Improving a Fuzzy Association Rule-Based Classification Model by Granularity Learning based on Heuristic Measures over Multiple Granularities

Michela Fazzolari and Rafael Alcalá
Dept. of Computer Science
and Artificial Intelligence
University of Granada
18071 Granada, Spain
Email: {fazzolari,alcala}@decsai.ugr.es

Yusuke Nojima and Hisao Ishibuchi
Dept. of Computer Science
and Intelligent Systems
Osaka Prefecture University
1-1 Gakuen-cho, Naka-ku, Sakai
Osaka 599-8531, Japan
Email: {nojima,hisaoi}@cs.osakafu-u.ac.jp

Francisco Herrera
Dept. of Computer Science
and Artificial Intelligence
University of Granada
18071 Granada, Spain
Email: herrera@decsai.ugr.es

***Abstract*—A multi-objective evolutionary fuzzy rule selection process extracts a subset of fuzzy rules from an initial set, by applying a multi-objective evolutionary algorithm. Two approaches can be used to determine the number of terms (i.e. the granularity) associated with the linguistic variables that appear in the rules: a pre-established single granularity can be chosen, or a multiple granularities approach can be preferred. The latter favors a reduction in the number of extracted rules, but it also brings to a possible loss of interpretability. To prevent this problem, suitable granularities can be determined by applying automatic techniques before the initial rule generation process.**

In this contribution, we investigate how the application of a single granularity learning approach influences the performance of fuzzy associative rule-based classifiers. The aim is to reduce the complexity of the obtained models, trying to maintain a good classification ability.

I. Introduction

A fuzzy rule-based system (FRBS) is composed of fuzzy rules. It has been frequently used in several fields and for a wide range of problems, such as classification, regression, modeling, control, etc. The fuzzy rules can be formulated by an expert or generated automatically considering numerical data that describe a certain phenomenon. To this aim, several techniques have been proposed in the past, but they usually aim to improve only the accuracy of the system, without considering the interpretability issue, which is one of the main advantages of FRBSs [1].

Among the proposed techniques, Evolutionary Algorithms (EAs) have been usefully applied, since they can generate automatically a fuzzy model by evolving its parameters in the evolutionary process. Initially, EAs only considered a single objective, but when the multi-objective problem was pointed out, they were extended to Multi-Objective Evolutionary Algorithms (MOEAs) [2], [3]. These algorithms can optimize multiple objectives and generate a group of solutions instead of a single one, in which each solution satisfies an objective with higher degree than the others.

Therefore, MOEAs have been used to address the accuracy-interpretability trade-off in FRBSs [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16]. The hybridization between MOEAs and FRBSs is called Multi-Objective Evolutionary Fuzzy Systems (MOEFSs).

In [10], [11], two of the most renowned works in the Evolutionary Multi-Objective Optimization (EMO) field, the authors have applied MOEAs to carry out a genetic fuzzy rule selection process, starting from an initial set of fuzzy rules. The aim is to obtain a set of Fuzzy Rule-Based Classifiers (FRBCs), considering two objectives at the same time: the classification accuracy and the number of rules. Then, in [13] a third objective was added to minimize the length of the rules.

When performing evolutionary or genetic fuzzy rule selection, the appropriate number of Membership Functions (MFs) for each variable, i.e. the granularity, is not known beforehand. The two possibilities to address this problem are to previously fix a single granularity [10], [11] or to adopt multiple granularities [13]. The former approach is simpler, although the choice of the granularity is often performed by hand, and usually induces the generation of a high number of fuzzy rules. On the other hand, the multiple granularities approach is useful to reduce the number of rules in the obtained models, but the interpretability loss has often been pointed out.

For this reason, in [17] the authors proposed a mechanism to identify appropriate single granularities while performing a multi-objective evolutionary fuzzy rule selection process, based on the proposal presented in [13]. The framework includes four steps: a) first, a heuristic procedure is used to create a pre-specified number of promising fuzzy rules; b) then, for each attribute a single granularity is learnt, considering the frequency of used partitions and the importance of the rules extracted in the previous step; c) next, these granularities are used to extract again a pre-specified number of fuzzy rules; d) finally, a multi-objective evolutionary algorithm is used to perform the rule selection process.

The present work proposes a method that combines the single granularity specification mechanism presented in [17]

978-1-4673-5899-6/13/$31.00 ©2013 IEEE

with a new multi-objective version of a fuzzy associative classification algorithm based on FARC-HD [18]. We name this method MO-FARCG. The aim is to avoid the use of multiple granularities but still reducing the complexity of the obtained classifier while maintaining high generalization ability, by considering both objectives within a multi-objective evolutionary framework. The method has been compared with the original FARC-HD and the results show that, by learning granularities, the complexity of the obtained models is significantly reduced while accuracy performance is maintained or slightly affected.

This proposal is arranged as follows. Section II introduces some preliminary concepts about FRBCs and fuzzy association rules. Section III describes in detail each stage of the proposed approach. Section IV presents the experimental setup and discusses the obtained results. Finally, in Section VI, some concluding remarks are pointed out.

II. Preliminaries

In this section, FRBCs are briefly introduced. Then, fuzzy association rules are described and their application to classification problems is explained.

A. Fuzzy Rule-Based Classifiers

Let us consider a set of m patterns $\mathbf{x}_p = (x_{p1}, ..., x_{pn})$, $p = 1, 2, ..., m$ to be classified and a set of M classes to be assigned. Considering an n-dimensional pattern space, x_{pi} is the attribute value of the p-th pattern for the i-th attribute ($i = 1, \ ..., \ n$). A classification problem is to assign a label to each pattern, in a way that it is consistent with some observed data we know about the problem (training data).

In this proposal, fuzzy rules of the following type are used:

$$\begin{aligned} R_q: \ & \text{If } x_1 \text{ is } A_{q1} \text{ and ... and } x_n \text{ is } A_{qn} \\ & \quad \text{then Class } C_q \text{ with } RW_q, \end{aligned} \tag{1}$$

where R_q is the label of the q-th fuzzy rule, $\mathbf{x} = (x_1, ..., x_n)$ is an n-dimensional pattern vector, A_{qi} is an antecedent fuzzy set ($i = 1, ..., n$), C_q is a class label, and RW_q is the rule weight. The antecedents fuzzy sets of R_q are denoted as a fuzzy vector $\mathbf{A}_q = (A_{q1}, \ A_{q2}, \ ..., \ A_{qn})$. Each new pattern is classied as the class with the maximum total strength of the vote.

The performance of FRBCs is highly affected by the rule weight RW_q associated with each fuzzy rule [19]. The rule weight can be defined in different ways and many mechanisms have been proposed in the literature. For example, in [20], the authors present several heuristic methods that can be used to specify the weight of fuzzy rules. In this study, the most common one has been chosen, i.e. the fuzzy confidence value or certainty factor (CF) [21].

The certainty factor is widely used for fuzzy classification as it just affects the strength of each fuzzy IF-THEN rule in the classification phase, without changing the positions of the antecedent fuzzy sets [19].

B. Fuzzy Association Rules for Classification

Association rules are used to represent and identify dependences between items in a database [22], [23]. They are expressions of the type $A \rightarrow B$, where A and B are sets of items, and $A \cap B = \oslash$. This means that if all the items in A exist in a transaction, then all the items in B with a high probability are also in the transaction, and A and B should have no common items [24].

Several studies have been presented in which association rules are extracted from databases with binary or discrete values, but many real-world problems include also quantitative data. Therefore, the new trend in this field is to design mining algorithms that can handle different types of data.

In this case, the fuzzy set theory represents an useful tool, due to its affinity with the human reasoning [25]. Fuzzy association rules can be used to describe associations between data in a more effective way than simple association rules do. In fact, the linguistic representation makes them more understandable for human experts and fuzzy sets avoid sharp boundaries when partitioning the domain of an attribute. Recently, the mining of fuzzy association rules from quantitative data has been investigated in several studies [26], [27], [28].

An example of fuzzy association rule is a rule of the type:

$$A \ is \ Middle \rightarrow B \ is \ High. \tag{2}$$

where A and B are the attributes present in the database and *Middle* and *High* are linguistic terms associated to these variables.

The interestingness of a fuzzy association rule is commonly evaluated by two measures, namely support and confidence, which can be defined as follows:

$$Support(A \rightarrow B) = \frac{\sum_{x_p \in T} \mu_{AB}(x_p)}{|N|} \tag{3}$$

$$Confidence(A \rightarrow B) = \frac{\sum_{x_p \in T} \mu_{AB}(x_p)}{\sum_{x_p \in T} \mu_{A}(x_p)} \tag{4}$$

where $|N|$ is the number of transactions that appear in the database T, $\mu_A(x_p)$ is the matching degree of the transaction x_p with the antecedent of the rule, and $\mu_{AB}(x_p)$ is the matching degree of the transaction x_p with the antecedent and consequent of the rule.

In recent years, fuzzy association rules have been investigated to be used as classification rules [29], [30], [31]. A fuzzy association rule can be used as a classification rule if its consequent part includes only one class label ($C = C_1, ..., C_j, ..., C_S$). This type of rule is called *fuzzy associative classification rule* and assumes the following form:

$$A \rightarrow C_j \tag{5}$$

As common fuzzy association rules, it can be evaluated in terms of support and confidence:

$$Support(A \rightarrow C_j) = \frac{\sum_{x_p \in ClassC_j} \mu_{A}(x_p)}{|N|} \tag{6}$$

$$Confidence(A \to C_j) = \frac{\sum_{x_p \in ClassC_j} \mu_A(x_p)}{\sum_{x_p \in T} \mu_A(x_p)} \quad (7)$$

III. Multi-Objective Fuzzy Association Rule-Based Classifier with Granularity Learning (MO-FARCG)

In this section, the integration between the granularity learning and a fuzzy associative classification algorithm will be described. In this case, the framework is organized into three stages:

1) Setting stage: Learning the Appropriate Granularities. First, for each class a fixed pre-specified number of rules with multiple granularities are obtained, according to well-known data mining rule evaluation measures [24]. Then, a single granularity for each attribute is chosen, depending on the frequency of extracted rules and some quality measures.
2) Learning stage: Extraction of Candidate Fuzzy Association Rules. All the possible frequent fuzzy itemsets are listed in a search tree and then fuzzy association rules are generated. Finally, rules are evaluated and sorted following a criterion and only the best rules are maintained in order to reduce the number of candidate rules.
3) Post-processing stage: Multi-Objective Evolutionary Fuzzy Rule Selection and MFs Tuning. The best cooperative rules are selected and tuned by using a Multi-Objective Evolutionary Algorithm based on the Strength Pareto Evolutionary Algorithm 2 (SPEA2) [32], exploiting the positive synergy of both techniques within the same process.

A. Setting stage: Learning the Appropriate Granularities

In this stage an initial set of rules is extracted, following the multiple granularity approach presented in [17]. Then, a single granularity for each attribute is chosen on the basis of some information included in the extracted rules. These granularities will be used in the next stage to obtain a new initial set of fuzzy rules.

1) Extracting Rules with Multiple Granularities: At the beginning, a multiple granularity is used for all attributes: fourteen fuzzy sets are considered, distributed into four fuzzy partitions (see Figure 1). An additional fuzzy set is also used, to represent a *don't care* condition (the domain interval $[0, 1]$), then the overall number of possible fuzzy rules is 15^n.

This number of rules is too large to be considered for the generation of the initial candidate rule set, therefore only rules with a small number of antecedent conditions are selected. This number is fixed on 3 for datasets with less than 30 attributes, and 2 for datasets with a number of attributes equal to or bigger than 30.

The heuristic procedure presented in [33] is used to determine the rule weight CF_q and the consequent class C_q for each fuzzy rule R_q. For the antecedent part $\mathbf{A}_q$, the confidence of each class is calculated as

$$c(\mathbf{A}_q \Rightarrow \text{Class } h) = \frac{\sum_{\mathbf{x}_p \in \text{ Class } h} \mu_{\mathbf{A}_q}(\mathbf{x}_p)}{\sum_{p=1}^{m} \mu_{\mathbf{A}_q}(\mathbf{x}_p)}, \quad h = 1, \dots, M. \quad (8)$$

Then, the consequent class C_q is specified according to the class with the maximum confidence:

$$c(\mathbf{A}_q \Rightarrow \text{Class } C_q) = \max_{h=1,2,\dots,M} \{c(\mathbf{A}_q \Rightarrow \text{Class } h)\}. \quad (9)$$

As previously said, the performance of FRBCs is highly influenced by the rule weight. In this proposal, the following definition of CF_q is used, it has led to good results [20]:

$$CF_q = c(\mathbf{A}_q \Rightarrow \text{Class}\, C_q) - \sum_{\substack{h=1 \\ h \neq C_q}}^{M} c(\mathbf{A}_q \Rightarrow \text{Class}\, h). \quad (10)$$

In order to maintain only useful rules, a fuzzy rule R_q is not selected as a candidate rule if its confidence is smaller than 0.5.

The heuristic technique described before generates a large number of short fuzzy rules as candidate rules, including not interesting rules. To decrease this number and to select the most useful rules, a preventive rule reduction is performed. To this end, rules are evaluated and sorted according to the product $p(R_q) = s(R_q) \cdot c(R_q)$, where $c(R_q)$ is the confidence and $s(R_q)$ is the support, that is the percentage of samples covered by R_q. Finally, for each class, the best 300 rules are chosen.

2) Specifying a Single Granularity for Each Attribute: The pool of extracted fuzzy rules contains information that can be used to choose a single granularity for each attribute. This can be done by considering how many times a granularity appears in the extracted rules (weighted by the corresponding rule importance). This approach has been already used in [17] to establish a single granularity.

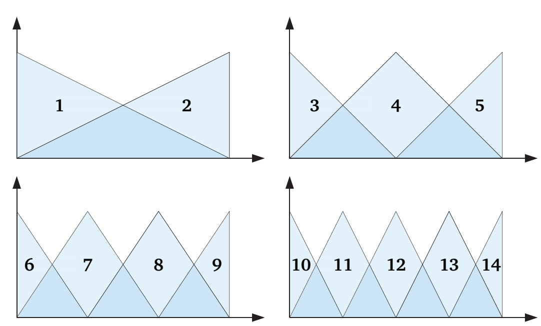

Fig. 1. The fourteen antecedent fuzzy sets considered for each variable.

In particular, we use the following specification:

$$Gr(i) = \underset{g=2,\ldots,5}{\operatorname{argmax}} \left\{ \sum_{Gran(\mathbf{A}_{qi})=g} \operatorname{Imp}(\mathbf{R}_q) \right\}, \qquad (11)$$

where $Gran(\mathbf{A}_{qi})$ is the granularity of the partition that include the fuzzy set used in attribute i of rule R_q and $\operatorname{Imp}(R_q)$ is a criterion to measure the importance of the rule. Frequency, confidence, weight, support and product of confidence and support are some of the criteria that can be used.

In order to promote more general rules, two different approaches were investigated in the original proposal [17]: 1-ALL approach and 1-2-3 approach. Both of them give priority to those rules which include a single antecedent condition in the attribute considered. The only difference is when all the rules presents more than one condition: in 1-ALL approach, the granularity is obtained by considering all the rules including the attribute itself, without taking into account the number of conditions, whereas, in 1-2-3 approach, the priority is given to shorter rules.

In this work, we focus on the product and confidence criteria, combined with 1-ALL approach, since these combinations provided better results. At the end of this stage, a single granularity is obtained for each attribute.

B. Learning stage: Extraction of Candidate Fuzzy Association Rules

In this stage, the attributes' granularities learnt in the previous phase are used to extract from the original data a new set of fuzzy association rules. Afterwards, this set is reduced by applying a prescreening procedure. These two steps are as follows.

1) Rule Extraction: For each class, all the possible fuzzy itemsets are constructed, using a search tree. The root level of each tree (level 0) is generated as an empty set. All attributes are assumed to have an order, which is the order of appearance in the training data, and all the one-itemsets constitute the first level of the search tree, following their order (level 1). The further level (level 2) for an attribute A is constructed by considering all the two-itemsets that combine the one-itemset of attribute A with all the one-itemsets for the other attributes in the order. The same procedure is used to construct the following levels of the tree. No repeated itemsets appear in the tree.

Moreover, each itemset is evaluated with respect to a minimum support and a minimum confidence: an itemset with a support higher than the minimum support is a frequent itemset and an itemset with a confidence higher than the confidence's threshold has reached the quality level demanded by the user. Therefore, if the support of an *n*-itemset in a node A is lower than the minimum support, the node is not extended anymore. At the same way, if the classification rule associated to an item set has a confidence higher than the minimum confidence, the correspondent node does not need to be extended further.

The procedure above lists all the frequent fuzzy itemsets, which are used to generate the candidate fuzzy association rules. Each rule will contain a frequent itemset in the antecedent and the corresponding class in the consequent and for all the classes this process is repeated. The number of frequent fuzzy itemsets extracted depends directly on the minimum support, which is defined for each class using the distributions of the classes over the data set. This stage generates a large number of rules, which can be hardly handled by human users. To simplify the understanding of the model, the depth of the trees ($Depth_{max}$), and so the length of the fuzzy rules, is limited to a fixed value.

2) Rule Prescreening: The rule extraction process generates a large number of rules, which can cause a problem of rule redundancy. To decrease this number by selecting only the best rules, a subgroup discovery technique is used, in particular the pattern weighting scheme described in [34].

Each pattern is associated to a weight $w(i) = \frac{1}{i+1}$, where i stores how many times the pattern has been covered by a rule. Initially, all the weights assume the same value $w(0) = 1$.

For each class, the algorithm selects the best rule, then the weights related to the patterns covered by this rule are decreased. In this way, the patterns that are still uncovered will have a greater possibility of being covered in the following iterations. When the i counter reach a threshold k_t, the correspondent pattern is deleted.

The remaining rules are sorted again and the procedure is repeated until either all patterns have been deleted, or there is no rule left in the rule base.

To evaluate the quality of fuzzy rules, we use a modification of the wWRAcc' measure described in [33]. The wWRAcc' measure has been modified to enable the handling of fuzzy rules. The new measure is defined as follows:

$$wWRAcc''(A \rightarrow C_j) = \frac{n''(A \cdot C_j)}{n'(C_j)} \cdot \left(\frac{n''(A \cdot C_j)}{n''(A)} - \frac{n(C_j)}{N}\right) \qquad (12)$$

where $n''(A)$ is the sum of the products of the weights of all covered patterns by their matching degrees with the antecedent part of the rule, $n''(A \cdot C_j)$ is the sum of the products of the weights of all correctly covered patterns by their matching degrees with the antecedent part of the rules, $n(C_j)$ is the number of patterns of class C_j and $n'(C_j)$ is the sum of the weights of patterns of class C_j.

C. Post-Processing Stage: Multi-Objective Evolutionary Fuzzy Rule Selection and Membership Functions Tuning

In the final step, a modification of the well-known SPEA2 is applied as post-processing algorithm to the KB generated by the previous stage. A similar version of this algorithm was already used for regression problems in [9], in which three objectives are considered instead of two. The SPEA2 was preferred to the well-known NSGA-II since in [4] the approaches based on SPEA2 have been demonstrated to be more appropriate when a tuning of MFs is performed.

This SPEA2 modification performs a fuzzy rule selection together with a tuning of the MFs, aiming to improve the model accuracy as the first objective and to reduce the model complexity as the second objective. In the next sections, the main components of this algorithm are described, and then, the specific characteristics and its main steps are presented.

1) Objectives: Each chromosome is associated with a bi-dimensional vector, whose elements express the fulfillment degree of the following two objectives, respectively:

- classification error minimization: it is represented by the complement of the number of the classification rate;
- complexity minimization: it is represented by the number of selected rules;

To compute the classification error, the following function has been used:

$$Fitness(C) = 1 - \frac{\#Hits}{N} \tag{13}$$

where $\#Hits$ is the number of patterns correctly classified and N is the total number of patterns.

2) Coding scheme and Initial Gene Pool: A double coding scheme for both *rule selection* (C_S) and *tuning* (C_T) is used: $C^p = C_S^p C_T^p$, where C^p is the chromosome representing the individual p. The $C_S^p = (c_{S1}, \ldots, c_{Sm})$ part is represented by a binary-coded string with m genes, where m is the number of initial rules. Each gene contains a values of "1" if the correspondent rule is selected, "0" otherwise. The C_T^p part uses a real coding scheme to codify the three definition parameters of the triangular MFs, where m^i is the number of labels in the database for each of the n variables.

$C_T^p = C_1 C_2 \ldots C_n$
$C_i = (a_1^i, b_1^i, c_1^i, \ldots, a_{m_i}^i, b_{m_i}^i, c_{m_i}^i)\ i = 1, \ldots, m$

The first individual of the first population codifies the KB obtained by the previous step. The remaining individuals of the first population are generated randomly, with each value within the corresponding variation intervals.

3) Crossover and Mutation: An intelligent crossover and mutation operators are described in this section. Each offspring is obtained in the following way.

- First, the C_T part of the offspring is obtained by applying blend crossover (BLX)-0.5 [35] to the C_T part of the parents.
- Then, the binary part C_S is generated depending on the C_T parts of parents and offspring: for each gene in the C_S part, the following steps are performed.
  - Each gene of the C_T part which represents the corresponding MFs of the rule, is considered for both parents and offspring. The MFs of these three rules are extracted.
  - Between the offspring rule and each parent rule, euclidean normalized distances are computed by considering the center points of the MFs involved in these rules. The differences between each pair of centers are normalized by the amplitudes of their respective variation intervals.
  - The parent's rule closer to the offspring's rule is selected and its value is duplicated in the C_S part of the offspring.

This process is repeated until each gene in the C_S part of the offspring is obtained. In each step four offspring are generated, although after applying mutation only the two best offspring are maintained. This type of crossover prevents the recovery of a bad rule already discarded, while permits the recovery of a rule that can be still considered good due to its MFs configuration.

The crossover operator performs a better exploration in the C_S part, therefore the mutation operator does not need to add rules. It simply changes randomly a gene value in the C_T part and sets to zero a random gene in the C_S part, with probability P_m.

The application of these operators brings some advantages: the crossover between individuals with very different rules allows the algorithm to explore different parts of the search space, while the mutation promotes rule extraction, since it is used to remove unnecessary rules.

4) Modifications of the classical SPEA2: Some changes have been introduced to the original selection mechanism of SPEA2, to improve the algorithm's search ability.

- A mechanism to prevent incest has been included, based on the concepts of CHC presented in [36]. This avoid premature convergence in the real coding (C_T) part, which has a greater influence on the algorithm convergence and represents a wider search space than the binary coding part (C_S). In the CHC approach, parents are crossed only if their Hamming distance divided by 4 exceeds a threshold. To follow this approach, the real coding scheme needs to be converted in a binary one, thus each gene is transformed using a gray code with a fixed number of bits per gene (BGene). The threshold value is initially set to $L = (\#CT \times BGene)/4$, where $\#CT$ is the number of genes in the C_T part of the chromosome. This value is decreased by 1 at each generation of the algorithm, therefore in further generations closer solutions can be crossed.
- A restart operator has been introduced to renew the external population when we detect that all the crossover are allowed. Actually, to prevent premature convergence, the first restart is applied if 50% of crossovers are detected at any generation (the required ratio can be defined as $\%_{required} = 0.5$). Each time the restart is performed, the required ratio is updated ad follows: $\%_{required} = (1 + \%_{required})/2$.
  The external population after the restart includes the individuals with the best value in each objective, and the remaining individuals are initialized as follows: the C_S part is copied from the most accurate individual, while the values in the C_T part are generated randomly. In this way, the most accurate and interpretable solutions obtained so far are preserved.
  Some constraints to the application of restart have been introduced: a) a new restart cannot be applied if the most accurate solution has not been improved; b) the restart is not applied at the end, when the approximation of the Pareto front is well formed and needs to be preserved; c) restart is disabled if the midpoint of the total evaluations number is reached and it has been never applied before.

- A mechanism to promote the most accurate solutions has been introduced. At each stage of the algorithm, between restarting points, the number of solutions in the external population ($\bar{P}_{t+1}$) that can be used to constitute the mating pool is reduced progressively and the most accurate solutions are preferred. To this end, solutions are sorted according their accuracy and the number of eligible solutions is reduced progressively from 100% at the beginning to 50% at the end of each stage. This mechanism is disabled in the last evaluations (when restart is disabled too), in order to obtain a wide and well-formed Pareto front.

IV. Experimental framework

To evaluate the performance of the proposed approach with respect to the original FARC-HD, we have considered 24 real-world datasets (Table I). Only datasets with continuous attributes have been considered, since the two first steps of the method have not been designed to handle nominal data. Moreover, in case of instances presenting missing values, they have been removed from the datasets (Cleveland).

All datasets can be downloaded from the Knowledge Extraction based on Evolutionary Learning (KEEL)-dataset repository (`http://sci2s.ugr.es/keel/datasets.php`).

TABLE I
Datasets considered in the study.

Name	Attributes (R/I/N)	Patterns	Classes
Appendicitis	7 (7/0/0)	106	2
Cleveland	13 (13/0/0)	297(303)	5
Ecoli	7 (7/0/0)	336	8
Glass	9 (9/0/0)	214	7
Heart	13 (1/12/0)	270	2
Iris	4 (4/0/0)	150	3
Magic	10 (10/0/0)	19020	2
Monks	6 (0/6/0)	432	7
Movement Libras	90 (90/0/0)	360	15
Page-blocks	10 (4/6/0)	5472	5
Penbased	16 (0/16/0)	10992	10
Phoneme	5 (5/0/0)	5404	2
Pima	8 (8/0/0)	768	2
Ringnorm	20 (20/0/0)	7400	2
Satimage	36 (0/36/0)	6435	6
Sonar	60 (60/0/0)	208	2
Spambase	57 (57/0/0)	4597	2
Spectfheart	44 (0/44/0)	267	2
Texture	40 (40/0/0)	5500	11
Twonorm	20 (20/0/0)	7400	2
Vowel	13 (10/3/0)	990	11
Wdbc	30 (30/0/0)	569	2
Wine	13 (13/0/0)	178	3
Yeast	8 (8/0/0)	1484	10

To carry the different experiments out, a ten-fold cross-validation model is considered: each dataset is randomly split into ten folds, each containing 10% of the patterns of the dataset. Then, a single fold is used for testing and the remaining for training. The cross-validation process is repeated ten times, with each fold used exactly once for testing. For each of the ten partitions, three trials of the algorithm are executed and finally the results are averaged out over 30 runs.

The proposed method is called MO-FARCG and it has been compared with the original approach FARC-HD [18]. Two versions of MO-FARCG have been considered, using product (MO-FARC-prod) and confidence (MO-FARC-conf) criteria, respectively. Due to the multi-objective nature of the SPEA2 included in MO-FARCG, the average of the most accurate solution from all the Pareto fronts is considered for the comparison.

Table II summarizes the parameters used for the methods' analysis. In this proposal, the maximum number of antecedents allowed for a fuzzy rule is restricted to 3 (*short* fuzzy rules), in order to encourage the generation of simpler models. While this was reported by the authors that the original FARC-HD did not present any change over 15000 evaluations, the proposed approach needs 20000 evaluations to reach the convergence. For this reason we fix 15000 for FARC-HD and 20000 for MO-FARCG, since no changes are obtained for the original approach over the specified number of evaluations. The parameter k_t is the threshold beyond which the covered patterns are completely eliminated.

TABLE II
Parameters of the methods considered for comparison.

Method	Parameters
FARC-HD	*Minsup* = 0.05, *Maxconf* = 0.85, $Depth_{max}$ = 3, k_t = 2, *Pop* = 50, *Evaluations* = 15000 *BITSGENE* = 30
MO-FARCG-prod	*Minsup* = 0.05, *Maxconf* = 0.85, $Depth_{max}$ = 3, k_t = 2, *Pop* = 50, *Evaluations* = 20000 *BITSGENE* = 30
MO-FARCG-conf	*Minsup* = 0.05, *Maxconf* = 0.85, $Depth_{max}$ = 3, k_t = 2, *Pop* = 50, *Evaluations* = 20000 *BITSGENE* = 30

Statistical analysis [37], [38] was adopted to evaluate the results, and in particular non-parametric tests, following the recommendations presented in [39], where a set of simple and robust non-parametric tests for statistical comparisons of classifiers has been described.

The Wilcoxons signed-ranks test [40], [41] for pair-wise comparison was used, with a confidence level of $\alpha = 0.05$ in all cases. A wider description of this test and a software to perform it can be found on the web site available at: `http://sci2s.ugr.es/sicidm/`.

V. Experimental Results

This section shows the results of the experiments described in the previous section. Table III shows the average number of rules/conditions (**#R/#C**) and classification percentages in training (**Tra**) and test (**Tst**) of the most accurate solution from each of the obtained Pareto fronts, for the two versions of MO-FARCG, and of the best solution for FARC-HD. The overall mean values for each method are highlighted in the last row.

A first comparison has been drawn between the two different approaches of MO-FARCG: the Wilcoxon's signed-ranks test has been applied to establish if the two versions are statistically equivalent (null-hypothesis). Table IV shows that the Wilcoxon test applied on the test classification percentage of the most accurate solutions rejects the null hypothesis, since $p - value \leq /alpha$. Therefore, the two approaches are not

TABLE III
RESULTS.

	MO-FARCG-prod				MO-FARCG-conf				FARC-HD			
DATASETS	#RULE	#COND	TRA	TST	#RULE	#COND	TRA	TST	#RULE	#COND	TRA	TST
Appendicitis	4.73	1.83	93.43	85.79	4.80	1.68	93.26	86.09	6.8	1.8	93.82	84.18
Cleveland	21.23	2.85	79.03	56.22	24.53	2.87	81.93	56.45	61.3	2.9	88.18	55.24
Ecoli	7.07	2.75	64.41	62.80	7.07	2.75	64.41	62.80	33.8	2.4	92.33	82.19
Glass	10.23	2.35	76.38	66.56	11.20	2.38	75.89	68.55	22.7	2.5	81.1	70.24
Heart	14.40	2.60	93.00	81.48	15.70	2.58	93.53	82.22	27	2.6	93.91	84.44
Iris	4.77	1.13	98.20	95.56	4.77	1.15	98.17	95.56	4	1.1	98.59	96
Magic	4.23	2.24	81.15	80.80	10.77	1.86	83.58	82.99	43.3	2.5	85.36	84.51
Monk2	13.53	1.92	100.00	99.47	12.43	1.65	100.00	99.92	14.2	2	99.92	99.77
Movement libras	46.33	2.94	94.31	71.67	46.40	2.94	94.72	72.87	83.1	2.9	95.52	76.67
Pageblocks	5.77	2.17	94.40	94.30	6.17	2.05	94.36	94.23	19.1	2.3	95.62	95.01
Penbased	55.57	2.91	95.38	94.25	55.70	2.92	95.07	93.89	152.8	2.8	97.04	96.04
Phoneme	3.87	2.35	80.09	79.13	8.00	2.08	81.79	80.53	17.8	2.2	83.52	82.14
Pima	4.23	1.75	80.04	74.93	8.77	2.35	82.17	76.01	22.7	2.4	82.9	75.66
Ring	8.27	1.50	84.82	83.72	14.07	1.65	92.54	91.52	24	1.9	95.13	94.08
Satimage	30.00	2.68	88.88	86.77	32.63	2.67	89.21	87.04	76.1	2.7	88.68	87.32
Sonar	4.73	2.74	88.98	79.09	6.50	2.69	92.70	78.81	18	2.3	98.77	80.19
Spambase	2.83	2.75	77.78	77.42	2.93	2.76	74.76	74.11	29.8	2.4	92.37	91.93
Spectfheart	3.50	1.00	79.40	79.42	3.50	1.00	79.40	79.42	12.9	2.8	91.43	79.83
Texture	21.20	2.75	93.59	92.37	21.67	2.71	94.45	93.08	54.5	2.7	93.71	92.89
Twonorm	14.87	2.96	97.06	96.06	14.87	2.96	97.06	96.06	60.9	2.6	96.64	95.28
Vowel	27.50	2.77	50.06	46.16	34.97	2.75	59.60	54.38	72.3	2.9	80.48	71.82
Wdbc	5.23	1.44	96.69	94.56	5.37	1.54	97.27	95.09	10.4	1.7	98.57	95.25
Wine	8.83	1.87	99.67	95.53	8.43	1.80	99.90	95.88	8.7	1.6	99.94	94.35
Yeast	12.17	2.53	58.69	56.34	15.77	2.53	58.24	56.11	35.2	2.6	63.81	58.5
MEANS	**13.96**	**2.28**	**85.23**	**80.43**	**15.71**	**2.26**	**86.42**	**81.40**	**37.98**	**2.36**	**91.14**	**84.31**

statistically equivalent and the version MO-FARCG-conf is to be preferred.

TABLE IV
COMPARISON ON TEST ACCURACY BETWEEN MO-FARCG-PROD AND MO-FARCG-CONF

Comparison (test accuracy)	**p value**	**Hypothesis**
MO-FARCG-prod vs MO-FARCG-conf	0.0084	Rejected

For this reason, this version of MO-FARCG has been chosen to be compared with the FARC-HD and the Wilcoxon's signed-ranks test has been applied again on the test classification percentage of the most accurate solutions. The null-hypothesis is rejected, since $p - value \leq /alpha$ (see Table V, hence the two methods are not statistically equivalents.

A further comparison has been drawn between the two algorithm with respect to the average number of rules of the most accurate solutions. Once more the null-hypothesis is rejected with $p-value = 1.35E-005$ and the two method are not statistically equivalent (Table V).

TABLE V
COMPARISON ON TEST ACCURACY AND COMPLEXITY BETWEEN FARC-HD AND MO-FARCG-CONF

Comparison (test accuracy)	**p value**	**Hypothesis**
FARC-HD vs MO-FARCG-conf	0.00754	Rejected

Comparison (complexity)	**p value**	**Hypothesis**
FARC-HD vs MO-FARCG-conf	1.35E-005	Rejected

By looking at the results reported in Table III, we can state that MO-FARCG-conf is outperformed by the original FARC-HD regarding the test accuracy, whereas the opposite is true when considering the complexity of the obtained systems. Nevertheless, against less than 3% loss on test accuracy, the complexity is reduced by more than 50%. This makes the models easier to understand but maintaining their accuracy at a similar level or even better for some datasets.

Ideally, when a multi-objective approach and a single-objective approach are applied to the same task, the solution set obtained by the former approach includes the solution obtained by the latter one. In the present study we cannot expect a similar result since the starting conditions are different for the multi-objective and the single-objective algorithms. Infact, FARC-HD uses an initial database generated by considering equidistributed granularities, while the initial database in MO-FARCG is constructed by considering the most promising granularities obtained in the initial steps of the algorithm. Therefore, the loss of accuracy and the complexity reduction are provoked by the initial granularities choice rather than by the application of the multi-objective algorithm.

VI. CONCLUSIONS

In this work, we have presented a study to investigate how the granularity learning affects the performance of a fuzzy associative classification algorithm, named FARC-HD. To this aim, a modification of the original algorithm has been created, named MO-FARCG. Two preliminary steps have been added: first a set of rules with multiple granularities is extracted and then a single granularity for each attribute is specified, depending on some measures performed on the extracted rules. Finally, the single-objective genetic algorithm included in the original FARC-HD has been replaced by a MOEA (that is a modification of SPEA2), in order to consider both accuracy and complexity of the obtained models.

After the analysis of results, we can conclude that the learning of the attributes' granularities combined with FARC-HD produces models with a slightly decreased accuracy, which is balanced by a considerable reduction of models' complexity.

REFERENCES

[1] J. Casillas, F. Herrera, O. Cordon, and L. Magdalena, *Interpretability issues in fuzzy modeling*. Secaucus, NJ, USA: Springer-Verlag, 2003.

[2] C. A. C. Coello, G. B. Lamont, and D. A. V. Veldhuizen, *Evolutionary Algorithms for Solving Multi-Objective Problems*. Secaucus, NJ, USA: Springer-Verlag, 2007.

[3] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*. New York, USA: John Wiley & Sons, Inc., 2001.

[4] R. Alcalá, M. J. Gacto, F. Herrera, and J. Alcalá-Fdez, "A multi-objective genetic algorithm for tuning and rule selection to obtain accurate and compact linguistic fuzzy rule-based systems," *Int. J. Uncertainty Fuzziness Knowledge Based Syst.*, vol. 15, no. 5, pp. 539–557, 2007.

[5] R. Alcalá, P. Ducange, F. Herrera, B. Lazzerini, and F. Marcelloni, "A multiobjective evolutionary approach to concurrently learn rule and data bases of linguistic fuzzy-rule-based systems," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 5, pp. 1106–1122, 2009.

[6] A. Botta, B. Lazzerini, F. Marcelloni, and D. C. Stefanescu, "Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index," *Soft Comput.*, vol. 13, no. 5, pp. 437–449, 2009.

[7] M. Cococcioni, P. Ducange, B. Lazzerini, and F. Marcelloni, "A Pareto-based multi-objective evolutionary approach to the identification of Mamdani fuzzy systems," *Soft Comput.*, vol. 11, no. 11, pp. 1013–1031, 2007.

[8] M. J. Gacto, R. Alcalá, and F. Herrera, "Adaptation and application of multi-objective evolutionary algorithms for rule reduction and parameter tuning of fuzzy rule-based systems," *Soft Comput.*, vol. 13, no. 5, pp. 419–436, 2009.

[9] ——, "Integration of an index to preserve the semantic interpretability in the multiobjective evolutionary rule selection and tuning of linguistic fuzzy systems," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 3, pp. 515–531, 2010.

[10] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, "Selecting fuzzy if-then rules for classification problems using genetic algorithms," *IEEE Trans. Fuzzy Syst.*, vol. 3, no. 3, pp. 260 –270, 1995.

[11] H. Ishibuchi, T. Murata, and I. B. Türksen, "Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems," *Fuzzy Sets Syst.*, vol. 89, no. 2, pp. 135–150, 1997.

[12] H. Ishibuchi and Y. Nojima, "Analysis of interpretability-accuracy trade-off of fuzzy systems by multiobjective fuzzy genetics-based machine learning," *Int. J. Approximate Reasoning*, vol. 44, no. 1, pp. 4–31, 2007.

[13] H. Ishibuchi and T. Yamamoto, "Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining," *Fuzzy Sets Syst.*, vol. 141, no. 1, pp. 59–88, 2004.

[14] P. Pulkkinen and H. Koivisto, "Fuzzy classifier identification using decision tree and multiobjective evolutionary algorithms," *Int. J. Approximate Reasoning*, vol. 48, no. 2, pp. 526–543, 2008.

[15] ——, "A dynamically constrained multiobjective genetic fuzzy system for regression problems," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 1, pp. 161–177, 2010.

[16] L. Sanchez, J. Otero, and I. Couso, "Obtaining linguistic fuzzy rule-based regression models from imprecise data with multiobjective genetic algorithms," *Soft Comput.*, vol. 13, no. 5, pp. 467–479, 2009.

[17] R. Alcalá, Y. Nojima, F. Herrera, and 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 Comput.*, vol. 15, pp. 2303–2318, 2011.

[18] J. Alcalá-Fdez, R. Alcalá, and F. Herrera, "A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 5, pp. 857–872, oct. 2011.

[19] H. Ishibuchi and T. Nakashima, "Effect of rule weights in fuzzy rule-based classification systems," *IEEE Trans. Fuzzy Syst.*, vol. 9, no. 4, pp. 506–515, 2001.

[20] H. Ishibuchi and T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 4, pp. 428–435, 2005.

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

[22] J. Han and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2006.

[23] C. Zhang and S. Zhang, *Association rule mining: models and algorithms*. Berlin: Springer-Verlag, 2002.

[24] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A. I. Verkamo, "Fast discovery of association rules," in *Advances in Knowledge Discovery and Data Mining*, 1996, pp. 307–328.

[25] H. Ishibuchi, T. Nakashima, and M. Nii, *Classification and modeling with linguistic information granules: advanced approaches advanced approaches to linguistic data mining*. London: Springer, 2004.

[26] M. Delgado, N. Marin, D. Sanchez, and M. A. Vila, "Fuzzy association rules: General model and applications," *IEEE Trans. Fuzzy Syst.*, vol. 11, no. 2, pp. 214–225, 2003.

[27] M. Kaya, "Multi-objective genetic algorithm based approaches for mining optimized fuzzy association rules," *Soft Comput.*, vol. 10, pp. 578–586, 2006.

[28] J. Alcalá-Fdez, R. Alcalá, M. J. Gacto, and F. Herrera, "Learning the membership function contexts for mining fuzzy association rules by using genetic algorithms," *Fuzzy Sets Syst.*, vol. 160, no. 7, pp. 905–921, 2009.

[29] Y.-C. Hu, R.-S. Chen, and G.-H. Tzeng, "Finding fuzzy classification rules using data mining techniques," *Pattern Recogn. Lett.*, vol. 24, pp. 509–519, 2003.

[30] F. P. Pach, A. Gyenesei, and J. Abonyi, "Compact fuzzy association rule-based classifier," *Expert Syst. Appl.*, vol. 34, pp. 2406–2416, 2008.

[31] Yi-Chung and Hu, "Determining membership functions and minimum fuzzy support in finding fuzzy association rules for classification problems," *Know. Based Syst.*, vol. 19, no. 1, pp. 57–66, 2006.

[32] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: improving the strength pareto evolutionary algorithm," Tech. Rep. 103, 2001.

[33] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, "Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms," *Fuzzy Sets Syst.*, vol. 65, no. 2-3, pp. 237–253, 1994.

[34] B. Kavšek, N. Lavrač, and V. Jovanoski, "APRIORI-SD: Adapting Association Rule Learning to Subgroup Discovery," in *Advances in Intelligent Data Analysis V*, M. R. Berthold, H.-J. Lenz, E. Bradley, R. Kruse, and C. Borgelt, Eds. Springer Berlin / Heidelberg, 2003, vol. 2810, pp. 230–241.

[35] L. J. Eshelman and J. D. Schaffer, "Real-coded genetic algorithms and interval-schemata," in *Foundation of Genetic Algorithms 2*, 1993, pp. 187–202.

[36] L. J. Eshelman, "The chc adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination." in *FOGA'90*, 1990, pp. 265–283.

[37] S. García and F. Herrera, "An Extension on Statistical Comparisons of Classifiers over Multiple Data Sets for all Pairwise Comparisons," *J. Mach. Learn. Res.*, vol. 9, pp. 2677–2694, 2008.

[38] S. García, A. Fernández, J. Luengo, and F. Herrera, "A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability," *Soft Comput.*, vol. 13, pp. 959–977, 2009.

[39] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, 2006.

[40] D. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures: Third Edition*. Boca Raton: Taylor & Francis Group, 2004.

[41] F. Wilcoxon, "Individual Comparisons by Ranking Methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.