Attributes regrouping in Fuzzy Rule Based Classification Systems: an intra-classes approach

Amel Borgi, Rim Kalaï, Hayfa Zgaya

▶ **To cite this version:**

Amel Borgi, Rim Kalaï, Hayfa Zgaya. Attributes regrouping in Fuzzy Rule Based Classification Systems: an intra-classes approach. In the 15th ACS/IEEE International Conference on Computer Systems and Applications AICCSA 2018, Oct 2018, Aqaba, Jordan. 10.1109/AICCSA.2018.8612802 . hal-02290131

HAL Id: hal-02290131
https://hal.science/hal-02290131

Submitted on 17 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Attributes regrouping in Fuzzy Rule Based Classification Systems: an intra-classes approach

Amel Borgi
Université de Tunis El Manar
Institut Supérieur d'Informatique
Faculté des Sciences de Tunis, LIPAH
Tunis, Tunisia
Amel.Borgi@insat.rnu.tn

Rim Kalai
Université de Tunis El Manar
Institut Supérieur d'Informatique
Faculté des Sciences de Tunis, LIPAH
Tunis, Tunisia
rym.kalai@gmail.com

Hayfa Zgaya
Faculty for Health engineering and management/ILIS
University Lille II
Lille, France
hayfa.zgaya@univ-lille2.fr

***Abstract*—Fuzzy rule-based classification systems (FRBCS) are able to build linguistic interpretable models, they automatically generate fuzzy if-then rules and use them to classify new observations. However, in these supervised learning systems, a high number of predictive attributes leads to an exponential increase of the number of generated rules. Moreover the antecedent conditions of the obtained rules are very large since they contain all the attributes that describe the examples. Therefore the accuracy of these systems as well as their interpretability degraded. To address this problem, we propose to use ensemble methods for FRBCS where the decisions of different classifiers are combined in order to form the final classification model. We are interested in particular in ensemble methods which split the attributes into subgroups and treat each subgroup separately. We propose to regroup attributes by correlation search among the training set elements that belongs to the same class, such an intra-classes correlation search allows to characterize each class separately. Several experiences were carried out on various data. The results show a reduction in the number of rules and of antecedents without altering accuracy, on the contrary classification rates are even improved.**

Keywords—Fuzzy rule based classification systems, supervised learning, automatic generation of fuzzy rules, ensemble learning methods, attributes regrouping, intra-classes correlation.

I. INTRODUCTION

One of the most used tools in supervised machine learning is Fuzzy Rule Based Classification Systems (FRBCS) [2, 8, 10, 14, 15, 29]. These systems are very powerful since, on the one hand, they provide an easily interpretable model composed by linguistic if-then rules, and on the other hand, they are able to deal with imprecise and noisy data.

The aim of a FRBCS is to construct a linguistic model which will be used to predict the class of new objects. In the literature, several methods have been proposed to generate fuzzy rules automatically from numerical data [8, 14, 29]. The big challenge of these systems is how to deal with high dimensional datasets. Indeed, a high number of predictive attributes leads to an exponential increase of the number of generated rules. Moreover it produces large premises, since they contain all the attributes that describes the observations. An important number of generated rules has a direct impact on training time as well as on storage capabilities It also has consequences on the transparency and the interpretability of the obtained results. Moreover it may affect the accuracy of the learning algorithms and their predictive capacity.

Thereby, reducing the number of rules and the number of antecedents without altering too much the classification performances appears as a key to improve FRBCS. To overcome this problem, several approaches have been proposed. A first possible approach is to select relevant rules. Many methods have been proposed in the literature to reduce rules number and to remove the useless. In [23], authors have suggested to reduce the number of rules by forgetting the weak ones. Other methods of rules selection based on genetic algorithm have been studied [16, 18, 36]. They explicitly consider a trade-off between the number of fuzzy if-then rules and the classification accuracy. Moreover, Multi-Objective Evolutionary Algorithms have also been used to design FRBS with a trade-off between the accuracy and the interpretability of the model [11, 30]. The second possible approach is feature selection which aim to remove the non-significant and redundant attributes. Indeed, since recent classification systems aim to deal with larger problems, the issue of feature selection is becoming increasingly important [6, 7, 9, 20, 25, 33, 35].

Another interesting solution is to decompose the set of attributes into subgroups and treat each subgroup separately. Based on the concept of ensemble learning methods, the learning problem which contains a big number of attributes is decomposed into sub-problems of lower complexity. Different classifiers are thus constructed with a different projection of the feature set. Each classifier generates his local rule base and the different rule bases are then combined to form the final classification model [3, 31].

The main issue of these methods is how to determine the attributes that will appear together in the same group, and consequently in the same if-then rules. In [31], a linear correlation search method was used to determine the related attributes: the attributes which are linearly correlated are grouped together and used by a classifier to construct the fuzzy if-then rules.

In this context, we study the attributes regrouping by correlation search among the training set elements that belong to the same class, it corresponds to an intra-classes correlation search. Thus, we propose a supervised learning method by automatic generation of fuzzy classification rules, the so called SIF-INTRA approach that allows reducing the number of rules without deteriorating classification accuracy, on the contrary it is even improved. In [31], the proposed approach called SIFCO is based on a correlation search among the components of all the training set elements, without any distinction. However, in our work, the linear correlation search is an intra-classes one: the training set elements are gathered together according to their class, and a linear correlation search is done among the components of the elements of each class considered separately. Thus, we may characterize each class and gather discriminating information for a further classification. As in [31], we

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

perform an ensemble methodology by combining several simple classification models: each model, here a FRBCS, uses a subset of the initial attributes, and solves a part of the original task. This is done in order to obtain a better composite global model, a set of classifiers, with more accurate and reliable decisions than a complex single model.

This paper is organized as follows. In section 2, we remind the classical process of FRBCS. Then, in section 3, we introduce the basic concepts of ensemble learning method and the SIFCO approach for regrouping attributes [31]. In section 4, we describe our proposed method of attributes regrouping by intra-classes correlation search. The tests and results obtained by computer simulations with different databases are provided in section 5. Finally, section 6 concludes this study.

II. Fuzzy Rule Based Classification Systems

A FRBCS is a supervised learning approach. It consists of two main phases: the learning phase and the classification phase. During the learning phase, fuzzy rules are automatically generated from the training data. In the classification phase, the above constructed rule base is used to classify new objects.

In the literature, several approaches have been proposed for automatically generating fuzzy if-then rules from numerical data [1, 10, 13, 14, 22]. The generation of fuzzy rules can be made by partitioning the training examples into fuzzy subsets by using membership functions and then constructing a fuzzy rule for each fuzzy grid [14-18]. This method is also known as fuzzy grid-type partition. Another way to generate fuzzy rules is the use of decision trees [13, 24]. Neural networks have also been used to construct neuro-fuzzy classification models [21].

In this section, we briefly describe the fuzzy grid-type partition approach that we use in this work [14, 17,18].

A. Learning phase with a simple fuzzy grid

We consider an n-dimensional classification problem, with m labelled examples $X^p= (X^p_1, X^p_2 \dots X^p_n)$, p=1, 2… m, the set of these examples is the training set denoted as TS. We denote the C classes as $y_1, y_2 \dots y_C$. X^p_i is the value of the i^{th} attribute of the p^{th} example. The generation of fuzzy if-then rules from numerical data includes two main phases [14, 17]. At first, a fuzzy partition of the pattern space into fuzzy subspaces is performed. So, each numerical attribute is partitioned into K fuzzy subsets $\{A_1, A_2, \dots A_K\}$, where each subset is defined by a membership function. Thus, we obtain a fuzzy grid (see Fig. 1). Then, a fuzzy if-then rule is constructed for each subspace of the fuzzy grid.

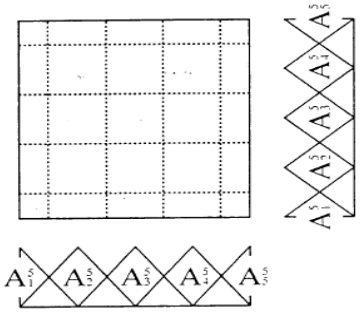

Fig. 1. Simple fuzzy grid

The fuzzy partition can be made by several approaches, such as the simple fuzzy grid [14], the efficient fuzzy partition [15] or the supervised fuzzy partition [31]. Fig. 1 represents an example of simple fuzzy grid where all the attributes are partitioned into the same number (K=5) of fuzzy subsets. Thus, the number of fuzzy rules generated is equal to 5^2. In general, if we consider an n-dimensional problem with K fuzzy subsets for the fuzzy partition, then we obtain K^n fuzzy rules. A triangular membership function is used in Fig. 1 [14-17]. Note that the membership function could have different shapes as triangular, trapezoidal, Gaussian, etc.

Let's consider a bi-dimensional pattern classification problem (with 2 attributes X_1 and X_2). A fuzzy rule, labeled R_{ij}^K and corresponding to the fuzzy subspace $A_i^K \times A_j^K$, is written as follows:

R_{ij}^K : If X_1 is A_i^K and X_2 is A_j^K then X= (X_1, X_2) belongs to y_{ij} with CF=CF_{ij} i=1… K and j=1… K

where X= (X_1, X_2) is a 2-dimentional pattern, K is the number of fuzzy subsets on each axis of the pattern space, y_{ij} corresponds to one among the C class labels, CF is the certainty factor which reflects the validity of the rule.

The conclusion and the certainty degree of each rule are determined as follows [18]:

1. For each class y_t (t=1,2,...,C), calculate the compatibility grade of each training example X^p belonging to this class, with the antecedent part of the fuzzy rule.

$$\beta_{yt} = \sum_{X^p \in yt} \mu_i^K(X_1^p) \times \mu_j^K(X_2^p) \quad (1)$$

where μ_i^k and μ_j^k are membership functions respective to A_i^K and A_j^K.

2. Find the class y_a that has the maximum value of the compatibility grade β_{yt} and assigns this class to y_{ij}

$$\beta_{ya} = \max\{\beta_{y1}, \beta_{y2}, \dots, \beta_{yC}\} \quad (2)$$

3. CF_{ij} is calculated as follows:

$$CF_{ij} = \frac{|\beta_{ya} - \beta|}{\sum_{t=1}^{C} \beta_{yt}} \text{ with } \beta = \sum_{yt \neq ya} \beta_{yt} / (C-1) \quad (3)$$

The rule base generated by the above procedure is denoted by S_R={R_{ij}^K , i=1,2,…,K and j=1,2,…,K}

B. Classification phase: fuzzy inference

In the classification phase, an inference system uses the rule base S_R generated in the learning phase to classify new objects. Let X'=(X'$_1$, X'$_2$) be a new observation. The steps, below, describe the inference method adopted in this paper [18].

1. For each class y_t, calculate α_{yt} (for t=1,2,…, C) as:

$$\alpha_{yt} = \max\{\mu_i^K(X_1') \times \mu_j^K(X_2') \times CF_{ij}^K / y_{ij} = y_t, R_{ij} \epsilon S_R\} \quad (4)$$

2. Find the class y_a which maximizes α_{yt}:

$$\alpha_{ya} = \max\{\alpha_{y1}, \alpha_{y2}, \dots, \alpha_{yC}\} \quad (5)$$

If two or more classes take the maximum value, then X' cannot be classified, else X' is assigned to y_a. Moreover, if there is no fuzzy rule with $\mu_i(X'_1)\times\mu_j(X'_1)\times CF_{ij}\neq 0$ the example X' cannot be classified.

The extension of this approach to n attributes is straightforward.

III. Regrouping Attributes in FRBCS

FRBCS with grid-type partition suffer from the explosion of the rules' number when the number of attributes is high. An interesting way to overcome this problem is the use of the regrouping attributes approach. It is based on the concept of ensemble learning machines.

Ensembles of learning machines constitute one of the main current themes in machine learning research, and have been applied to various real problems. They correspond to a set of supervised learning machines whose decisions are combined to make the final decision. Taking into account the opinions of several experts rather than one single opinion (which cannot always be the best one) can improve the performance of the overall system [27, 32]. Several algorithms based on this concept have been developed, like Bagging, Boosting and Stacking [4, 5, 12, 26, 27, 32]. The following steps describe the process of these methods:

- information, which may correspond to the examples, the attributes or the class variable, is distributed between different learners,
- each learner performs the learning phase using its own information,
- the decisions of the different learners are combined to make the final decision and classify a new observation.

In this paper, we focus on the regrouping attributes approach proposed in [31] in the context of FRBCS. This method, called SIFCO, combines different classifiers by giving to each one a different projection of the set of initial attributes. The idea is to consider different classifiers, and to give to each classifier a different projection of the training set elements. This is done by decomposing the initial set of attributes into subsets of correlated components; each classifier, here a FRBCS, uses a subset of attributes and generates his local rule base. Thus, the correlated attributes will appear in the same premise. Such an approach, inspired by Vernazza, allows considering the eventual discrimination power of the union of simultaneous selected features [4, 34]. Then the different rule bases are combined to form the final model or the global rule base. In opposition to a feature selection, the idea here is to preserve all attributes. The advantage, as mentioned in [33], is to exploit the redundancy of information in order to reduce noise on each variable. Another objective is to improve the prediction ensured by the diversity of built learners for each group [26].

As represented in Fig. 2, SIFCO method is composed of 3 phases [31]:

- a pre-treatment phase (attributes regrouping phase)
- a learning phase
- a classification phase

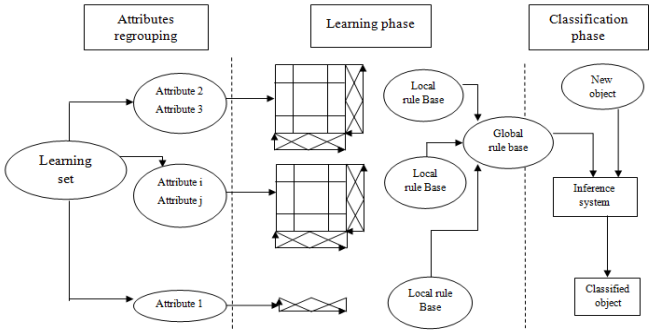

Fig. 2. Diagram describing the different steps of SIFCO [31]

For the pre-treatment phase, the authors proposed to regroup the attributes using linear correlation search among the training set elements [31]. Each group of high correlated attributes is treated separately by a FRBCS to generate a local rule base. During the learning phase the method described in section 2.1. was used to construct the local rule bases. In the classification phase the inference method described in section 2.2. was applied to classify a new object using the global rule base (union of all the local bases).

IV. Proposed Approach: Intra-Classes Attributes Regrouping in Fuzzy Rule Based Classification Systems

We have drawn on the work of [31] and we belief that the regrouping attributes approach proposed in [31] can be very interesting if the information about the class attribute is taken into account in the process of the determination of the related attributes.

For that purpose, we proceed to the attributes regrouping by correlation search among the training set elements that belong to the same class, it corresponds to an intra-classes correlation search. In [31], the authors have proposed to regroup attributes by using a correlation search among the components of all the training set elements, without any distinction. However, in our work, the linear correlation search is an intra-classes one: the training set elements are gathered together according to their class, and a linear correlation search is done among the components of the elements of each class considered separately. Thus, we may characterize each class and gather discriminating information for a further classification.

On the first step we rely on the class as a criterion for splitting the initial training set TS into several subsets. This step leads to a partitition of TS into C (total number of classes) training subsets denoted by TS_1, TS_2, ..., TS_C, such as TS_t contains all the training examples of the class y_t. The second step consists in computing, for each subset TS_t, the correlation matrix $R^t = (r^t_{i,j})_{n\times n}$, with n the number of attributes. We denote by $r^t_{i,j}$ the coefficient of linear correlation between the components X_i and X_j of the TS_t vectors (t=1,2,...,C).

$$R^t = \begin{bmatrix} 1 & r^t_{1,2} & \cdots & r^t_{1,n} \\ r^t_{2,1} & 1 & \cdots & \\ \cdots & & & \\ r^t_{1,n} & \cdots & \cdots & 1 \end{bmatrix} \quad where \quad t \in \{1,2,...,C\}$$

Next step in attributes regrouping consists of thresholding each matrix R_t with a prefixed threshold denoted θ, same threshold being used for all the matrices. As in [31], we use the following procedure to decide if two attributes X_i and X_j are correlated:

if $| r^t_{i,j} | < \theta$
then X_i and X_j are not correlated and $r^t_{i,j}$ is set to 0
else X_i and X_j are correlated and $r^t_{i,j}$ is set to 1

For each thresholded matrix $R_\theta{}^t$ obtained from R^t, we extract the largest subsets of correlated components, these subsets form a partition P^t of the components set $\{X_1, X_2, ..., X_n\}$. All the subsets, obtained from the different matrices, are regrouped by the union operator to obtain a final set of correlated components subsets. As a same correlated components subset can be found from two or more different training subsets, we use the union operator to eliminate such redundancies. It should be noted that this final set is not necessarily a partition of $\{X_1, X_2, ..., X_n\}$.

Each subset of correlated components will be treated separately by a FRBCS. Fig. 3 and Fig. 4 summarize these different steps of our approach SIF-INTRA.

In order to illustrate this procedure, let us consider an example of a 5-dimensional classification problem with 2 classes y_1 and y_2, and a training set TS containing m labeled patterns $X^p = (X^p_1, X^p_2 \ldots X^p_n)$, p=1, 2… m.

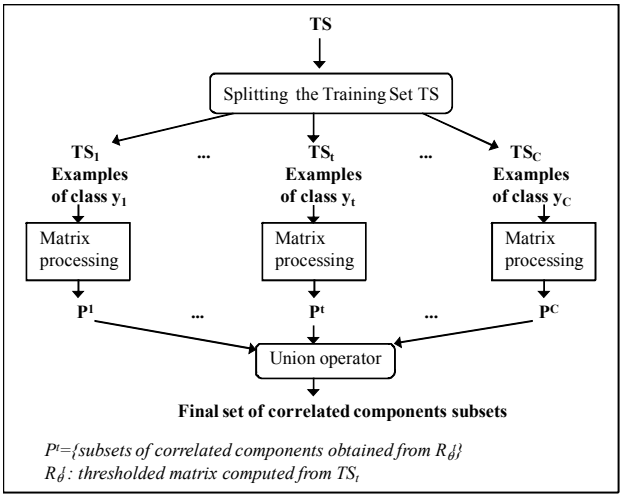

Fig. 3. Attributes regrouping steps in SIF-INTRA

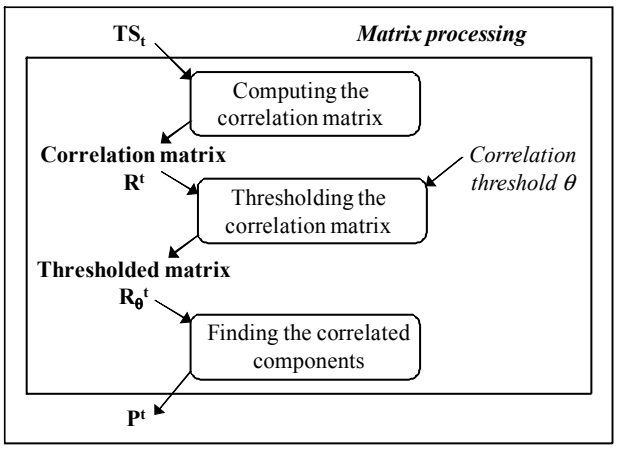

Fig. 4. Matrix processing in SIF-INTRA

TS is decomposed into 2 subsets: TS_1 containing the observations of class y_1 and TS_2 containing the observations of class y_2. Then we compute the correlation matrix R_1 among the examples of TS_1, and R_2 among the examples of TS_2. We obtain, for example, the following correlation matrices:

$$R^1 = \begin{bmatrix} 1 & 0.8 & 0.1 & 0.2 & 0.5 \\ 0.8 & 1 & 0.9 & 0.1 & 0.1 \\ 0.1 & 0.9 & 1 & 0.3 & 0.2 \\ 0.2 & 0.1 & 0.3 & 1 & 0.8 \\ 0.5 & 0.1 & 0.2 & 0.8 & 1 \end{bmatrix}$$

$$R^2 = \begin{bmatrix} 1 & 0.8 & 0.5 & 0.2 & 0.3 \\ 0.8 & 1 & 0.4 & 0.1 & 0.1 \\ 0.5 & 0.4 & 1 & 0.3 & 0.2 \\ 0.2 & 0.1 & 0.3 & 1 & 0.8 \\ 0.3 & 0.1 & 0.2 & 0.8 & 1 \end{bmatrix}$$

with a threshold θ=0.7, we obtain the following thresholded matrices:

$$R^1_\theta = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \qquad R^2_\theta = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

So, for the training subset TS_1, the partition P^1 of the initial set of attributes obtained from $R_\theta{}^1$ is: $P^1=\{\{X_1,X_2,X_3\}, \{X_4,X_5\}\}$, it contains the largest subsets of correlated components. For the training subset TS_2, the obtained partition P^2 of the initial set of attributes obtained from $R_2{}^1$ is: $P^2=\{\{X_1,X_2\},\{X_3\},\{X_4,X_5\}\}$. The final set of correlated attributes obtained through the union of all these subsets is: $\{\{X_1,X_2,X_3\}, \{X_4,X_5\},\{X_1,X_2\},\{X_3\}\}$.

Each of these four groups of attributes will be treated separately by a FRBCS, so the correlated components will be brought together in the premises of the rules. Thus the number of attributes present in the premises of the generated rules is smaller than the total number of attributes describing the learning examples. Thereby our method allows to reduce the size of the premises of the rules generated compared with FRBCS without grouping of attributes. Moreover, having less attributes in each group than the total number of attributes allows reducing the number of rules comparing to a single FRBCS that treats all the attributes. This is confirmed by the experimental tests.

In our example, and with simple fuzzy grids divided into K=5 fuzzy subsets, a single FRBCS handling all the attributes $\{X_1,X_2,X_3,X_4,X_5\}$ would lead to at most $5^5 = 3125$ rules. With a FRBCS treating each group of attributes we will obtain at most only 180 rules (125+25+25+5) as explained below:

- a FRBCS with $\{X_1,X_2,X_3\}$ will lead to $5^3 = 125$ rules
- a FRBCS with $\{X_4,X_5\}$ will lead to $5^2 = 25$ rules
- a FRBCS with $\{X_1,X_2\}$ will lead to $5^2 = 25$ rules
- a FRBCS with $\{X_3\}$ will lead to 5^1 rules

V. EXPERIMENTATION AND RESULTS

We implemented our intra-classes regrouping attributes approach in a so called SIF-INTRA system. As SIFCO, SIF-INTRA is composed of 3 phases: the pre-processing phase, the learning phase and the classification phase (see Fig. 2). In the pre-processing phase, we used our method of regrouping attributes based on an intra-classes correlation search (described in section 4): this point is different from the work of [31] as the authors used a correlation search among all the examples, without distinction of classes. The learning and the classification phases are kept the same as in [31].

We tested our system with well-known data bases that differ in terms of their attributes number, instances number and number of classes (Table 1). To evaluate the generalization performance of our method, we used the 10-fold cross-validation technique [19]. Different parameters values were tested, for the correlation threshold (θ): 0.5, 0.8, 0.9, 0.95, and for the fuzzy grid size (K): 3, 4, 5, 6.

Let us begin by analyzing the influence of the threshold, with a fuzzy grid size K=3. Experimental results are presented in Table 2, every cell contains the good classification rates followed in brackets by the number of generated rules, and in braces by the size of the premises. The best classifications rates are highlighted in bold. "Imp" refers to the impossibility of generating the fuzzy rule base due to the explosion of the rules' number (in our experiments, we note this problem from a number of rules about 10^5).

For Iris, Lupus and Wine data sets, the classification accuracy degrades or remains the same with the increase of the correlation threshold. For a high threshold, the number of subsets of correlated attributes is less than the one determined with a low threshold: the correlated attributes are those strongly correlated. Thus, we can lose information on the weak correlation of certain attributes. This is not the case for the Vehicle base: with a lower correlation threshold, information about the correlation of some discriminating attributes is lost and the classification rates are worse. To understand whether weak or strong correlations should be favoured a detailed analysis of the nature of data should be performed in a future work.

For the Vehicle base, we note the impossibility of generating a rule base with low correlation thresholds (0.5 and 0.8). This is due to a very large number of correlated attributes regrouped together.

We also notice that the number of generated rules (in brackets) decreases with the increase of the correlation threshold. This decrease is very important in the case of the Vehicle base: a difference of about 36116 between a threshold of 0.9 and a threshold of 0.95. This is due to the decrease in the number of correlated attributes in the subsets of correlated components found with a higher threshold.

TABLE 1. DESCRIPTION OF THE USED DATA SET

Data Base	Number of instances	Number of attributes	Number of classes
Iris	150	4	3
Lupus	87	3	2
Wine	178	13	3
Vehicle	846	18	4

TABLE 2. SIF-INTRA: INFLUENCE OF THE CORRELATION THRESHOLD

θ	0.5	0.8	0.9	0.95
Iris	**96.66** (97.5) {3.29}	95.33 (21.9) {1.44}	94.66 (12) {1}	94.66 (12) {1}
Lupus	**78.08** (11) {1.75}	**78.08** (11) {1.75}	**78.08** (11) {1.75}	**78.08** (11) {1.75}
Wine	**90.48** (1435.8) {4.72}	88.82 (44.7) {1.15}	88.82 (39) {1}	88.82 (39) {1}
Vehicle	Imp	Imp	52.24 (36652.8) {10.45}	**53.16** (536) {4.97}

For the two data sets Iris and Wine, we obtain the same results with the two different correlation thresholds 0.9 and 0.95, and for the Lupus data, the same results are obtained whatever the correlation threshold is. It means that these thresholds lead to the same groups of correlated attributes that is to say to the same grouping of attributes in the premises. The rule bases are then identical, hence the same classification results.

For the four data sets, the number of attributes (in braces) present in the premises of the generated rules is smaller than the total number of attributes describing the learning examples. Thus, our method allows to reduce the size of the premises of the rules generated compared with FRBCS without grouping of attributes.

Our method SIF-INTRA is strongly inspired by Fuzzy Rule Based Systems with simple fuzzy grid [14] denoted by SIF in Table 3, as well as SIFCO method [31]. Therefore, we give Table 3 as a comparative table between our method and these two approaches (rate of good classification followed in brackets by the rules' number). For each method, Table 3 shows the best results obtained with different values of the parameter K (3, 4, 5, 6).

For Iris and Lupus data sets, SIF-INTRA allows a slight improvement of classification results compared to SIF, the number of rules is also lower. Moreover SIF-INTRA is able to process the two datasets Wine and Vehicle that can't be dealt with SIF (referred by “Imp” in Table 3). The reason is an important number of attributes that lead to a huge number of rules.

Let us now compare the two linear correlation methods: the intra-classes correlation search ensured by SIF-INTRA and the correlation search among all classes without distinction processed by SIFCO. SIF-INTRA generally outperforms SIFCO method, except with Lupus data where SIF-INTRA leads to a small deterioration of classification rates. Nevertheless the classification rates are generally similar, except with Vehicle data where we note a significant improvement of 24% with an important reduction in the rules number when using SIF-INTRA. In general, the number of generated rules is less important with the intra-classes method than with the linear correlation search upon all classes, and this for the same fixed parameters values (except for Wine data set). Indeed, in the case of the intra-classes method, the subsets of correlated attributes used to build the premises are more numerous but smaller (smaller number of attributes in each subset) than those obtained by SIFCO. So the number and size of the rules are reduced with Iris, Lupus and Vehicle data.

TABLE 3. COMPARISON BETWEEN SIF, SIFCO AND SIF-INTRA

	SIF	SIFCO	SIF-INTRA
Iris	96 (46)	92.67 (30)	**96.66** (24)
Lupus	77.01 (11)	**79.31** (12)	78.08 (9)
Wine	Imp	93.26 (60)	**93.82** (3289.3)
Vehicle	Imp	43.5 (94514)	**54** (1459.7)

In the case of Wine data, we notice an important increase of the number of rules with SIF-INTRA: the obtained subsets of correlated attributes remain in this case too large. In all cases, it would be interesting to consider a post treatment that eliminates the neutral rules and optimizes the whole rule base.

A linear correlation search among the vectors of each class separately allows to access to more discriminating information: each class is characterized by its own matrix and correlated attributes. However, if the linear correlation is done among the attributes of all the training set vectors, without any distinction of classes, as in SIFCO, some discriminating information may be lost and classification rates are worse. Nevertheless, we have to note, in consideration to this last method (correlation search upon all classes), that its results remain satisfying, and that its complexity is weaker than the intra-classes method complexity (less matrices to treat and less rules to generate). As the attributes subsets are treated separately, one may use, in future work, distributed approaches for modelling the overall system and for parallelizing tasks when possible.

VI. CONCLUSION

In this paper, we propose a supervised regrouping attributes approach based on intra-classes correlation search. Our main goal is to take into account the class of the labelled data while detecting the correlations between the attributes. Each obtained group of attributes is then used by a classical FRBCS to generate a local rule base, thus, the correlated attributes will appear in the same premises. Then, the different rule bases are combined to form the final model.

The proposed approach reduces complexity compared to classical FRBCS: the number of rules and the number of antecedent conditions are reduced, and thus, comprehensibility is improved. Moreover our method handles fairly high-dimensional data (such as Wine and Vehicle), unlike classical FRBCS that cannot provide classification results. It was shown through computer simulations that this complexity reduction is done without deteriorating classification accuracy, on the contrary it is even improved.

These encouraging results lead us to complete computer simulations with other datasets and real data. We also plan to continue the experimentation tests on data bases with larger dimensionality. Moreover, it would be interesting to study more precisely the choicie of the threshold in order to find the relevant value for each data base.

In this paper, we propose an intra-classes correlation search, it would be interesting to study a different method of attributes regrouping to characterize a class versus all others, known as OAA (One-Against-All) or OVA (One-Vs-All) approaches [28, 37]. Another attractive perspective consists in reducing the number of rules in SIF-INTRA by introducing rules' selection methods.

As further work, one may also study other methods to find dependencies within data sets in order to regroup dependent attributes. We have already begun such a study by using Association Rules, more precisely a frequent itemsets mining algorithm was used to detect associations between the attributes [3]. In the same vein, it would be interesting to extend our approach SIF-INTRA to handle symbolic features. Indeed, in this work SIF-INTRA is limited to data with numerical features since it uses linear correlation to find dependencies between attributes. We have already started work in this direction using three methods to find associations between symbolic features: the chi-squared independence test, the Cramer V coefficient and the symmetrical uncertainty factor [38].

REFERENCES

[1] Angelov P, Lughofer E, Zhou X (2008). Evolving fuzzy classifiers using different model architectures. Fuzzy Sets Syst 159:3160–3182.

[2] M. Antonelli, P. Ducange, and F. Marcelloni (2014), A fast and efficient multi-objective evolutionary learning scheme for fuzzy rule-based classifiers. Information Sciences, vol. 283, pp. 36-54.

[3] Ben Slima I., Borgi A. (2015) Attributes regrouping by association rules in fuzzy inference systems. "Regroupement d'attributs par règles d'association dans les systèmes d'inférence floue". EGC 2015, vol. RNTI-E-28, pp. 317-328.

[4] J. Cao, H. Wang, S. Kwong, and K. Li, (2011). Combining interpretable fuzzy rule-based classifiers via multi-objective hierarchical evolutionary algorithm. In IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1771-1776.

[5] Breiman L (1996) Bagging predictors. Machine Learning 24(2):123–140.

[6] G. Biagetti, et al. (2016) Multivariate direction scoring for dimensionality reduction in classification problems. Smart Innovation, Systems and Technologies, 56, pp. 413-423.

[7] P. Crippa et al. (2015). Multi-class ECG beat classification based on a Gaussian mixture model of Karhunen-Loève transform. International Journal of Simulation: Systems, Science and Technology, 16 (1), pp. 2.1-2.10.

[8] O. Dehzangi, M. J. Zolghadri, S. Taheri, and S. M. Fakhrahmad. (2007). Efficient fuzzy generation: A new approach using data mining principles and rule weighting. Fuzzy Sys. and Knowledge Discovery, vol. 2, pp. 134–139.

[9] Dhir CS, Lee J, Lee S-Y (2011) Extraction of independent discriminant features for data with asymmetric distribution. Knowl Inf Syst. doi:10.1007/s10115-011-0381-9.

[10] Elkano M., Galar M., Sanz J., Bustince H.(2016) Fuzzy Rule-Based Classification Systems for multi-class problems using binary decomposition strategies: On the influence of n-dimensional overlap functions in the Fuzzy Reasoning Method. Information Sciences, Volume 332, 1, Pages 94-114.

[11] M. Fazzolari, R. Alcalá, and F. Herrera (2014). A multi-objective evolutionary method for learning granularities based on fuzzy discretization to improve the accuracy-complexity trade-off of fuzzy rule-based classification systems: D-MOFARC algorithm. Applied Soft Computing, vol. 24, pp. 470-481.

[12] Y. Freund (1995). Boosting a weak learning algorithm by majority. Information and Computation, 121(2):256–285.

[13] Hefny, H. A., Ghiduk, A. S., Wahab, A. A., & Elashiry, M. (2010). Effective Method for Extracting Rules from Fuzzy Decision Trees based on Ambiguity and Classifiability. Universal Journal of Computer Science and Engineering Technology 1 (1), 55-63, Oct. 2010.

[14] Ishibuchi H, Nozaki K, Tanaka H (1992) Distributed representation of fuzzy rules and its application to pattern classification. Fuzzy Sets Syst 52:21–32.

[15] Ishibuchi H, Nozaki K, Tanaka H (1993) Efficient fuzzy partition of pattern space for classification problems. Fuzzy Sets Syst 59:295–304.

[16] Ishibuchi H, Yamamoto T (2004) Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining. Fuzzy Sets Syst 141(1):59–88.

[17] Ishibuchi, H., & Yamamoto, T. (2005). Rule weight specification in fuzzy rule-based classification systems. IEEE transactions on fuzzy systems, 13(4), 428-435.

[18] Ishibuchi H, Nojima Y (2007) Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning. Int J Approx Reason 44(1):4–31.

[19] Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th international joint conference on artificial intelligence (2), Canada.

[20] Lee HM, Chen C-M, Chen J-M et al (2001) An efficient fuzzy classifier with feature selection based on fuzzy entropy. IEEE Trans Syst Man Cybern Part B Cybern 31(3):426–432.

[21] Mitra, S., & Hayashi, Y. (2000). Neuro-fuzzy rule generation: survey in soft computing framework. IEEE transactions on neural networks, 11(3), 748-768.

[22] T. T. Nguyen, A. W. C. Liew, C. To, X. C. Pham, and M. P. Nguyen. (2014). Fuzzy If-Then rules classifier on ensemble data. In International Conference on Machine Learning and Cybernetics, pp. 362-370, Springer, 2014, July.

[23] Nozaki K, Ishibuchi H, Tanaka H (1994) Selecting fuzzy rules with forgetting in fuzzy classification systems. In: 3rd IEEE international conference on fuzzy systems (1), pp 618–623.

[24] Pal, N. R., & Chakraborty, S. (2001). Fuzzy rule extraction from ID3-type decision trees for real data. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 31(5), 745-754.

[25] P. D. Phong (2015). An application of feature selection for the fuzzy rule based classifier design based on an enlarged hedge algebras for high-dimensional datasets Journal of Science and Technology, vol. 53, no. 5, pp. 583-597.

[26] Prudhomme E, Lallich S (2008) Maps ensemble for semi-supervised learning of large high dimensional dataset. In: 19th international symposium ISMIS 2008, LNAI 4994, Springer, Berlin, pp 100–110.

[27] Y. Ren, L. Zhang, & P.N. Suganthan (2016) Ensemble classification and regression-recent developments, applications and future directions. IEEE Computational Intelligence Magazine, vol. 11, no. 1, pp. 41-53.

[28] Rifkin R., and Klautau A (2004). In Defense of One-Vs-All Classification. Journal of Machine Learning Research, 2004, pp. 101–141.

[29] Riza L. S., Bergmeir C., Herrera F. and Benitez J. M. (2015) frbs: Fuzzy Rule-Based Systems for Classification and Regression in R. Journal of Statistical Software, May 2015, Volume 65, Issue 6.

[30] F. Rudziński (2016). A multi-objective genetic optimization of interpretability-oriented fuzzy rule-based classifiers. Applied Soft Computing, vol. 38, pp. 118-133.

[31] Soua, B., A. Borgi, and M. Tagina (2013) An ensemble method for fuzzy rule-based classification systems. Knowledge and Information Systems 36 (2): 385–410, DOI 10.1007/s10115-012-0532-7.

[32] Valentini G,Masulli F (2002) Ensembles of learning machines. In: Marinaro M, Tagliaferri R (eds) Neuralnets WIRN Vietri-02, LNCS 2486, Springer, Berlin, pp 3–19.

[33] Verleysen M, François D, Simon G et al (2003). On the effects of dimensionality on data analysis with neural networks. Int. work-conf. on artificial and natural neural networks, LNCS 2687. Springer, pp 105–112.

[34] Vernazza G (1993) Image classification by extended certainty factors. In: Pattern recognition 26(11). Pergamon Press Ltd, Oxford, pp 1683–1694.

[35] Villacampa O. (2015) Feature Selection and Classification Methods for Decision Making: A Comparative Analysis. Doctoral dissertation, Nova Southeastern University. Retrieved from NSUWorks, College of Engineering and Computing.

[36] P. Villar, A. Fernández, and F. Herrera (2010). A genetic algorithm for feature selection and granularity learning in fuzzy rule-based classification systems for highly imbalanced data-set. Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Methods, pp. 741-750.

[37] Yang L., Rui W. and Zeng Y. (2007) An Improvement of One-against-all Method for Multi-class Support Vector Machine. IEEE Machine Learning and Cybernetics, Hong Kong. DOI: 10.1109/ICMLC.2007.4370646.

[38] A. Borgi, D. Lahbib (2013). Ensemble learning methods and attribute grouping for rule generation "*Ensembles d'apprentissage et regroupement d'attributs pour la génération de règles"*. 8th International Conference on Intelligent Systems: Theories and Applications, SITA'2013, 08-09 May 2013, Rabat, Morocco, pp. 214-221 (co-sponsored by IEEE).