Analysing the Hierarchical Fuzzy Rule Based Classification Systems with Genetic Rule Selection

A. Fernández
Department of Computer Science
and Artificial Intelligence, CITIC-UGR
University of Granada, Spain
Email: alberto@decsai.ugr.es

M. J. del Jesus
Department of Computer Science
University of Jaén, Spain
Email: mjjesus@ujaen.es

F. Herrera
Department of Computer Science
and Artificial Intelligence, CITIC-UGR
University of Granada, Spain
Email: herrera@decsai.ugr.es

Abstract—This contribution is focused on the enhancement of the precision for Fuzzy Rule Based Classification Systems by the refinement of the Knowledge Base. Specifically, we make use of a Hierarchical Fuzzy Rule Based Classification System, which consists in the application of a thicker granularity in order to generate the initial Rule Base, and to reinforce those problem subspaces that are specially difficult by means of the application of rules with a higher granularity. Furthermore, we will perform a genetic rule selection process in order to obtain a compact and accurate model.

Our experimental results show the goodness of this approach, especially when the number of classes is high, which usually implies a higher difficulty in the separability of the examples. Our conclusions are supported by means of the corresponding statistical tests.

I. INTRODUCTION

Linguistic Fuzzy Rule Based Classification Systems (FR-BCSs) [1] are a very useful tool in the field of Data Mining. They provide an accurate model which is also easily interpretable by the end-user or expert by means of the use of linguistic labels. The main handicap in the application of linguistic systems is the hard restrictions on the fuzzy rule structure,which may suppose a loss in accuracy when dealing with some complex systems, i.e. high dimensional problems, when the classes are overlapped or in the presence of noise.

In this contribution, we propose the use of a hierarchical environment in order to improve the behaviour of linguistic FRBCSs [2]. This approach preserves the original descriptive power and increases its accuracy by reinforcing those problem subspaces that are specially difficult. Therefore, we focus our efforts in enhancing the classification performance in the boundary areas of the problem, obtaining a good separability among the classes.

We consider the modification of the Knowledge Base (KB) structure using the concept of "layers" that was introduced in [3], defined by the authors as Hierarchical Knowledge Base (HKB). In order to obtain a Hierarchical Fuzzy Rule Base Classification System (HFRBCS), we will apply a two-level learning methodology:

1) A Linguistic Rule Generation (LRG) method is used to create the initial Rule Base (RB), from which we extract the Hierarchical Rule Base (HRB).

2) A Genetic Algorithm (GA) is employed to select the best cooperative rules from the HRB.

To obtain the initial linguistic fuzzy models, we will employ a simple inductive LRG-method, the Chi et al.'s method [4], using triangular membership functions for the fuzzy partitions and rule weights in the consequent of the rules [5].

Our objective is to analyse the behaviour of the HFRBCS by contrasting its performance with the one obtained by the simple Chi algorithm with different granularity levels in standard classification. We have selected 15 data-sets from UCI repository [6] within the experimental framework. The measure of performance is based on accuracy rate and the significance of results is supported by the proper statistical analysis as suggested in the literature [7], [8].

This contribution is set up as follows. Section II we give a brief introduction to linguistic FRBCS and the description of the fuzzy algorithm selected for our study. Then, in Section III, we describe the concept of a HFRBCS. In Section IV we present the two-level methodology to automatically design an HFRBCS from a generic LRG-method. In Section V we include our experimental analysis. Finally, in Section VI some concluding remarks are pointed out.

II. FUZZY RULE BASED CLASSIFICATION SYSTEMS AND LINGUISTIC RULE GENERATION METHOD

Any classification problem consists of m training patterns $x_p = (x_{p1}, \ldots, x_{pn})$, $p = 1, 2, \ldots, m$ from M classes where x_{pi} is the ith attribute value ($i = 1, 2, \ldots, n$) of the p-th training pattern.

In this work we use fuzzy rules of the following form for our FRBCSs:

$$\text{Rule } R_j : \text{ If } x_1 \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn} \\ \text{ then Class} = C_j \text{ with } RW_j \quad (1)$$

where R_j is the label of the jth rule, $x = (x_1, \ldots, x_n)$ is an n-dimensional pattern vector, A_{ji} is an antecedent fuzzy set, C_j is a class label, and RW_j is the rule weight [9].

Fuzzy learning methods are the basis to build a FRBCS. The algorithm used in this work is the method proposed in [4], that we have called the Chi et al.'s rule generation. To generate the fuzzy RB this FRBCSs design method determines the relationship between the variables of the problem and

establishes an association between the space of the features and the space of the classes by means of the following steps:

1) *Establishment of the linguistic partitions.* Once the domain of variation of each feature A_i is determined, the fuzzy partitions are computed.
2) *Generation of a fuzzy rule for each example $x_p = (x_{p1}, \ldots, x_{pn}, C_p)$.* To do this is necessary:
   2.1 To compute the matching degree $\mu(x_p)$ of the example to the different fuzzy regions using a conjunction operator.
   2.2 To assign the example x_p to the fuzzy region with the greatest membership degree.
   2.3 To generate a rule for the example, whose antecedent is determined by the selected fuzzy region and whose consequent is the label of class of the example.
   2.4 To compute the rule weight.

Rules with the same antecedent can be generated during the learning. If they have the same class in the consequent we remove one of the duplicated rules, but if it is different, only the rule with the highest weight is kept in the RB.

III. HIERARCHICAL FUZZY RULE BASED CLASSIFICATION SYSTEM

The HFRBCS approach [3] presents a more flexible KB structure that allows to improve the accuracy of the FRBCSs without losing their interpretability: the HKB, which is composed of a Hierarchical Data Base (HDB) and an HRB.

The HKB is composed of a set of layers, and each layer is defined by its components in the following way:

$$layer(t, n(t)) = DB(t, n(t)) + RB(t, n(t)), \qquad (2)$$

with $n(t)$ being the number of linguistic terms in the fuzzy partitions of layer t, $DB(t, n(t))$ being the Data Base (DB) which contains the linguistic partitions with granularity level $n(t)$ of layer t (t-linguistic partitions), and $RB(t, n(t))$ being the RB formed by those linguistic rules whose linguistic variables take values in $DB(t, n(t))$ (t-linguistic rules). For the sake of simplicity in the descriptions, the following notation equivalences are established:

$$DB(t, n(t)) \equiv DB^t \text{ and } RB(t, n(t)) \equiv RB^t. \qquad (3)$$

We must point out that we are using *linguistic partitions* with the same number of linguistic terms for all input variables, composed of symmetrical triangular-shaped and uniformly distributed membership functions (see Fig. 1). The number of linguistic terms in the t-linguistic partitions is defined in the following way:

$$n(t) = (n(1) - 1) \cdot 2^{t-1} + 1 \ , \qquad (4)$$

with $n(1)$ being the granularity of the initial fuzzy partitions.

Figure 1 (left) graphically depicts the way in which a linguistic partition in DB^1 becomes a linguistic partition in DB^2. Each term of order k from DB^t, $S_k^{n(t)}$ ($S_k^{n(1)}$ in the

figure), is mapped into the fuzzy set $S_{2k-1}^{2 \cdot n(t)-1}$, preserving the former modal points, and a set of *n(t)-1* new terms is created, each one between $S_k^{n(t)}$ and $S_{k+1}^{n(t)}$ ($k = 1, ..., n(t) - 1$) (see Figure 1 right).
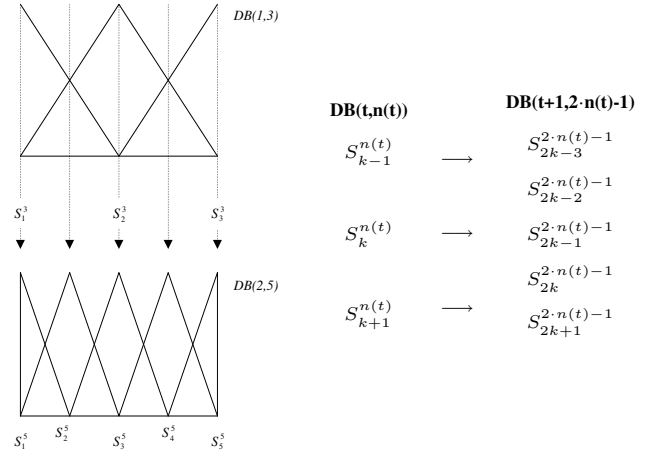

Fig. 1. Two layers of linguistic partitions which compose the HDB and mapping between terms from successive DBs.

The main purpose of developing an HRB is to divide the problem space in a more accurate way. To do so, those linguistic rules from $RB(t, n(t))$ –RB^t– that classify a subspace with bad performance are expanded into a set of more specific linguistic rules, which become their image in $RB(t + 1, 2 \cdot n(t) - 1)$ –RB^{t+1}–. This set of rules classify the same subspace that the former one and replaces it. As a consequence of the previous definitions, we could now define the HKB as the union of every layer t:

$$HKB = \cup_t layer(t, n(t)). \qquad (5)$$

In this contribution, we will just consider a two-layer HKB which allows us to produce a refinement of simple FRBCS to increase their accuracy, preserving their structure and descriptive power, and reinforcing only the classification of those problem subspaces with more difficulties by a hierarchical treatment of the rules generated in these zones.

IV. TWO-LEVEL LEARNING METHOD FOR BUILDING HFRBCSs

In this section we will describe the algorithm to obtain an HFRBCS, which is based on a two-stage methodology:

1) HKB Generation Process: An HRB is created from a simple RB obtained by an LRG-method.
2) HRB Genetic Selection Process: The best cooperative rules are selected by means of a GA.

A. Hierarchical Knowledge Base Generation Process

We use an existing inductive LRG-method and a previously defined DB^1. Specifically, as we state in Section II, we consider as LRG-method the Chi et al. [4] approach, that will

lead us to obtain simple linguistic fuzzy models, although any other technique could be used.

Two measures of error are used in the algorithm: a global measure, which is used to evaluate the complete RB, and a local measure, used to determine if an individual rule is expanded. Their expressions are defined below:

1) **Global measure**. We will employ the accuracy per class, computed as:

$$Acc_i\left(X_i, RB\right) = \frac{|\{x_p \in X_i \ / \ FRM(x_p, RB) = Class(x_p)\}|}{|X_i|} \tag{6}$$

where $| \cdot |$ is the number of patterns, with X_i being the subset of examples of the i-th class ($i \in 1...M$), $FRM(x_p, RB)$ is the output class computed following the Fuzzy Reasoning Method (FRM) using the current RB and $Class(x_p)$ is the class label for example x_p.

2) **Local measure**. The accuracy for a simple rule, $R_j^{n(1)}$, calculated over X, is shown as follows:

$$Acc\left(X, R_j^{n(1)}\right) = \frac{|X^+(R_j^{n(1)})|}{|X(R_j^{n(1)})|} \tag{7}$$

$$X^+(R_j^{n(1)}) = \{x_p \in X \ / \ \mu_{R_j^{n(1)}}(x_p) > 0 \\ \text{and } Class(x_p) = Class(R_j^{n(1)})\} \tag{8}$$

$$X(R_j^{n(1)}) = \{x_p \in X \ / \ \mu_{R_j^{n(1)}}(x_p) > 0\} \tag{9}$$

where $Class(\cdot)$ is a function that provides the class label for a pattern, or for a rule. We must note that $X^+(R_j^{n(1)})$ and $X(R_j^{n(1)})$ only include those examples that the rule actually classifies, since we will use as FRM the winning rule approach.

TABLE I
TWO-LEVEL LEARNING METHOD

HIERARCHICAL KNOWLEDGE BASE GENERATION PROCESS
    Step 0. *RB(1,n(1)) Generation Process*
    Step 1. *RB(2,2·n(1)-1) Generation Process*
    Step 2. *Summarization Process*
HIERARCHICAL RULE BASE GENETIC SELECTION PROCESS
    Step 3. *HRB Genetic Selection Process*

Now we will describe the HKB generation process (summarized in Table I), which basically consists of the following steps:

Step.0 RB^1 **Generation.** Generate the rules from DB^1 by means of an existing LRG-method: $RB^1 = LRG\text{-}method(DB^1, X)$.

Step.1 RB^2 **Generation**. Generate RB^2 from RB^1, DB^1 and DB^2.

  a) *Calculate the global error of RB^1 per class: $Acc_i\left(X_i, RB^1\right), i = 1, \ldots, M$.*

  b) *Calculate the local error of each 1-linguistic rule: $Acc\left(X, R_j^{n(1)}\right)$.*

  c) *Select the 1-linguistic rules with bad performance which will be expanded (the expansion factor α*

may be adapted in order to have more or less expanded rules):

$$R'(i,j) = \begin{cases} RB_{bad}^1, & \text{If } Acc\left(X, R_j^{n(1)}\right) \leq \\ & (1-\alpha) \cdot Acc_i\left(X_i, RB^1\right) \\ RB_{good}^1, & \text{otherwise.} \end{cases} \tag{10}$$

where $Class(R_j^{n(1)}) = i$.

  d) *Create DB^2.*

  e) *For each bad performance 1-linguistic rule to be expanded, $R_j^{n(1)} \in RB_{bad}$:*

    i) *Select the 2-linguistic partitions terms from DB^2 for each rule. For all linguistic terms considered in $R_j^{n(1)}$, i.e., $S_{jk}^{n(1)}$ defined in DB^1, select those terms $S_h^{2 \cdot n(1)-1}$ in DB^2 that significantly intersect them. We consider that two linguistic terms have a "significant intersection" between each other, if the maximum cross level between their fuzzy sets in a linguistic partition overcomes a predefined threshold δ:*

$$I(S_{jk}^{n(1)}) = \{S_h^{2 \cdot n(1)-1} \in DB^2 \ / \max_{u \in U_k} \\ \min\{\mu_{S_{jk}^{n(1)}}(u), \mu_{S_h^{2 \cdot n(1)-1}}(u)\} \geq \delta\} \tag{11}$$

where $\delta \in [0, 1]$.

    ii) *Combine the previously selected s sets $I(S_{jk}^{n(1)})$ by the following expression:*

$$I(R_j^{n(1)}) = I(S_{j1}^{n(1)}) \times \cdots \times I(S_{js}^{n(1)}) \tag{12}$$

    iii) *Extract 2-linguistic rules, which are the expansion of the bad 1-linguistic rule $R_j^{n(1)}$. This task is performed by the LRG-method, which takes $I(R_j^{n(1)})$ and the set of examples $X(R_j^{n(1)})$ as its parameters:*

$$CLR(R_j^{n(1)}) = LRG\text{-}method(I(R_j^{n(1)}), X(R_j^{n(1)})) \\ = \{R_{j_1}^{2 \cdot n(1)-1}, ..., R_{j_L}^{2 \cdot n(1)-1}\} \tag{13}$$

with $CLR(R_j^{n(1)})$ being the image of the expanded linguistic rule $R_j^{n(1)}$, i.e., the candidates to be in the HRB from rule $R_j^{n(1)}$.

Step.2 **Summarization.** *Obtain a Joined set of Candidate Linguistic Rules (JCLR), performing the union of the group of the new generated 2-linguistic rules and the former good performance 1-linguistic rules:*

$$JCLR = RB_{good}^1 \cup (\cup_j CLR(R_j^{n(1)})), \ R_j^{n(1)} \in RB_{bad}^1.$$

Step.3 **HRB Selection**. *Simplify the set JCLR by removing the unnecessary rules from it and generating an HRB with good cooperation. In JCLR −where rules of different hierarchical layers coexist−, it may happen*

that a complete set of *2-linguistic rules* which replaces an expanded *1-linguistic rule* does not produce good results. However, a subset of this set of *2-linguistic rules* may work properly. A genetic process is considered to put this task into effect, which is explained on detail in the next subsection.

B. Hierarchical Rule Base Genetic Rule Selection Process

In the previous part of this section we have mentioned that an excessive number of rules may not produce a good performance and it makes difficult to understand the model behaviour. We may find different types of rules in a large fuzzy rule set: irrelevant rules, which do not contain significant information; redundant rules, whose actions are covered by other rules; erroneous rules, which are wrong defined and distort the performance of the FRBCS; and conflicting rules, which perturb the performance of the FRBCS when they coexist with others.

In this work, we consider the CHC genetic model [10] in order to make the rule selection process, since it has achieved good results for binary selection problems [11]. In the following, the main characteristics of this genetic approach are presented.

1) *Coding Scheme and Initial Gene Pool:* It is based on a binary coded GA where each gene indicates whether a rule is selected or not (alleles '1' or '0' respectively). Considering that N rules are contained in the preliminary/ candidate rule set, the chromosome $C = (c_1, \ldots, c_N)$ represents a subset of rules composing the final HRB, such that:

$$\text{If } c_i = 1 \text{ Then } (R_i \in \text{ HRB }) \text{ Else } (R_i \notin \text{ HRB }),$$

with R_i being the corresponding i-th rule in the candidate rule set and HRB being the final hierarchical rule base.

The initial pool is obtained with an individual having all genes with value '1' and the remaining individuals generated at random in $\{0, 1\}$, so that the initial HRB is taking into account in the genetic selection process.

2) *Chromosome Evaluation:* The fitness function is simply the accuracy rate of the HRB.

3) *Crossover Operator:* The half uniform crossover scheme (HUX) is employed. An incest prevention mechanism is considered in order to apply this operator, i.e., two parents are crossed if their hamming distance divided by 2 is higher than a predetermined threshold, L. The threshold value is initialized as: $L = (\#Genes/4.0)$ and decremented by one when the population does not change in one generation.

4) *Restarting approach:* To get away from local optima, this algorithm uses a restart approach. In this case, the best chromosome is maintained and the remaining are generated at random in $\{1,0\}$. The restart procedure is applied when the threshold value is reached, which means that all the individuals coexisting in the population are very similar. We will stop the genetic process if more than 3 restarts are performed without including any new chromosome in the population

V. EXPERIMENTAL STUDY

In this section we will first present our experimental framework, providing details of the real-world problems chosen for the experimentation, the configuration parameters and the statistical tests applied to compare the results obtained along the experimental study. Then, we will present our empirical results and we will extract the conclusions derived from the performance obtained by the HFRBCS.

A. Experimental Framework

Table II summarizes the properties of the selected data-sets. It shows, for each data-set, the number of examples (#Ex.), the number of attributes (#Atts.), and the number of classes (#Cl.). The *magic*, *page-blocks*, *penbased* and *ringnorm* data-sets have been stratified sampled at 10% in order to reduce their size for training. In the case of missing values (*wisconsin*) we have removed those instances from the data-set.

To develop the different experiments we consider a *10-folder cross-validation model*, i.e., 10 random partitions of data with a 10%, and the combination of 9 of them (90%) as training and the remaining one as test. For each data-set we consider the average results of the ten partitions.

TABLE II
SUMMARY DESCRIPTION OF THE DATA-SETS

id	Data-set	#Ex.	#Atts.	#Cl.
bup	bupa (liver disorder)	345	6	2
eco	ecoli	336	7	8
gla	glass identification	214	9	7
iri	iris	150	4	3
mag	magic	1,902	10	2
new	new-thyroid	215	5	3
pag	page-blocks	548	10	5
pen	pen-based recognition	1,099	16	10
pim	pima	768	8	2
rin	ringnorm	740	20	2
shu	shuttle	2,175	9	7
veh	vehicle	846	18	4
win	wine	178	13	3
wis	wiscosin	683	9	2
yea	yeast	1,484	8	10

We will use the following configuration for the FRBCS approach: product T-norm as conjunction operator, together with the Penalized Certainty Factor heuristic [5] for the rule weight and the winning rule approach for the FRM. Furthermore, we selected the following values for the parameters in the learning method for building HFRBCSs:

- Rule Generation:
  - δ,*n(t+1)-linguistic partition* terms selector: 0.1
  - α, used to decide the expansion of the rule: 0.2
- GA Selection:
  - Number of evaluations: 10,000
  - Population length: 61

Finally, we must point out that in this contribution, we use the hypothesis testing techniques to provide statistical

support to the analysis of the results [12], [13]. Specifically, we will use non-parametric tests, due to the fact that the initial conditions that guarantee the reliability of the parametric tests may not be satisfied, making the statistical analysis to lose credibility with these type of tests [7], [8]. Specifically, we use the Iman-Davenport test [13] to detect statistical differences among a group of results and the Holm post-hoc test [14] in order to find which algorithms reject the hypothesis of equality with respect to a selected control method.

Furthermore, we consider the average ranking of the algorithms in order to show graphically how good a method is with respect to its partners. This ranking is obtained by assigning a position to each algorithm depending on its performance for each data-set. The algorithm which achieves the best accuracy on a specific data-set will have the first ranking (value 1); then, the algorithm with the second best accuracy is assigned rank 2, and so forth. This task is carried out for all data-sets and finally an average ranking is computed as the mean value of all rankings.

B. Empirical Analysis

Table III shows the performance results, ordered incrementally by the number of classes, for the three methods used in the experimentation, the Chi et al.'s method with 3 and 5 labels and the HFRBCS. We can observe that in almost all data-sets the HFRBCS obtains better results, which means that this approach benefits both the learning phase of the FRBCS (improvement in train) and the generalisation of the output model (improvement in test).

TABLE III
ACCURACY RESULTS FOR CHI WITH 3 AND 5 LABELS AND THE HFRBCS

		Chi3		Chi5		HFRBCS	
Data-Set	#Cl.	Train	Test	Train	Test	Train	Test
bup	2	60.03	57.87	73.60	59.42	84.24	62.59
mag	2	75.91	75.24	82.89	76.82	85.66	78.55
pim	2	75.72	72.40	86.12	69.80	91.23	72.14
rin	2	59.44	52.57	98.72	56.76	99.73	92.97
wis	2	98.09	91.21	100.0	66.77	100.0	91.07
iri	3	93.73	94.00	96.49	95.33	94.93	92.67
new	3	85.87	84.24	92.83	91.26	97.61	94.46
win	3	98.74	92.68	100.0	75.95	99.94	91.60
veh	4	65.87	60.77	88.37	64.06	93.66	67.26
pag	5	92.60	91.61	94.98	92.70	95.14	92.70
gla	7	65.86	60.04	74.63	58.71	82.05	61.73
shu	7	80.27	80.23	84.08	84.00	84.08	84.05
eco	8	79.46	78.33	90.74	84.53	93.26	87.19
pen	10	98.60	94.82	100.0	78.36	99.94	95.00
yea	10	29.63	29.18	59.34	56.34	63.99	57.48
Mean	X	77.32	74.35	88.19	74.05	91.03	81.43

In order to validate our results, we show the ranking of the different models by means of the procedure described in subsection V-A. Figure 2 shows the average ranking computed for the three different alternatives: the two basic approaches with 3 and 5 labels and the HFRBCS. We can observe that the HFRBCS is the best option, as the basic FRBCSs obtains the worst ranking with a much higher value than the former.

Next, we perform a Iman-Davenport test for detecting significant differences among the results of these approaches.
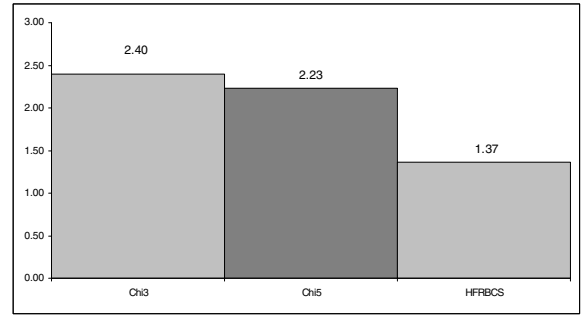

Fig. 2. Ranking in accuracy for the Chi et al.'s method with 3 and and the HFRBCS approach

The associated p-value is equal to 0.0058, which implies that there are significant differences among the results and thus, we should apply a post-hoc test, in this case the Holm test. We want to compare the HFRBCS versus the FRBCS with 3 and 5 labels.

The result of this test is shown in Table IV, in which the algorithms are ordered with respect to the z value obtained. Thus, by using the normal distribution, we can obtain the corresponding p-value associated with each comparison and this can be compared with the associated α/i in the same row of the table to show whether the associated hypothesis of equal behaviour is rejected in favour of each corresponding algorithm against the control method or not.

TABLE IV
HOLM TEST FOR THE HFRBCS AND FRBCS (CHI WITH 3 AND 5 LABELS). HFRBCS IS THE CONTROL METHOD

i	algorithm	z	p	α/i	Hypothesis ($\alpha = 0.05$)
2	Chi3	2.82990	0.00466	0.025	Rejected for HFRBCS
1	Chi5	2.37346	0.01762	0.05	Rejected for HFRBCS

We observe from the results of this test that the HFRBCS outperforms the basic Chi et al.'s method for both granularity levels. We can conclude then that the use of a hierarchical system approach derives in a benefit in performance for linguistic FRBCSs.

Finally, we present the number of rules obtained with each method in Table V. We can observe that the number of rules increases when a fine granularity is applied. In the case of the HFRBCS we have also a higher number of rules in the second level of granularity (5 labels per variable) and for this reason there are a loss in the interpretability in the use of the two-partitions scheme.

In order to detect whether this high number of rules is caused by the complexity of the problems or to the effectiveness of the genetic optimisation process, we show in Table VI the number of rules before and after the genetic rule selection process, together with the percentage of reduction both globally and in each granularity level. We observe that in average we remove approximately a 30% of the total rules, independently of their granularity. Furthermore, in Table VII we also show the accuracy obtained before and after the genetic

rule selection, observing that the percentage of enhancement in train is quite low due to the fact that in some cases the initial accuracy is above the 90% and the generalisation capability (test improvement) is above 1.5 points in only a third part of the cases. This issues allow us to direct our future work to develop new mechanisms for the genetic rule selection that will enable the final KB to have a lower complexity together with a higher precision.

VI. CONCLUDING REMARKS

In this work we have analysed a hierarchical methodology for FRBCSs in order to increase the precision of the output model. Our empirical results have shown that the use of a two-level HKB derives on a benefit in the performance since we apply a fine granularity specifically on those areas of the problem in which the classes are difficult to discriminate. We have observed that the HFRBCS have a good behaviour especially when the number of classes is high (4 or more classes).

Our future work will be oriented to analyse the use of the hierarchical model over a wider benchmark of data-sets and with other LRG-methods. We will also focus on the reduction of the length of the final HRB and the integration of the genetic rule selection with a genetic tuning of the HDB in order to obtain a higher improvement of the fuzzy system.

TABLE V
AVERAGE #RULES FOR CHI WITH 3 AND 5 LABELS AND HFRBCS

Data-Set	#Cl.	Chi3	Chi5	HFRBCS All	3labels	5labels
bup	2	43.3	111.4	120.0	12.7	107.3
mag	2	153.0	599.1	782.6	56.9	725.7
pim	2	105.2	401.5	531.0	45.9	485.1
rin	2	89.9	590.0	633.4	47.0	586.4
wis	2	224.0	296.6	613.4	216.3	397.1
iri	3	16.5	41.2	18.4	14.0	4.4
new	3	18.4	45.1	38.0	8.2	29.8
win	3	120.2	159.4	144.9	74.7	70.2
veh	4	227.8	640.4	2041.5	118.7	1922.8
pag	5	19.6	53.0	39.4	12.9	26.5
gla	7	27.1	75.0	83.3	15.4	67.9
shu	7	6.8	21.7	31.2	5.8	25.4
eco	8	43.5	112.8	116.6	19.8	96.8
pen	10	685.8	950.7	1307.0	510.4	796.6
yea	10	25.4	137.9	216.0	9.9	206.1
Mean	X	120.4	282.4	447.8	77.9	369.9

ACKNOWLEDGMENT

REFERENCES

[1] H. Ishibuchi, T. Nakashima, and M. Nii, *Classification and modeling with linguistic information granules: Advanced approaches to linguistic Data Mining*. Springer-Verlag, 2004.

[2] A. Fernández, M. J. del Jesus, and F. Herrera, "Hierarchical fuzzy rule based classification systems with genetic rule selection for imbalanced data-sets," *International Journal of Approximate Reasoning*, vol. 50, pp. 561–577, 2009.

[3] O. Cordón, F. Herrera, and I. Zwir, "Linguistic modeling by hierarchical systems of linguistic rules," *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 1, pp. 2–20, 2002.

TABLE VI
#RULES GENERATED BY THE HFRBCS WITHOUT GENETIC RULE SELECTION (HFRBCS NO GS) AND WITH RULE SELECTION (HFRBCS)

Data-Set	#Cl.	HFRBCS No GS Total	3labels	5labels	HFRBCS Total	3labels	5labels	% Reduction Total	3labels	5labels
bup	2	222.3	25.0	197.3	120.0	12.7	107.3	46.02	49.20	45.62
mag	2	1322.2	112.6	1209.6	782.6	56.9	725.7	40.81	49.47	40.00
pim	2	841.2	77.4	763.8	531.0	45.9	485.1	36.88	40.70	36.49
rin	2	1011.7	79.2	932.5	633.4	47.0	586.4	37.39	40.66	37.12
wis	2	613.4	216.3	397.1	613.4	216.3	397.1	0.00	0.00	0.00
iri	3	24.6	16.1	8.5	18.4	14.0	4.4	25.20	13.04	48.24
new	3	67.5	14.0	53.5	38.0	8.2	29.8	43.70	41.43	44.30
win	3	241.4	119.1	122.3	144.9	74.7	70.2	39.98	37.28	42.60
veh	4	2840.7	176.6	2664.1	2041.5	118.7	1922.8	28.13	32.79	27.83
pag	5	62.7	18.8	43.9	39.4	12.9	26.5	37.16	31.38	39.64
gla	7	131.5	21.0	110.5	83.3	15.4	67.9	36.65	26.67	38.55
shu	7	31.2	5.8	25.4	31.2	5.8	25.4	0.00	0.00	0.00
eco	8	195.3	36.4	158.9	116.6	19.8	96.8	40.30	45.60	39.08
pen	10	1745.2	675.3	1069.9	1307.0	510.4	796.6	25.11	24.42	25.54
yea	10	369.6	16.2	353.4	216.0	9.9	206.1	41.56	38.89	41.68
Mean	X	648.03	107.32	540.71	447.78	77.91	369.87	30.90	27.41	31.60

TABLE VII
ACCURACY OBTAINED BY THE HFRBCS WITHOUT GENETIC RULE SELECTION (HFRBCS NO GS) AND WITH RULE SELECTION (HFRBCS)

Data-Set	#Cl.	HFRBCS No GS Train	Test	HFRBCS Train	Test	% Enhancement Train	Test
bup	2	74.51	61.13	84.24	62.59	13.06	2.39
mag	2	82.06	78.13	85.66	78.55	4.38	0.54
pim	2	85.31	71.63	91.23	72.14	6.95	0.72
rin	2	97.65	90.00	99.73	92.97	2.13	3.30
wis	2	100.00	91.07	100.00	91.07	0.00	0.00
iri	3	94.33	95.33	94.93	92.67	0.63	-2.80
new	3	91.38	89.85	97.61	94.46	6.82	5.13
win	3	99.43	93.24	99.94	91.60	0.51	-1.75
veh	4	88.93	65.96	93.66	67.26	5.32	1.97
pag	5	93.99	92.70	95.14	92.70	1.23	0.00
gla	7	75.78	60.58	82.05	61.73	8.27	1.90
shu	7	84.08	84.05	84.08	84.05	0.00	0.00
eco	8	90.68	86.34	93.26	87.19	2.85	0.99
pen	10	99.84	95.73	99.94	95.00	0.10	-0.76
yea	10	59.70	56.54	63.99	57.48	7.19	1.67
Mean	X	87.84	80.82	91.03	81.43	3.63	0.76

[4] Z. Chi, H. Yan, and T. Pham, *Fuzzy algorithms with applications to image processing and pattern recognition*. World Scientific, 1996.

[5] H. Ishibuchi and T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems," *IEEE Transactions on Fuzzy Systems*, vol. 13, pp. 428–435, 2005.

[6] A. Asuncion and D. Newman, "UCI machine learning repository," 2007. [Online]. Available: http://www.ics.uci.edu/~mlearn/MLRepository.html

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

[8] 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. 2607–2624, 2008.

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

[10] L. J. Eshelman, *Foundations of Genetic Algorithms*. Morgan Kaufman, 1991, ch. The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination, pp. 265–283.

[11] J. R. Cano, F. Herrera, and M. Lozano, "Using evolutionary algorithms as instance selection for data reduction in KDD: An experimental study," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 6, pp. 561–575, 2003.

[12] 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 Computing*, vol. 13, no. 10, pp. 959–977, 2009.

[13] D. Sheskin, *Handbook of parametric and nonparametric statistical procedures*, 2nd ed. Chapman & Hall/CRC, 2006.

[14] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, vol. 6, pp. 65–70, 1979.