

An Extraction Method for the Characterization of the Fuzzy Rule Based Classification Systems' Behavior using Data Complexity Measures: A case of study with FH-GBML

Julián Luengo and Francisco Herrera

Abstract—When dealing with problems using Fuzzy Rule Based Classification Systems it is difficult to know in advance whether the model will perform well or badly.

In this work we present an automatic extraction method to determine the domains of competence of Fuzzy Rule Based Classification Systems. As a case of study we use the Fuzzy Hybrid Genetic Based Machine Learning method. We consider twelve metrics of data complexity in order to analyze the behavior patterns of this method, obtaining intervals of such data complexity measures with good or bad performance of it. Combining these intervals we obtain rules that describe both good or bad behaviors of the Fuzzy Rule Based Classification System mentioned.

These rules allow describe both good or bad behaviors of the Fuzzy Rule Based Classification Systems mentioned, allowing us to characterize the response quality of the methods from the data set complexity metrics of a given data set. Thus, we can establish the domains of competence of the Fuzzy Rule Based Classification Systems considered, making it possible to establish when the method will perform well or badly prior to its application.

I. INTRODUCTION

Fuzzy Rule Based Classification Systems (FRBCSs) [12], [15] are widely employed due to their capability to build a linguistic model interpretable to the users and the possibility of mixing different information such as that coming from expert knowledge and information coming from mathematical models or empiric measures. There is a vast literature in the field of FRBCSs, which is very active at this time. New FRBCS models have been proposed on standard classification [10], [11], [17] and data streams [2] among others. They have also been applied widely including, but not limited to, the detection of intrusions [20], medical applications [1], [21], [19] and in the imbalanced data framework [7].

On the other hand, the prediction capabilities of classifiers are often strongly dependent on the problem's characteristics. An emergent field has arisen that uses a set of complexity measures [4] applied to the problem to describe their difficulties [4], [9]. These measures quantify particular aspects of the problem which are considered relevant to the classification task [4]. Studies of data complexity metrics applied to

J. Luengo is with the Department of Computer Science and Artificial Intelligence, University of Granada, Granada, 18071, Granada, Spain. e-mail: julianlm@decsai.ugr.es.

F. Herrera is with the Department of Computer Science and Artificial Intelligence, University of Granada, Granada, 18071, Granada, Spain. e-mail: herrera@decsai.ugr.es.

particular classification algorithms can be found in [6], [5], [8], [18].

The complexity in the data can be used for characterizing FRBCSs' performance and it can be considered a new trend in the use of FRBCSs in pattern recognition. No data complexity metrics have been analyzed together with FRBCSs up to now, except in our previous study [16].

In this work we propose a novel automatic extraction method for characterizing FRBCSs by means of data complexity measures. In particular we consider twelve types of data complexity measures based on the overlaps in feature values from different classes; separability of classes; and measures of geometry, topology, and density of manifolds.

In order to perform this study, we consider the Fuzzy Hybrid Genetic Based Machine Learning (FH-GBML) method proposed by Ishibuchi et al. [14]. We have created 438 binary classification data sets from real world problems, and computed the value of the twelve metrics proposed by Ho and Basu [9].

The automatic extraction method of the domains of competence consider the following steps:

- It obtains intervals that describe when the FRBCS perform well or badly attending to the data complexity values.
- It formulates one rule for each interval where some information and conclusions about the behavior of these methods can be stated.
- It combines the individual rules in order to improve their support and interpretability.
- Finally, one rule which discriminates the good or bad behavior of the FRBCS is obtained.

The intervals which describe the performance are based on the following average values:

- Classification ratio, considering the average accuracy and test rate, and its difference with the global behavior.
- Detection of the over-learning, by means of the distance between the training accuracy and test accuracy ratio.

The rest of this contribution is organized as follows. In Section II the description of the FH-GBML method and the considered complexity measures are introduced. Section III describes the automatic method proposed in this work. In Section IV we include the experimental framework and the analysis of the obtained rules with the automatic method along with their analysis. Finally, in Section V some concluding remarks are pointed out.

II. PRELIMINARIES

In this section we introduce the FRBCS used and the parameters used in Subsection II-A. The data complexity metrics considered in this contribution are described in Subsection II-B.

A. FH-GBML Description and Parameters

The FH-GBML FRBCS is a Pittsburgh method in which each rule set is an individual. Besides it contains a cooperative-competitive learning approach (on individual represents one rule), which is used as an heuristic mutation in order to partially modify each rule set.

This method uses standard fuzzy rules with weights [13] in which each input variable x_i is represented by one linguistic term or label. The system define 14 possible linguistic terms for each attribute, and a special “don’t care” set.

In the learning process, N_{pop} rule sets are created by selecting randomly N_{rule} training examples. Then a fuzzy rule is generated from each training pattern by means of choosing one fuzzy set as antecedent randomly from the 14 candidates ($P(B_k) = \frac{\mu_{B_k}(x_{pi})}{\sum_{j=1}^{14} \mu_{B_j}(x_{pi})}$); and each antecedent (fuzzy set) of the generated fuzzy rule is replaced by *don’t care* using a fixed probability $P_{don't\ care}$.

$N_{pop} - 1$ rule sets are generated by the selection, crossover and mutation mechanisms as it is usually performed in Pittsburgh approaches. Next, with fixed probability, one single iteration of the cooperative-competitive algorithm is applied to each generated rule set.

Finally the best rule set is added, in the current population, to the new generated rule sets ($N_{pop} - 1$) in order to create the new population. Then if the stop criterion is not satisfied, the genetic process is repeated again. The winning rule fuzzy reasoning method is used in order to classify the examples.

We have used fixed parameters due to the necessity of analyze the characterization of FH-GBML independently to the procedure of adjusting of them, and therefore this task is out of the scope of this contribution. However, we have performed several experiment and adjusted the parameters empirically to obtain a good performance. In Table I we have summarized the parameters.

TABLE I

PARAMETERS USED BY FH-GBML (p IS THE NUMBER OF ATTRIBUTES IN THE DATA SET)

Number of fuzzy rules: $5 \times p$ rules.
Number of rule sets (N_{pop}): 200 rule sets.
Crossover probability: 0.9.
Mutation probability: $1/p$.
Number of replaced rules: All rules except the best-one (Pittsburgh-part, elitist approach), number of rules/5 (Michigan-part).
Total number of generations: 1,000 generations.
Don’t care probability: 0.5.
Probability of the application of the Michigan iteration: 0.5

B. Data Complexity Measures

In our analysis we have used the 12 data complexity measures proposed by Ho and Basu [9] that quantify difficult characteristics of the data sets for the classification task. They are summarized in Table II, and a deeper description can be found in [4].

TABLE II

COMPLEXITY METRICS USED IN THIS STUDY

Type	Id.	Description
Measures of Overlaps in Feature Values from Different Classes	F1	maximum Fisher’s discriminant ratio
	F2	volume of overlap region
	F3	maximum (individual) feature efficiency
Measures of Separability of Classes	L1	minimized sum of error distance by linear programming
	L2	error rate of linear classifier by Linear Programming
	N1	fraction of points on class boundary
	N2	ratio of average intra/inter class NN distance
Measures of Geometry, Topology and Density of Manifolds	N3	error rate of 1NN classifier
	L3	nonlinearity of linear classifier by linear programming
	N4	non-linearity of 1NN classifier
	T1	fraction of points with associated adherence subsets retained
	T2	average number of points per dimension

III. AUTOMATIC EXTRACTION METHOD

In this section the motivation behind the proposal is presented in Subsection III-A. In Subsection III-B the definition of the good or bad behavior points and intervals is given. The method is described in Subsection III-C.

A. Problem Analysis

To determine when a method will perform good or bad is not a trivial task, considering the accuracy as performance measure. One primary indicative of the method’s performance is the training accuracy. However, this is not always a precise measure. Figure 1 contains the accuracy results in training and test for FH-GBML over all the 438 data sets, plotted in ascending training accuracy value. We would

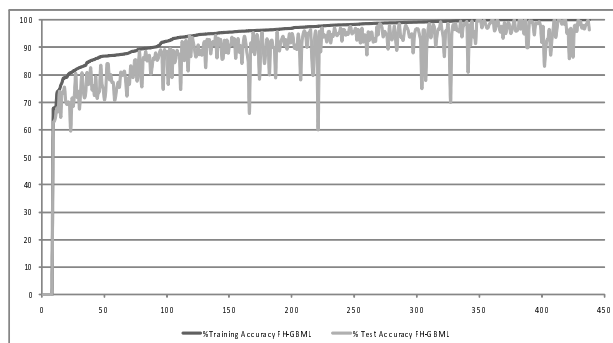


Fig. 1. Accuracy in Training/Test for FH-GBML sorted by training accuracy

like to point out how over-fitting is continuously present in them. Therefore the necessity of other kinds of tools for characterizing the behavior of the methods appears.

The data complexity measures presented in the previous section can be used to carry out the characterization of the FRBCSs. One direct approach is to analyze the relationship between the data complexity value for a given data set and the performance obtained by the method. Using enough data sets sorted by a particular data complexity measure, we could observe regions in which the method is performing noticeably well or badly. In Figure 2 the FH-GBML method’s results sorted by the N1 data complexity measure can be easily separated between good and bad results. However, this case does not usually appear when considering all the data

complexity measures. Figure 3 shows an example of a data complexity measure in which no significant regions could be found neither for good or bad behavior for the FH-GBML method.

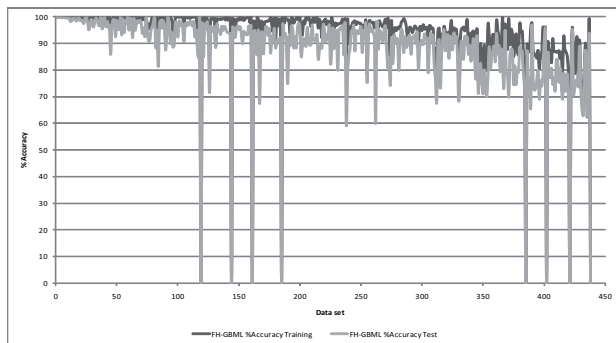


Fig. 2. Accuracy in Training/Test for FH-GBML sorted by N1

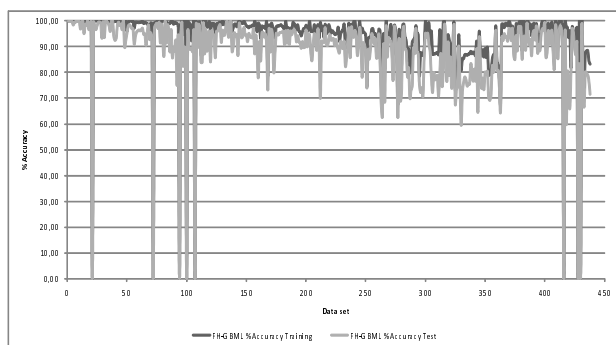


Fig. 3. Accuracy in Training/Test for FH-GBML sorted by T1

In the study performed in [16] an ad-hoc method used for extracting intervals for the FH-GBML was proposed. The intervals were extracted over the sorted data sets as described above. This ad-hoc method was based on the selection of intervals of data complexity metrics' values with some interest for the user according to the graphical window of the good results. With the combination of the manually extracted intervals, the characterization of the behavior of the FH-GBML for the considered data sets was achieved.

Two main problems were found when dealing with the ad-hoc process of extracting intervals in [16]:

- The cut points which define the intervals were arbitrary selected according to the graphics.
- It is possible to omit intervals with similar characteristics to the extracted ones. That is, the user is not using a formal description of the bad or good behavior intervals.

These issues can be tackled by the rigorous definition of the good and bad intervals, and creating an automatic search method to finding them.

B. Good or Bad Behavior Definitions

In order to overcome the issues indicated in the previous subsection, we propose an automatic method that can

extract the intervals in which the learning method shows a prominent behavior. The automatic method decides which data complexity measures are useful as they contain interesting intervals, and which measures are discarded (without providing any interval for them).

The automatic method analyzes a list of data sets sorted by a particular data complexity measure, where each data set has an associated performance measure of the FRBCSs (typically the training and test accuracy rates). Then it extracts intervals of values of such data complexity measure. These intervals correspond to the given definition of good and bad behavior of the FRBCSs performance.

Definition 1: Let $U = \{u_1, u_2, \dots, u_n\}$ be a list of paired training and test accuracy values for n different data sets, where u_i^{tra} is the training accuracy value associated to the data set u_i and u_i^{tst} is the test accuracy value associated to the data set u_i . Each data set u_i has associated a particular value of every data complexity measure considered.

Definition 2: Given a list of paired training and test accuracy values $U = \{u_1, u_2, \dots, u_n\}$, we define the **average training accuracy** over U as $\bar{U}^{tra} = \frac{1}{n} \sum_{i=1}^n u_i^{tra}$ and the **average test accuracy** as $\bar{U}^{tst} = \frac{1}{n} \sum_{i=1}^n u_i^{tst}$.

Definition 3: Given a list of paired training and test accuracy values $U = \{u_1, u_2, \dots, u_n\}$, we define an **interval** $V = \{u_i, \dots, u_j\} \subseteq U$. The lower and upper bound values of V are defined as:

- $M_{low}(V) = \min_{v_k \in V} \{V\}$.
- $M_{up}(V) = \max_{v_k \in V} \{V\}$.

In our proposal, we distinguish between two kinds of good and bad behavior elements.

- Based on quality points, which take into account the presence of overlearning.
- Based on intervals V , which take into account differences in their average accuracy \bar{V} with respect to the global one \bar{U} .

Both kind of elements are defined next.

Definition 4: A **good behavior point** u_i is such that

- 1) $u_i^{tra} - u_i^{tst} \leq overLearningDifference$; and
- 2) $u_i^{tra} \geq minGoodTraining$.

Definition 5: A **bad behavior point** u_i is such that

- 1) $u_i^{tra} - u_i^{tst} > overLearningDifference$; or
- 2) $\bar{U}^{tra} - u_i^{tra} \geq minBadGlobalDifference$.

Definition 6: An **interval of good behavior** $V = \{u_i, \dots, u_j\}$ is such that

- 1) $\bar{V}^{tra} - \bar{U}^{tra} \geq intTrainDifference$; and
- 2) $\bar{V}^{tst} - \bar{U}^{tst} \geq intTestDifference$.

Definition 7: An **interval of bad behavior** $V = \{u_i, \dots, u_j\}$ is such that

- 1) $\bar{U}^{tra} - \bar{V}^{tra} \geq minBadGlobalDifference$; or
- 2) $(\bar{V}^{tra} - \bar{V}^{tst}) - (\bar{U}^{tra} - \bar{U}^{tst}) \geq overLearningDifference$; or
- 3) $\bar{U}^{tst} - \bar{V}^{tst} \geq intTestDifference$.

The definitions of points and intervals of good or bad behavior are parameterized, and they can be adjusted by the user. The meaning and adjusted experimental values of the parameters used are shown next:

- *overLearningDifference* = 5. It describes the difference limit between training and test accuracy in order to consider the presence of overlearning.
- *minGoodTraining* = 90. It defines the minimum training accuracy of a point in order to consider it a good behavior point.
- *minBadGlobalDifference* = 10. It defines the minimum difference with the global training accuracy in order to be a bad behavior point.
- *intTrainDifference* = 3. This parameter indicates the minimum difference with the global training accuracy needed to consider a good behavior interval.
- *intTestDifference* = 6. This parameter indicates the minimum difference with the global test accuracy needed to consider a good behavior interval.

C. Automatic Extraction Method Description

For each FRBCSs considered, we obtain a list of paired training and test accuracy values U . With the previous definitions, the automatic method is capable of extracting a series of intervals. In order to do that, it is necessary that the data sets in U are sorted by the data complexity measures. That is, for a given data complexity measure CM , $\forall i < j; CM(u_i) \leq CM(u_j); u_i, u_j \in U; i, j \in \{1, \dots, n\}$.

We present the main outline of the automatic method in Algorithm 1. The functions present in it are defined as follows:

- *nextImportantGoodPoint*(u_i, U): Looks for the index of the next *good behavior point* u_k in the subset $V = \{u_i, \dots, u_n\} \subseteq U$. If no good behavior point could be found it returns -1 .
- *nextImportantBadPoint*(u_i, U): Looks for the index of the next *bad behavior point* u_k in the subset $V = \{u_i, \dots, u_n\} \subseteq U$. If no bad behavior point could be found it returns -1 .
- *extendGoodInterval*(pos, U): From the reference point u_{pos} this method creates a new *interval of good behavior* $V = \{u_l, \dots, u_m\} \subseteq U$ from the initial interval such that $u_{pos} \in V$.
- *extendBadInterval*(pos, U): From the reference point u_{pos} this method creates a new *interval of bad behavior* $V = \{u_l, \dots, u_m\} \subseteq U$ from the initial interval such that $u_{pos} \in V$.
- *mergeOverlappedIntervals*(A): In this function, an interval V_k is dropped from A if $\exists V_m \in A; M_{up}(V_m) \geq M_{up}(V_k)$ and $M_{low}(V_m) \leq M_{low}(V_k)$. Moreover it tries to merge overlapped intervals or intervals separated by a maximum gap of 5 points, provided that the new merged intervals satisfies the previous definitions of good or bad behavior.

IV. EXPERIMENTAL STUDY

In this section the analysis of the automatic extraction method is presented. This study begins with the intervals obtained by the automatic extraction method for the three FRBCSs for the initial 438 data sets considered. From

Algorithm 1 Automatic Method

Input: A set of performance results in training and test of the learning method $U = \{u_1, u_2, \dots, u_n\}$ sorted by a particular data complexity measure

Output: A set of intervals G in which the learning method shows good or behavior, and a set of intervals B where the learning method shows bad behavior

Steps:

$i \leftarrow 1$

$G \leftarrow \{\}$

$B \leftarrow \{\}$

while $i < n$ **do**

$pos \leftarrow nextImportantGoodPoint(u_i, U)$

if $pos \neq -1$ **then**

$interval \leftarrow extendGoodInterval(pos, U)$

$G \leftarrow G \cup \{interval\}$

$u_i \leftarrow M_{up}(interval)$

end if

end while

$i \leftarrow 1$

while $i < n$ **do**

$pos \leftarrow nextImportantBadPoint(u_i, U)$

if $pos \neq -1$ **then**

$interval \leftarrow extendBadInterval(pos, U)$

$B \leftarrow B \cup \{interval\}$

$u_i \leftarrow M_{up}(interval)$

end if

end while

$G \leftarrow mergeOverlappedIntervals(G)$

$B \leftarrow mergeOverlappedIntervals(B)$

return $\{G, B\}$

these intervals we construct several rules that model the performance of the used FRBCSs.

In order to perform this analysis, we have divided this section into the following two subsections.

- 1) Data sets generation for the experimental study in Subsection IV-A.
- 2) Extraction of individual rules based on the FRBCSs' behavior in Subsection IV-B.
- 3) Analysis of the disjunctive rules and their conjunctive combination in Subsection IV-C.

A. Data Sets Generation

We evaluate the FH-GBML method on a set of 438 binary classification problems. These problems are generated from pairwise combinations of the classes of 21 problems from the University of California, Irvine (UCI) repository [3]. These are *iris*, *wine*, *new-thyroid*, *solar-flare*, *led7digit*, *zoo*, *yeast*, *tae*, *balanced*, *car*, *contraceptive*, *ecoli*, *hayes-roth*, *shuttle*, *australian*, *pima*, *monks*, *bupa*, *glass*, *haberman* and *vehicle*.

In order to do that, first we take each data set and extract the examples belonging to each class. Then we construct a new data set with the combination of the examples from two different classes. This will result in a new data set with only 2 classes and the examples which have two such classes as output. We perform this process for every possible pairwise

combination of classes. However, if an obtained data set with this procedure proves to be linearly-separable, we discard it (since we could classify it with a linear classifier with no error). The complexity measure L1 from [9] indicates if a problem is linearly-separable if its value is zero, so every data set with a L1 value of zero will be discarded.

This method for generating binary data sets is limited by the proper combinatorics, and we can only obtain over 200 new data sets with the original 20 data sets with this first approach. In order to obtain more data sets, we group the classes two by two, that is, we create a new binary data set, and each of its two classes are the combination of two original classes each. For this second approach we have used *ecoli*, *glass* and *flare* data sets, since they have a high number of class labels. Again, those data sets with a L1 value of zero are discarded. Finally, all these combinations resulted in 438 binary classification problems.

To estimate the learning methods' accuracy we have used a 10-fold cross validation scheme once all the measures are computed. We have taken the average accuracy of training and test of the 10 partitions as a representative measure of the FH-GBML's performance.

In Table III we have summarized the global Training and Test accuracy obtained by the FH-GBML. The high standard deviation (std. dev.) present in the results (over the 10% in test) agrees with the irregularity observed in Figure 1.

TABLE III

GLOBAL AVERAGE TRAINING AND TEST ACCURACY/STD. DEV. FOR FH-GBML OVER THE 438 DATA SETS

	Global % Accuracy Training Global Training std. dev.	Global % Accuracy Test Global Test std. dev.
FH-GBML	85.39% 11.70	85.39% 12.30

B. Intervals and Rule Extraction with the Automatic Extraction Method

Once we have obtained the average FH-GBML's accuracy in training and test for every data set (of the 438), we have sorted them according to the different data complexity measures. Then we have used the automatic extraction method described above in order to obtain a series of intervals of such data complexity measures in which FH-GBML has obtained a significant good or bad behavior with respect to the global performance. In Table IV we have summarized the intervals obtained by the automatic method.

In Table V we depict the rules derived from the intervals obtained, and it is organized with the following columns for each rule:

- The first column corresponds to the identifier of the rule for further references.
- The "Range" column presents the domain of the rule.
- The third column "Support" presents the percentage of data sets which verify the antecedent part of the rule.
- The column "% Training, Std. Dev." shows the average accuracy in training of all the examples which are covered by the rule. The standard deviation of the average training accuracy is computed as well.

TABLE IV

INTERVALS OBTAINED BY THE AUTOMATIC METHOD FOR THE 438 DATA SETS FOR FH-GBML

Good behavior		Bad behavior	
Measure	Range	Measure	Range
N1	[0.001,0.05]	F1	[0.03,0.86]
N2	[0.01,0.24]	N1	[0.30,1.00]
L1	[0.03,0.22]	N2	[0.59,1.05]
L2	[0.00,0.12]	N3	[0.16,0.54]
		N4	[0.19,0.49]
		L2	[0.25,0.56]
		T1	[0.96,1.00]
		T2	[0.56,14.00]

- The column "Training Diff." contains the difference between the training accuracy of the rule and the training accuracy across all 438 data sets.
- The column "% Test, Std. Dev." shows the average accuracy in test of all the examples which are covered by the rule. The standard deviation of the average test accuracy is computed as well.
- The column "Test Diff." contains the difference between the test accuracy of the rule and the test accuracy across all 438 data sets.

As we can observe in this table, the positive rules (denoted with a "+" symbol in their identifier) always show a positive difference with the global average, both in training and test accuracy. The negative ones (with a "-" symbol in their identifier) verify the opposite case.

From these three tables we can point out the following:

- The support of the rules shows us that we can characterize a wide range of data sets and obtain significant differences in accuracy.
- The measures N1 and N2 and L2 are used in order to find both good and bad behavior intervals. Therefore they appear to be interesting measures for characterizing both aspects of FH-GBML.
- Most of the data complexity measures used to describe the good behavior belong to the *measures of separability of classes* category, which appears to be very interesting to describe the good data sets for the FH-GBML.
- The data complexity measures F2, F3 and L3 have not been used by the automatic method.
- The automatic method use more data complexity measures for the bad behavior intervals, while the support are similar to the good behavior intervals.

With these simple and individual rules, an initial characterization of the good or bad data sets for the FRBCSs can be considered.

C. Combination of the Individual Rules

The objective of this section is to analyze the effect of combining the rules of good behavior, and the combination of the rules of bad behavior is considered as well. By means of merging the individual rules we can arrive at a more general description, with a wider support, of the behavior of the FH-GBML method.

Therefore we have considered the disjunctive combination of all the positive rules to obtain a single rule (Positive Rule

TABLE V
RULES OBTAINED FOR FH-GBML FROM THE AUTOMATIC INTERVALS

Good behavior rules						
Id.	Range	% Support	% Training Accuracy	Training Difference	% Test Accuracy	Test Difference
R1+	N1 ∈ [0.001,0.05]	17.352	99.768	6.199	98.222	9.981
R2+	N2 ∈ [0.01,0.24]	26.941	99.195	5.625	96.654	8.412
R3+	L1 ∈ [0.03,0.22]	22.603	98.764	5.195	95.754	7.512
R4+	L2 ∈ [0.00,0.12]	40.411	98.014	4.444	94.890	6.648
Bad behavior rules						
Id.	Range	% Support	% Training Accuracy	Training Difference	% Test Accuracy	Test Difference
R1-	F1 ∈ [0.034,0.86]	24.201	89.202	-4.367	82.182	-6.060
R2-	N1 ∈ [0.30,1.00]	21.005	83.645	-9.924	74.959	-13.283
R3-	N2 ∈ [0.59,1.05]	15.982	82.182	-11.387	73.561	-14.681
R4-	N3 ∈ [0.16,0.54]	21.005	83.531	-10.038	74.521	-13.721
R5-	N4 ∈ [0.19,0.49]	21.689	87.069	-6.501	80.847	-7.394
R6-	L2 ∈ [0.25,0.56]	30.137	87.911	-5.658	79.299	-8.943
R7-	T1 ∈ [0.96,1.00]	27.397	89.572	-3.998	82.201	-6.041
R8-	T2 ∈ [0.56,14.00]	23.973	89.878	-3.692	82.235	-6.007

Disjunction -PRD-), that is, we use the *or* operator to combine the individual positive rules. The same procedure is done with all the negative ones so we obtain another rule (Negative Rule Disjunction -NRD-). The new disjunctive rules will be activated if any of the component rules' antecedents are verified.

The PRD and NRD rules may present overlapping in their support, and a mutual exclusive description of the good and bad regions is desirable. We consider the conjunctive operator *and* and the difference *and not* between the PRD and NRD rules in order to tackle this issue. The difference will remove the data sets for which FH-GBML presents good or bad behavior from the disjunctive negative or positive rule, respectively. That is, by means of the difference we have tried to remove the data sets of the opposite type from the considered disjunctive rule. Thus we obtain three different kinds of intersections and an extra region:

- Intersection of positive disjunction *and* the negative disjunction ($PRD \wedge NRD$).
- Intersection of positive disjunction *and not* the negative disjunction ($PRD \wedge \neg NRD$).
- Intersection of negative disjunction *and not* the positive disjunction ($NRD \wedge \neg PRD$).
- *Not characterized* region, in which no rule covers its data sets.

In Table VI we have depicted the new collective rules for the FH-GBML method.

From the new obtained rules, we can point out the following for FH-GBML:

- The Positive Rule Disjunction (PRD) offers a high support and it also gives a good training and test accuracy (over the 98% and 94% respectively).
- The Negative Rule Disjunction (NRD) obtains a wide support as well (over the 66%). However, the differences in both training and test have decreased due to this increment in support with respect to the single rules of bad behavior.
- The Positive and Negative Rule Disjunction ($PRD \wedge NRD$) is more specific than PRD in isolation. It

is also similar to PRD in the training and test accuracy difference. This rule obtains positive differences in training and test accuracy, representing good data sets for the FRBCSs covered by the rules of bad behavior (NRD rule). That means that these data sets could not be recognized as good data sets for the FH-GBML as they were mixed with bad data sets.

- The Positive and Not Negative Rule Disjunction ($PRD \wedge \neg NRD$) has a lower support than $PRD \wedge NRD$. Its difference is higher than PRD and $PRD \wedge NRD$ rules, since the data sets with low accuracy for the FRBCSs present in $PRD \wedge NRD$ have been removed from PRD, becoming a more specific rule.
- The Negative and Not Positive Rule Disjunction ($NRD \wedge \neg PRD$) is a good rule to describe the bad behavior of the FRBCSs. It has a high support and both a high difference in training and test sets. When removing the good data sets of $PRD \wedge NRD$, the $NRD \wedge \neg PRD$ rule becomes more accurate.
- The data sets not characterized neither by the PRD rule or the NRD rule present a positive difference with the global accuracy both in training and test.

From all the disjunctive and new conjunctive rules, we can present PRD as a representative description of good data sets, and $NRD \wedge \neg PRD$ as a representative description for bad data sets. On the other hand, the not characterized data sets can be defined as good data sets for FH-GBML due to their positive difference in test accuracy. Therefore the union of PRD and not characterized data sets can be considered as the final characterization of the good data sets for the FRBCSs.

In Figure 4 we have depicted the two block regions for the FH-GBML method. On the left side the data sets not covered by $NRD \wedge \neg PRD$ are depicted, distinguishing between those covered by the PRD rule and the not covered ones. On the right side the data sets covered by $NRD \wedge \neg PRD$ are plot. The behavior of the FH-GBML is clearly differentiated in the two main distinguished regions.

TABLE VI
DISJUNCTION AND INTERSECTION RULES FROM ALL SIMPLE RULES FOR FH-GBML

Id.	Range	% Support	% Training Accuracy	Training Difference	% Test Accuracy	Test Difference
PRD	If R1+ or R2+ or R3+ or R4+ then good behavior	42.24	98.08	4.51	94.96	6.72
NRD	If R1- or R2- or R3- or R4- or R5- or R6- or R7- or R8- then bad behavior	66.89	91.06	-2.51	84.72	-3.53
$PRD \wedge NRD$	If PRD and NRD then good behavior	22.60	97.20	3.63	93.40	5.16
$PRD \wedge \neg NRD$	If PRD and not NRD then good behavior	19.63	99.09	5.52	96.75	8.51
$NRD \wedge \neg PRD$	If NRD and not PRD then bad behavior	44.29	87.93	-5.64	80.28	-7.96
not characterized	If not (PRD or NRD) then good behavior	13.47	97.97	4.40	93.35	5.11

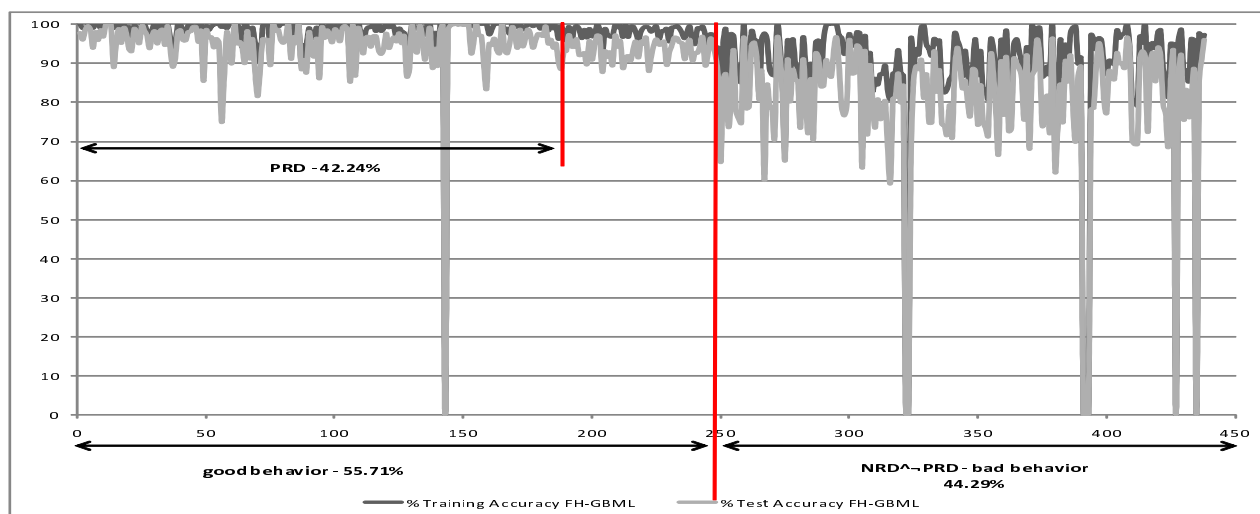


Fig. 4. FH-GBML block representation for PRD, not characterized and $NRD \wedge \neg PRD$ covered data sets

V. CONCLUDING REMARKS

We have proposed a new automatic method to characterize the domains of competence of FRBCSs. In order to analyze it, we have performed a study over a large set of binary data sets with the FH-GBML method. First we have computed twelve data complexity measures for each data sets. Next we have used the automatic method and we have obtained a set of intervals different for the FRBCS in which its performance is significantly good or bad. Then we have constructed descriptive rules from this intervals and we have studied the interaction between the proper rules and we have obtained one rule which is precise to describe the both good and bad performance of each of FH-GBML.

We present the possibility of determining automatically which data sets would prove to be good or bad for this FRBCS prior to its execution, using the data complexity measures. Moreover, the definition of good or bad domains of competence is provided and they can be adjusted to the user's needs.

We must point out that this is a study for a particular method. As a future work, more FRBCSs needs to be used

to analyze the generalization properties of our proposal. An extra amount of data sets is also needed in order to validate the obtained domains of competence. This work presents a new challenge that could be extended to other learning methods and to automatically analyze their domains of competence.

ACKNOWLEDGMENT

Supported by the Spanish Ministry of Science and Technology under Project TIN2008-06681-C06-01. J. Luengo holds a FPU scholarship from Spanish Ministry of Education and Science.

REFERENCES

- [1] M. R. Akbarzadeh-Totonchi and M. Moshtagh-Khorasani, "A hierarchical fuzzy rule-based approach to aphasia diagnosis," *Journal of Biomedical Informatics*, vol. 40, no. 5, pp. 465–475, 2007.
- [2] P. P. Angelov and X. Zhou, "Evolving fuzzy-rule-based classifiers from data streams," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 6, pp. 1462–1475, 2008.
- [3] A. Asuncion and D. Newman, "UCI machine learning repository," 2007. [Online]. Available: <http://archive.ics.uci.edu/ml/>
- [4] M. Basu and T. K. Ho, *Data Complexity in Pattern Recognition (Advanced Information and Knowledge Processing)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

- [5] R. Baumgartner and R. L. Somorjai, "Data complexity assessment in undersampled classification of high-dimensional biomedical data," *Pattern Recognition Letters*, vol. 12, pp. 1383–1389, 2006.
- [6] E. Bernadó-Mansilla and T. K. Ho, "Domain of competence of xcs classifier system in complexity measurement space," *IEEE Transactions on Evolutionary Computation*, vol. 9, no. 1, pp. 82–104, 2005.
- [7] A. Fernández, S. García, M. J. del Jesús, and F. Herrera, "A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets," *Fuzzy Sets and Systems*, vol. 159, no. 18, pp. 2378–2398, 2008.
- [8] S. García, J. R. Cano, E. Bernadó-Mansilla, and F. Herrera, "Diagnose of effective evolutionary prototype selection using an overlapping measure," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, no. 8, pp. 2378–2398, 2009.
- [9] T. K. Ho and M. Basu, "Complexity measures of supervised classification problems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 289–300, 2002.
- [10] J. Hühn and E. Hüllermeier, "FR3: A fuzzy rule learner for inducing reliable classifiers," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 1, pp. 138–149, 2009.
- [11] —, "FURIA: an algorithm for unordered fuzzy rule induction," *Data Mining and Knowledge Discovery*, vol. 19, no. 3, pp. 293–319, 2009.
- [12] H. Ishibuchi, T. Nakashima, and M. Nii, *Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining (Advanced Information Processing)*. Springer-Verlag New York, Inc., 2004.
- [13] 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.
- [14] H. Ishibuchi, T. Yamamoto, and T. Nakashima, "Hybridization of fuzzy GBML approaches for pattern classification problems," *IEEE Transactions on System, Man and Cybernetics B*, vol. 35, no. 2, pp. 359–365, 2005.
- [15] L. Kuncheva, *Fuzzy Classifier Design*. Springer, Berlin, 2000.
- [16] J. Luengo and F. Herrera, "Domains of competence of fuzzy rule based classification systems with data complexity measures: A case of study using a fuzzy hybrid genetic based machine learning method," *Fuzzy Sets and Systems*, vol. 161, no. 1, pp. 3–19, 2010.
- [17] E. G. Mansoori, M. J. Zolghadri, and S. D. Katebi, "SGERD: A steady-state genetic algorithm for extracting fuzzy classification rules from data," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 4, pp. 1061–1071, 2008.
- [18] J. S. Sánchez, R. A. Mollineda, and J. M. Sotoca, "An analysis of how training data complexity affects the nearest neighbor classifiers," *Pattern Analysis & Applications*, vol. 10, no. 3, pp. 189–201, 2007.
- [19] G. Schaefer, M. Zaviscek, and T. Nakashima, "Thermography based breast cancer analysis using statistical features and fuzzy classification," *Pattern Recognition*, vol. 42, no. 6, pp. 1133–1137, 2009.
- [20] C. Tsang, S. Kwong, and H. Wang, "Genetic-fuzzy rule mining approach and evaluation of feature selection techniques for anomaly intrusion detection," *Pattern Recognition*, vol. 40, no. 9, pp. 2373–2391, 2007.
- [21] S. A. Vinterbo, E.-Y. Kim, and L. Ohno-Machado, "Small, fuzzy and interpretable gene expression based classifiers," *Bioinformatics*, vol. 21, no. 9, pp. 1964–1970, 2005.