp. 723–728.

[29] P. Bulacio, S. Guillaume, E. Tapia, and L. Magdalena, "A selection approach for scalable fuzzy integral combination," *Information Fusion*, vol. 11, no. 2, pp. 208–213, 2010.

[30] Y. Lu and F. Yamaoka, "Fuzzy integration of classification results," *Pattern Recognition*, vol. 30, no. 11, pp. 1877–1891, 1997.

[31] J. C. Hühn and E. Hüllermeier, "FURIA: an algorithm for unordered fuzzy rule induction," *Data Mining and Knowledge Discovery*, vol. 19, no. 3, pp. 293–319, 2009.

[32] ——, "An analysis of the FURIA algorithm for fuzzy rule induction," in *Advances in Machine Learning I*, 2010, pp. 321–344.

[33] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[34] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Transactions on Neural Networks*, vol. 5, no. 4, pp. 537–550, 1994.

[35] H. Ishibuchi, T. Nakashima, and T. Morisawa, "Voting in fuzzy rule-based systems for pattern classification problems," *Fuzzy Sets and Systems*, vol. 103, no. 2, pp. 223–238, 1999.

[36] T. Dietterich, "Approximate statistical test for comparing supervised classification learning algorithms," *Neural Computation*, vol. 10, no. 7, pp. 1895–1923, 1998.

[37] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

[38] S. García and F. Herrera, "An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons," *Journal of Machine Learning Research*, vol. 9, pp. 2677–2694, 2008.

[39] S. García, A. Fernández, J. Luengo, and F. Herrera, "Advanced non-parametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power," *Information Sciences*, vol. 180, no. 10, pp. 2044–2064, 2010.

[40] H. Ishibuchi and Y. Nojima, "Repeated double cross-validation for choosing a single solution in evolutionary multi-objective fuzzy classifier design," *Knowledge-Based Systems*, vol. 54, pp. 22–31, 2013.