On the Usefulness of Fuzzy Rule Based Systems Based on Hierarchical Linguistic Fuzzy Partitions

Alberto Fernández, Victoria López, María José del Jesus, and Francisco Herrera

Summary. In the recent years, a high number of fuzzy rule learning algorithms have been developed with the aim of building the Knowledge Base of Linguistic Fuzzy Rule Based Systems. In this context, it emerges the necessity of managing a flexible structure of the Knowledge Base with the aim of modeling the problems with a higher precision. In this work, we present a short overview on the Hierarchical Fuzzy Rule Based Systems, which consists in a hierarchical extension of the Knowledge Base, preserving its structure and descriptive power and reinforcing the modeling of those problem subspaces with more difficulties by means of a hierarchical treatment (higher granularity) of the rules generated in these areas. Finally, this methodology includes a summarisation step by means of a genetic rule selection process in order to obtain a compact and accurate model. We will show the goodness of this methodology by means of a case of study in the framework of imbalanced data-sets in which we compare this learning scheme with some basic Fuzzy Rule Based Classification Systems and with the well-known C4.5 decision tree, using the proper statistical analysis as suggested in the specialised literature. Finally, we will develop a discussion on the usefulness of this methodology, analysing its advantages and proposing some new trends for future work on the topic in order to extract the highest potential of this technique for Fuzzy Rule Based Systems.

Keywords: Fuzzy Rule Based Classification Systems, Hierarchical Fuzzy Partitions, Hierarchical Systems of Linguistic Rules Learning Methodology, Granularity, Imbalanced Data-sets.

Alberto Fernández · María José del Jesus
Dept. of Computer Science, University of Jaén, Spain
e-mail: alberto.fernandez@ujaen.es, mjjesus@ujaen.es

Victoria López · Francisco Herrera
Dept. of Computer Science and A.I., University of Granada, Spain
e-mail: vlopez@decsai.ugr.es, herrera@decsai.ugr.es

W. Pedrycz and S.-M. Chen (Eds.): Granular Computing and Intell. Sys., ISRL 13, pp. 155–184.
springerlink.com

1 Introduction

Linguistic Fuzzy Rule Based Systems (FRBSs) (Yager and Filev, 1994) have demonstrated their ability for control problems (Palm et al, 1997), modeling (Pedrycz, 1996), classification or data mining (Kuncheva, 2000; Ishibuchi et al, 2004) in a huge number of applications. 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 (Bastian, 1994), which may suppose a loss in accuracy when dealing with some complex systems, i.e. high dimensional problems, in the presence of noise or when the classes are overlapped (in classification tasks).

It is possible to make some considerations to face this drawback. Many different possibilities to improve the linguistic fuzzy modeling have been considered in the specialised literature. All of these approaches share the common idea of improving the way in which the linguistic fuzzy model performs the interpolative reasoning by inducing a better cooperation among the rules in the Knowledge Base (KB). This rule cooperation may be induced acting on three different model components:

- *Approaches acting on the Data Base (DB).* For example a priori granularity learning (Cordón et al, 2001b) or membership function tuning (Alcalá et al, 2007).
- *Approaches acting on the Rule Base (RB).* The most common approach is rule selection (Ishibuchi et al, 1995; Gacto et al, 2009) but also multiple rule consequent learning (Cordón and Herrera, 2000) could be considered.
- *Approaches acting on the whole KB.* This includes the KB derivation (Magdalena and Monasterio-Huelin, 1997) and a hierarchical linguistic rule learning (Ishibuchi et al, 1993; Cordón et al, 2002).

In this work we will focus on this last issue, studying the use of a hierarchical environment in order to improve the behaviour of linguistic FRBSs. This approach has been first proposed by Herrera and Martínez (Herrera and Martínez, 2001) in the field of Decision Making and later by Cordón et al. (Cordón et al, 2002) in the scenario of regression problems. The hierarchical model preserves the original descriptive power of FRBS and increases its accuracy by reinforcing those problem subspaces that are especially difficult by means of a hierarchical treatment of the rules generated in these areas producing a more general and well defined structure, the Hierarchical Knowledge Base (HKB).

Our aim is to provide a wide overview on the hierarchical methodology for linguistic fuzzy systems, describing the different approaches that have been developed on the topic including the basic hierarchical systems of linguistic rules learning methodology (HSLR-LM) (Cordón et al, 2002), the hybridization of weighted rule learning with the hierarchical approach (Alcalá et al, 2003) and the iterative scheme through different granularity levels of the HSLR-LM (Cordón et al, 2003). In order to show their usefulness, we will present a case of study on classification with imbalanced data-sets (He and Garcia, 2009; Sun et al, 2009), in which we have made use of the adaption of Hierarchical Fuzzy Rule Based Systems (HFRBSs) to this scenario (Fernández et al, 2009).

According to all these points, this work is organised as follows. First, Section 2 introduces the concept of hierarchal fuzzy partitions and the definition of the HKB. Next, Section 3 describes the learning methodology for HFRBSs and some extensions that have been developed this approach. In Section 4 we present the framework of imbalanced data-sets and the specific hierarchical fuzzy methodology that was designed for this scenario. Then, we provide a case of study for imbalanced data-sets in Section 5, showing some experimental results on this new topic. Finally, in Section 6 we will point out some concluding remarks about the study carried out and we will discuss some new challenges on the topic that can support further work from the basis previously presented.

2 Hierarchical Linguistic Fuzzy Partitions

As we have stated in the introduction of this work, the KB structure usually employed in the field of linguistic modeling has the drawback of its lack of accuracy when working with very complex systems. This fact is due to some problems related to the linguistic rule structure considered, which are a consequence of the inflexibility of the concept of linguistic variable (Zadeh, 1975). A summary of these problems may be found in (Bastian, 1994; Carse et al, 1996), and it is briefly enumerated as follows.

- There is a lack of flexibility in the FRBSs because of the rigid partitioning of the input and output spaces.
- When the system input variables are dependent themselves, it is very hard to fuzzy partition the input spaces.
- The homogenous partitioning of the input and output spaces when the input-output mapping varies in complexity within the space is inefficient and does not scale to high-dimensional spaces.
- The size of the RB directly depends on the number of variables and linguistic terms in the system. Obtaining an accurate FRBS requires a significant granularity amount, i.e., it needs of the creation of new linguistic terms. This granularity increase causes the number of rules to rise significantly, which may take the system to lose the capability of being interpretable for human beings.

At least two things could be done to solve many of these problems and to improve the model accuracy. On the one hand, we can use approximative fuzzy modeling, with the consequence of losing the model interpretability. On the other hand, we can refine a linguistic model trying not to change too much the meaning of the linguistic variables neither the descriptive power of the final FRBS generated.

Related to the previous issue, a crucial task for dealing with linguistic information is to determine the granularity of uncertainty, i.e., the cardinality of the fuzzy linguistic term set used to assess the linguistic variables. Depending on the uncertainty degree held by a source of information qualifying a phenomenon, the linguistic term set will have more or less terms (Bonissone and Decker, 1985; Herrera et al, 2000).

In order to overcome this drawback, Herrera and Martínez proposed in (Herrera and Martínez, 2001) the use of a set of multigranular linguistic contexts that they denoted as linguistic hierarchies term sets. A linguistic hierarchy is a set of levels, where each level is a linguistic term set with different granularity to the rest of levels of the hierarchy. The purpose of this extension is the flexibilisation of the KB to become an HKB. This is possible by the development of a new KB structure, where the linguistic variables of the linguistic rules could take values from fuzzy partitions with different granularity levels. An HKB is said to be composed of a set of layers ("levels" in the notation of Herrera and Martínez), and each layer is defined by its components in the following way:

$$layer(t,n(t)) = DB(t,n(t)) + RB(t,n(t)), \quad (1)$$

with $n(t)$ being the number of linguistic terms in the fuzzy partitions of layer t, $DB(t,n(t))$ being the 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:

$$DB(t,n(t)) \equiv DB^t \text{ and } RB(t,n(t)) \equiv RB^t \quad (2)$$

At this point, we should note that, At this point, we should note that, in this work, we are considering *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). This type of membership functions is the most suitable for this environment easing the mapping between the different layers of the HKB. Furthermore, this environment can be extended to interval-valued fuzzy sets adding a degree of uncertainty in the definition of the support of each fuzzy term (Sanz et al, 2010).

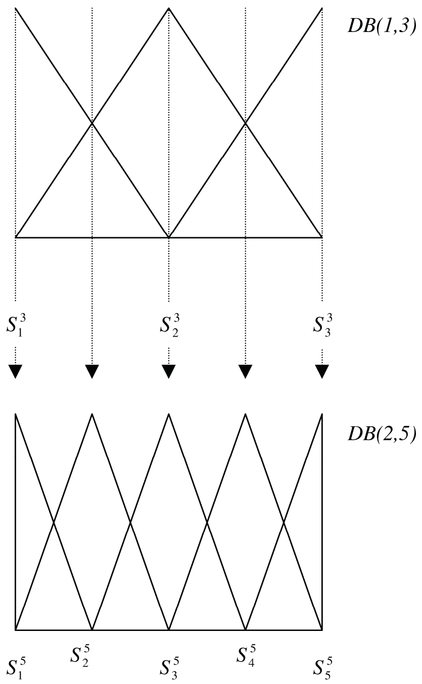
Specifically, the number of linguistic terms in the t-linguistic partitions is defined in the following way:

$$n(t) = (n(1) - 1) \cdot 2^{t-1} + 1, \quad (3)$$

with $n(1)$ being the granularity of the initial fuzzy partitions, linguistic hierarchy basic rules. This structure must satisfy the following rules:

1. To preserve all former modal points of the membership functions of each linguistic term from one level to the following one.
2. To make smooth transitions between successive levels. The aim is to build a new linguistic term set, new linguistic term will be added between each pair of terms belonging to the term set of the previous level. To carry out this insertion, we shall reduce the support of the linguistic labels in order to keep place for the new one located in the middle of them.

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



DB(t, n(t))		**DB(t+1, 2·n(t)-1)**
$S_{k-1}^{n(t)}$	$\rightarrow$	$S_{2k-3}^{2\cdot n(t)-1}$
		$S_{2k-2}^{2\cdot n(t)-1}$
$S_{k}^{n(t)}$	$\rightarrow$	$S_{2k-1}^{2\cdot n(t)-1}$
		$S_{2k}^{2\cdot n(t)-1}$
$S_{k+1}^{n(t)}$	$\rightarrow$	$S_{2k+1}^{2\cdot n(t)-1}$

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

The main purpose of developing a Hierarchical Rule Base (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:

$$HKB = \bigcup_t layer(t,n(t)) \tag{4}$$

3 Hierarchical Fuzzy Rule Based Systems

In the previous section we have introduced the concept of hierarchical fuzzy partitions. As explained, this approach presents a more flexible KB structure that allows improving the accuracy of the FRBCSs without losing their interpretability: the HKB, which is composed of a Hierarchical Data Base (HDB) and an HRB.

In this section, we will first introduce the basic two-level HSLR-LM to generate an HFRBS (Cordón et al, 2002). Next, we describe the hybridization of weighted rule learning with the hierarchical approach which was developed with the aim of improving the system accuracy (Alcalá et al, 2003). Finally, an extension of the two-level learning method is presented as an iterative scheme through different granularity levels (Cordón et al, 2003). We must point out that in all these three cases, the scenario in which these approaches have been proposed is devoted to regression problems.

3.1 A Two-Level HSLR Learning Methodology

The first methodology to build an HFRBS was proposed by Cordón et al. (Cordón et al, 2001a) as a strategy to improve simple linguistic models preserving their structure and descriptive power, reinforcing only the modeling of those problem subspaces with more difficulties. For the sake of maintaining the interpretability of the final model, this basic HSLR was only based on two hierarchical levels, i.e., two layers.

In the following, the structure of the learning methodology and its most important components are described in detail. Specifically, the algorithm to obtain an HFRBS is based on two processes:

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 Genetic Algorithm (GA).

To do so, it is needed to use an existing inductive LRG-method based on the existence of a set of input-output training data $X = \{x_1, \ldots, x_p, \ldots, x_m\}$ with $x_p = (x_{p1}, \ldots, x_{pn}, y_p)$, $p = 1,2,\ldots,m$ where x_{pi} is the ith attribute value ($i = 1,2,\ldots,n$) of the p-th training pattern y_p is the output value, and a previously defined DB^1. Usually, the LRG-method selected is aimed to obtain simple linguistic fuzzy models, such as the Wang and Mendel's algorithm (Wang and Mendel, 1992), or the Thrift's algorithm (Thrift, 1991). Two measures of error are used in the algorithm:

1. Global measure (used to evaluate the complete RB): The Mean Square Error (MSE) for a whole RB, calculated over X, is defined as:

$$MSE(X, RB) = \frac{\sum_{x_p \in X} (y_p - s(x_p))^2}{2 \cdot |X|}$$

with $s(x_p)$ being the output value obtained from the HLSR using the current RB when the input variable values are $x_p = (x_{p1}, \ldots, x_{pn}, y_p)$, and y_p is the known desired value.

2. Local measure (used to determine if an individual rule is expanded): The MSE for a simple rule[1], $R_i^{n(1)}$, calculated over X_i, is showed as follows:

$$MSE(X_i, RB_i^{n(1)}) = \frac{\sum_{x_p \in X_i} (y_p - s_i(x_p))^2}{2 \cdot |X_i|}$$

with X_i being a set of the examples matching the i-th rule antecedents to degree $\tau \in (0,1]$ and $s_i(x_p)$ being the output value from this rule.

Table 1 Two-level learning method

HIERARCHICAL KNOWLEDGE BASE
Step 0. $RB(1, n(1))$ *Generation Process*
Step 1. $RB(2, 2 \cdot n(1) - 1)$ *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 (summarised in Table 1), which basically consists of the following steps:

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

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

a) *Calculate the error of* RB^1: $MSE(X, RB^1)$.

b) *Calculate the error of each 1-linguistic rule:* $MSE(X_i, RB_i^{n(1)})$.

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):*

If $MSE(X_i, R_i^{n(1)}) \geq \alpha \cdot MSE(X, RB^1)$ *Then* $R_i^{n(1)} \in RB^1_{bad}$

Else $R_i^{n(1)} \in RB^1_{good}$.

[1] Notice that other local error measures, such as the one showed in (Yen et al, 1998) could also be considered.

d) *Create* DB^2.

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

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 :

$$I(S_{jk}^{n(1)}) = \{S_h^{2\cdot n(1)-1} \in DB^2 / \max_{u \in U_k} \min\{\mu_{S_{jk}^{n(1)}}(u), \mu_{S_h^{2\cdot n(1)-1}}(u)\} \geq \delta\} \tag{5}$$

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

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

$$I(R_j^{n(1)}) = I(S_{j1}^{n(1)}) \times \ldots \times I(S_{js}^{n(1)}) \tag{6}$$

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:

$$CLR(R_j^{n(1)}) = LRG-method(I(R_j^{n(1)}), X(R_j^{n(1)})) = \\ = \{R_{j_1}^{2\cdot n(1)-1}, \ldots, R_{j_L}^{2\cdot n(1)-1}\} \tag{7}$$

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*:

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

Example: In the following, we show an example of the whole expansion process. Let us consider $n(1) = 3$ and the following linguistic partitions:

$$DB_{x_1}(1,3) = DB_{x_2}(1,3) = DB_y(1,3) = \{S^3, M^3, L^3\},$$

$$DB_{x_1}(2,5) = DB_{x_2}(2,5) = DB_y(2,5) = \{VS^5, S^5, M^5, L^5, VL^5\},$$

where S stands for Small, M for Medium, L for large, and V for Very. Let us consider the following bad performance *1-linguistic rule* to be expanded (see Fig. 2):

$$R_i^3 : \text{IF } x_1 \text{ is } S_{i1}^3 \text{ and } x_2 \text{ is } S_{i2}^3 \text{ THEN } y \text{ is } B_i^3$$

where the linguistic terms are, $S_{i1}^3 = S^3$, $S_{i2}^3 = S^3$, $B_i^3 = S^3$, and the resulting sets I with $\delta = 0.5$ are:

$$I(S_{i1}^3) = \{VS^5, S^5\}, I(S_{i2}^3) = \{VS^5, S^5\}, I(B_i^3) = F(\cdot) \subseteq D_y(2,5)$$

$$I(R_i^3) = I(S_{i1}^3) \times I(S_{i2}^3) \times I(B_i^3) .$$

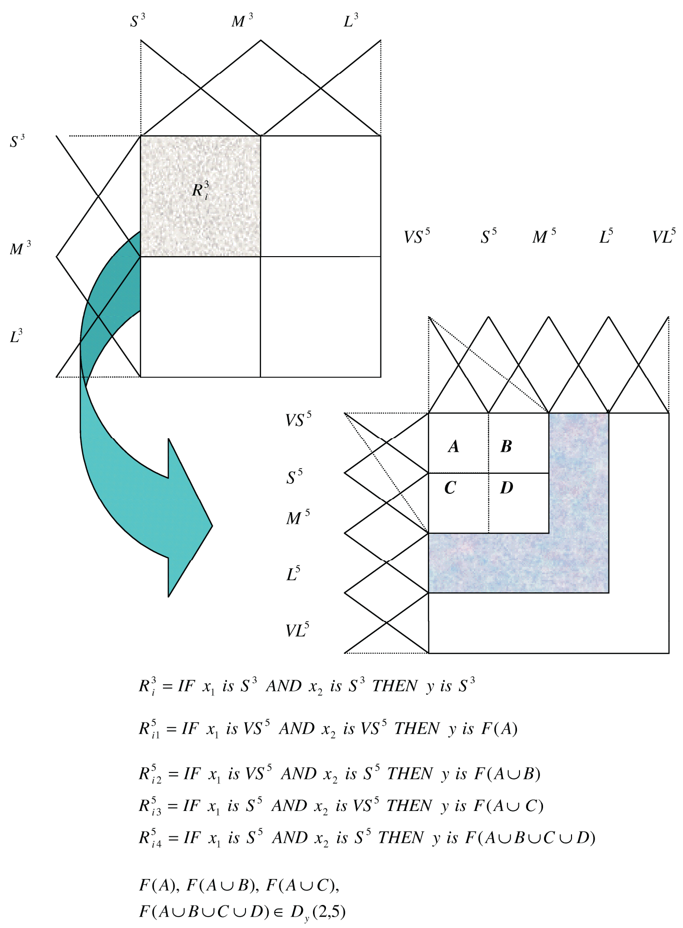


Fig. 2 Example of the HRB Generation Process.

Therefore, it is possible to obtain at most four *2-linguistic rules* generated by the LRG-method from the expanded R_i^3:

$$LRG(I(R_i^3), E_i) = \{R_{i1}^5, R_{i2}^5, R_{i3}^5, R_{i4}^5\}$$

This example is graphically showed in Fig. 2. In the same way, other bad performance neighbor rules could be expanded simultaneously.

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.

$$HRB = Selection\Pr ocess(JCLR)$$

After applying this algorithm, the HKB is obtained as:

$$HKB = HDB + HRB$$

3.2 Linguistic Modeling with Hierarchical Systems of Weighted Linguistic Rules

In (Alcalá et al, 2003), Alcalá et al. proposed the hybridization of the hierarchical scheme with the use of rule weights by extending the two-level HSLR-LM proposed in (Cordón et al, 2002). The resulting Hierarchical System of Weighted Linguistic Rules (HSWLR), presents a model structure which is extended by permitting the use of weighted hierarchical linguistic rules. Besides, the summarization component –which has the aim of selecting the subset of rules best cooperating among the rules generated to obtain the final HKB– was modified by allowing it to jointly perform the rule selection and the rule weight derivation. A GA (Michalewicz, 1996) performing the rule selection together with the derivation of rule weights was developed for this task.

Hence, this extended methodology was intended as a meta-method over any other LRG-method, developed to improve simple linguistic fuzzy models by only reinforcing the modeling of those problem subspaces with more difficulties whereas the use of rule weights improved the way in which they interact. This extension of the learning methodology was named two-level HSWLR Learning Methodology (HSWLR-LM) and consists of two modifications:

- *Modification of the HRB structure and Inference System*, in order to consider the use of weights, obtaining Weighted HKB s (WHKB s).
- *Modification of the rule selection process (Step 3 of the two-level HSLR-LM algorithm),* to consider the derivation of rule weights.

Weighted Hierarchical Knowledge Base

In this case, only the rule structure in the HRB has to be modified. The same structure of the weighted linguistic rules will be used to form the Weighted HRB (WHRB) and then the WHKB:

$$WHKB = HDB + WHRB$$

Therefore, the fuzzy reasoning process must be extended as in the case of weighted linguistic rules, considering the matching degree of the rules fired.

In this way, we can define the WHRB as a whole HRB together with their corresponding rule weights:

$$WHKB = \bigcup_t RB^t + \bigcup_t W^t \ .$$

with W^t being the set of weights associated to the rules from layer t. We should notice that these weights are obtained over the whole HRB (and not over the isolated layers) since they must consider the way in which all the rules interact, i.e., the weights considered in the different layers, W^t, are interdependent. Therefore, *they must be jointly derived once the whole HRB is available.*

Algorithm

The same operation mode of the two-level HSLR-LM algorithm will be followed to generate linguistic fuzzy models with this new structure, but including the weight learning. Again, the Wang and Mendel's algorithm (Wang and Mendel, 1992) was considered as LRG-method to obtain simple linguistic fuzzy models, although any other technique could be used. Therefore, the main steps of the extended algorithm are the following ones:

HIERARCHICAL KNOWLEDGE BASE GENERATION PROCESS
Step 0. $RB(1, n(1))$ *Generation Process*
Step 1. $RB(2, 2 \cdot n(1) - 1)$ *Generation Process*
Step 2. *Summarization Process* → **(Extract Repeated (JCLR) + Weights).**

HIERARCHICAL RULE BASE GENETIC SELECTION PROCESS
Step 3. Genetic Weight Derivation and Rule Selection Process
- Genetic selection of a subset of rules presenting good cooperation.
- Genetic derivation of the weights associated to these rules.

Fig. 3 presents the flowchart of this algorithm. Specifically, at Step 3 of the two-level HSWLR-LM, a GA with double coding scheme ($C = C_1 + C_2$) was employed for both *rule selection* and *weight derivation*.

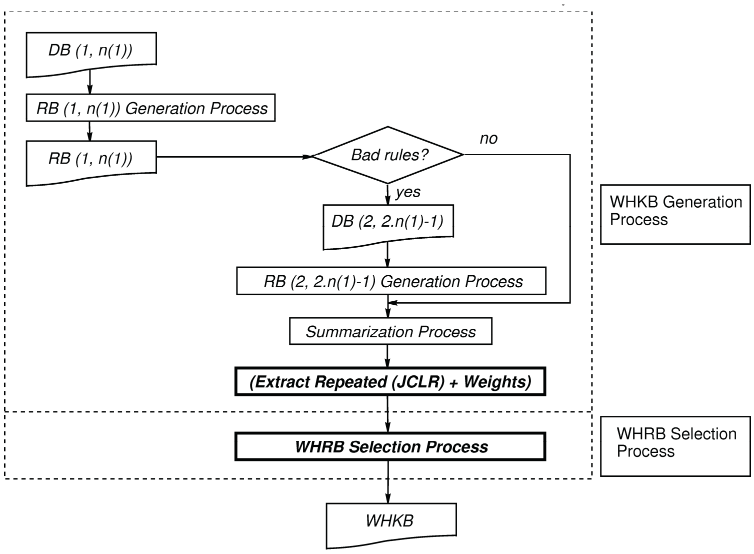


Fig. 3 HSWLR Learning Methodology.

3.3 An Iterative Methodology for Hierarchical Linguistic Modeling

In the beginning of this section we have introduced a basic approach to develop hierarchical models from a limited HKB focused on interpretability: the HSLRs of two levels (Cordón et al, 2002). In (Cordón et al, 2003), Cordón et al. extended the former model structure, i.e., the HKB, and proposed an iterative HSLR learning methodology to learn it from examples. As the name suggests, this methodology iteratively selects bad performance linguistic rules, which need more specificity, and expands them locally through different granularity levels. This fact produces a wide spectrum of solutions—from high interpretable to high accurate, and tradeoff solutions—and avoids typical drawbacks of prototype-based linguistic rule generation methods (LRG-methods).

The iterative HSLR-LM was developed as a parametrised methodology. The factor of expansion controls the level of bad performance that a rule should have to be expanded into more specific ones. Thus, a low factor implies a small expansion, a smaller number of rules, and a more interpretable model. In this sense, the basic approach (Cordón et al, 2002) is a special case which makes use of this parameter to obtain interpretable hierarchical models. Another parameter to be considered is the iteration of the algorithm. It is used to control the granularity level that more specific hierarchical rules, which replace those ones with bad performance, should have (see Fig. 4).

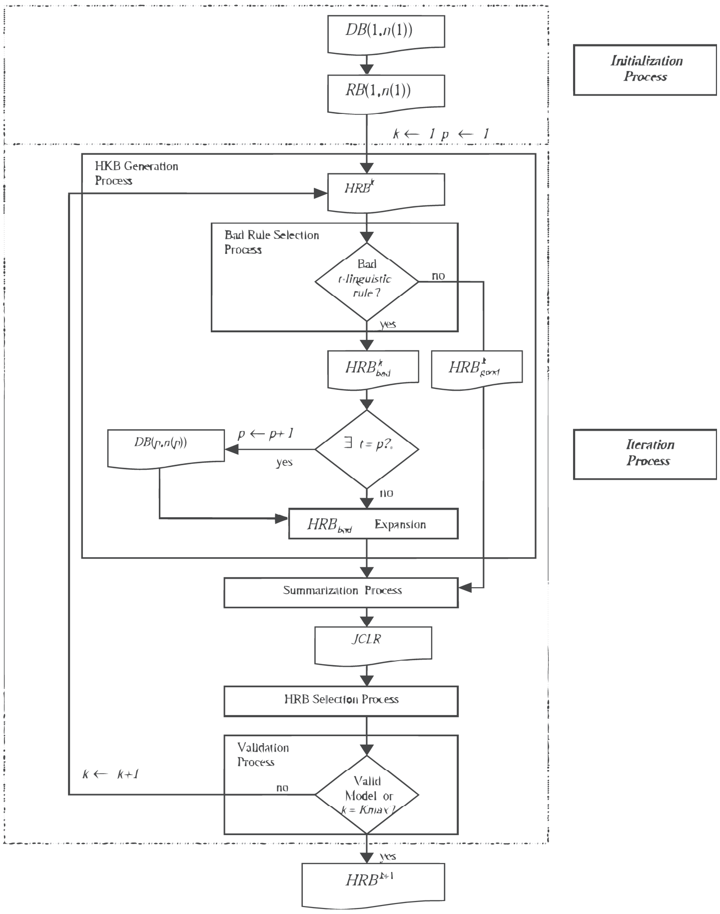


Fig. 4 Algorithm of the iterative HSLR-LM design process.

As the reader may have already noticed, the main change in the structure of the algorithm presented in the beginning of this section is the enabling of having several granularity levels and therefore to iterate steps 1 to 3 several times depending on a parameter k. According to this, it performs gradual and local-oriented refinements on problem subspaces that are badly modeled by previous models rather than in the whole problem domain. Furthermore, it integrates the improved local behavior with the whole model by summarization processes which ensure a good global performance.

4 On the Use of Hierarchical Fuzzy Rule Based Classification Systems on Imbalanced Data-Sets

In this section we will address a new and interesting problem, named as classification with imbalanced data-sets (He and Garcia, 2009; Sun et al, 2009) which consists in learning from highly skewed data, in general terms. Indeed, this problem has been recently identified as one of the hot topics in data mining (Yang and Wu, 2006) and therefore we must emphasise its significance and guide our efforts in dealing with this type of applications.

Regarding linguistic fuzzy systems, in a previous study of some of the authors we have shown the goodness of this type of models to deal with imbalanced data-sets (Fernández et al, 2008). In this section we will introduce a solution we have provided to improve the accuracy of Fuzzy Rule Based Classification Systems (FRBCS) in this framework (Fernández et al, 2009). Specifically we proposed a hierarchical environment, by means of the HSLR-LM described in the previous section, by increasing the granularity of the fuzzy partitions on the boundary areas between the classes, in order to obtain a better separability.

In order to do so, we will first introduce in this section the scenario of classification with imbalanced data-sets, that is, its main features and how to deal with this problem. Then, we will describe how to adapt the HSLR-LM to this specific framework.

4.1 An Introduction to Classification with Imbalance Data-Sets

Learning from imbalanced data is an important topic that has recently appeared in the Machine Learning community (Chawla et al, 2004; He and Garcia, 2009; Sun et al, 2009). This problem is very representative since it appears in a variety of real-world applications including, but not limited to, medical applications, finance, telecommunications, biology and so on.

In this framework, the class distribution is not uniform, resulting on a high number of examples belonging to one or more classes and a few from the rest. The minority classes are usually associated to the most interesting concepts from the point of view of learning and, due to that fact, the cost derived from a misclassification of one of the examples of these classes is higher than that of the majority classes. In this work we will focus on binary problems where there is just one positive and negative class.

Standard classifier algorithms usually have a bias towards the majority class, since the rules which predict the higher numbers of examples are positively weighted during the learning process in favour of the standard accuracy rate metric, which does not take into account the class distribution of the data. Consequently, the instances belonging to the minority class are misclassified more often than those belonging to the majority class.

Another important issue of this problem are the small disjuncts that can be found in the data set (Weiss and Provost, 2003) and the difficulty of most learning algorithms in detecting these areas (Orriols-Puig et al, 2009; Orriols-Puig and Bernadó-Mansilla, 2009). In fact, learning algorithms try to benefit those models with a higher degree of coverage and these small disjuncts imply the application of very specific models which are discarded in favour or more general ones.

Furthermore, another handicap of imbalanced data sets, which is related to the apparition of small disjuncts, is the overlapping between the examples of the positive and the negative class (García et al, 2008), in which the minority class examples can be simply treated as noise and ignored by the learning algorithm. These phenomena are depicted in Figure 5.a and 5.b respectively.

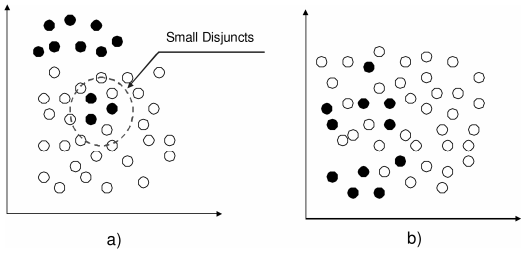


Fig. 5 Example of the imbalance between classes: a) small disjuncts b) overlapping between classes

A large number of approaches have previously been proposed for dealing with the class-imbalance problem. These approaches can be categorised into two groups: the internal approaches which create new algorithms or modify existing ones to take the class-imbalance problem into consideration (Barandela et al, 2003; Wu and Chang, 2005; Xu et al, 2007) and the external approaches which pre-process the data in order to diminish the effects of their class imbalance (Batista et al, 2004; Estabrooks et al, 2004). Furthermore, cost-sensitive learning solutions incorporating both the data and algorithmic level approaches assume higher misclassification costs with samples in the minority class and seek to minimise the high cost errors (Domingos, 1999; Zhou and Liu, 2006; Sun et al, 2007).

The great advantage of the external approaches is that they are more versatile, since their use is independent of the classifier selected. Furthermore, we may preprocess all data sets beforehand in order to use them to train different classifiers. In this manner, the computation time needed to prepare the data is lower.

Specifically, in the framework of fuzzy classification we analyzed the cooperation of some preprocessing methods with FRBCSs (Fernández et al, 2008), showing a good behaviour for the oversampling methods, especially in the case of the SMOTE methodology (Chawla et al, 2002).

4.2 Adaptation of the Hierarchical Learning Process for Imbalanced Data

As we stated previously, the framework of imbalanced data-sets requires some specific adaptations for the algorithms in order to obtain a good performance for both classes. In the remaining of this section we will first introduce the type of rules, rule weights and inference model used for standard classification tasks. Next, we will highlight the main changes carried out in the hierarchical learning process for its application to imbalanced problems.

Fuzzy Rule Based Classification Systems

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:

Rule R_j : If x_1 is A_{j1} and $\ldots$ and x_n is A_{jn} then Class = C_j with RW_j (8)

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 (we use triangular membership functions), C_j is a class label, and RW_j is the rule weight.

In the specialised literature rule weights have been used in order to improve the performance of FRBCSs (Ishibuchi and Nakashima, 2001). For the framework of imbalanced data-sets and following the conclusions extracted in (Fernández et al, 2008), as heuristic method for the rule weight the Penalised Certainty Factor (Ishibuchi and Yamamoto, 2006) was employed:

$$RW_j = \frac{\sum_{x_p \in ClassC_j} \mu_{A_j}(x_p)}{\sum_{p=1}^{m} \mu_{A_j}(x_p)} - \frac{\sum_{x_p \notin ClassC_j} \mu_{A_j}(x_p)}{\sum_{p=1}^{m} \mu_{A_j}(x_p)} \tag{9}$$

The Fuzzy Reasoning Method of the winning rule (classical approach) (Cordón et al, 1999) was used for classifying new patterns by the RB. The single winner rule R_w is determined for a new pattern $x_p = (x_{p1},\ldots,x_{pn})$ as

$$\mu_w(x_p) \cdot RW_w = \max\{\mu_j(x_p) \cdot RW_j\}; x_p \in X, j = 1\ldots L \tag{10}$$

The new pattern x_p is classified as Class C_w , which is the consequent class of the winner rule R_w . If multiple fuzzy rules have the same maximum value but different consequent classes for the new pattern x_p in the previous equation, the classification of x_p is rejected. The classification is also rejected if no fuzzy rule is compatible with the new pattern x_p.

Two-Level Learning Method for Building HFRBCSs in Imbalanced Domains

The main scheme of the two-level HSLR-LM was maintained in this case, following exactly the same steps. In this case, as LRG-method the Chi et al. (Chi et al, 1996) approach must be considered for building a FRBCS.

The main change refers to the measures of error are used in the algorithm used both to evaluate the complete RB and 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:

$$Acc_i(X_i, RB) = \frac{|\{x_p \in X_i / FRM(x_p, RB) = Class(x_p)\}|}{|X_i|} \tag{11}$$

where $|\cdot|$ is the number of patterns, with X_i being the subset of examples of the i-th class ($i \in 1 \ldots M$), $FRM(x_p, RB)$ is the output class computed following the fuzzy reasoning process 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 showed as follows:

$$Acc(X, R_j^{n(1)}) = \frac{|X^+(R_j^{n(1)})|}{|X(R_j^{n(1)})|} \tag{12}$$

$$X^+(R_j^{n(1)}) = \{x_p \in X / \mu_{R_j^{n(1)}}(x_p) > 0 .and. Class(x_p) = Class(R_j^{n(1)})\} \tag{13}$$

$$X(R_j^{n(1)}) = \{x_p \in X / \mu_{R_j^{n(1)}}(x_p) > 0\} \tag{14}$$

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, because we are using as Fuzzy Reasoning Method the winning rule approach.

Additionally, another significant modification must be developed in the HRB genetic selection. Specifically, during the chromosome evaluation the fitness function must be in accordance with the framework of imbalanced data-sets and therefore the geometric mean of the true rates (Barandela et al, 2003) was used. This metric defined as:

$$GM = \sqrt{\frac{TP}{TP + FN} \cdot \frac{TN}{FP + TN}} \tag{15}$$

where TP and TN are the true rate for the positive and negative instances and FP and FN the rate for the false positives and negatives respectively. This metric attempts to maximise the accuracy of each one of the two classes with a good balance, being a performance metric that links both objectives.

5 Case of Study: Hierarchical Fuzzy Rule Based Classification Systems for Imbalanced Data-Sets

In the previous part of this work we have introduced the problem of imbalanced data-sets and the solution proposed to increase the performance of linguistic FRBCSs by means of the use of the HSLR-LM adapted for imbalanced problems (Fernández et al, 2009). Now, this section has the aim of presenting a case of study in which we show the goodness of the proposed methodology in contrast with other well-known fuzzy approaches and with C4.5 (Quinlan, 1993).

According to this, we will first introduce the configuration of the two-level learning method, determining all the parameters used in this experimental study and the selected benchmark data-sets. Next, we will present the statistical tests used in all our analysis. Finally, we will analyze the results of the HFRBCS when applied to imbalanced data-sets globally, and considering two different degrees of imbalance. This last part of the study is divided into two sections:

- A comparative study is carried out between the HFRBCS model and other fuzzy learning methodologies, including Chi et al.'s (Chi et al, 1996) and Ishibuchi et al.'s (Ishibuchi and Yamamoto, 2005) rule learning algorithms, and an approach proposed by Xu et al. for imbalanced data-sets, called E-Algorithm (Xu et al, 2007).
- The performance of the HFRBCSs is compared against the well-known C4.5 algorithm (Quinlan, 1993) as a well-known classifier that has been widely used for this type of problems (Batista et al, 2004; Estabrooks et al, 2004; Orriols-Puig et al, 2009; Su et al, 2006; Su and Hsiao, 2007; Sun et al, 2007).

5.1 Experimental Set-Up: Parameters and Data-Sets

In our former studies (Fernández et al, 2008) we selected as a good FRBCS model the use of the product T-norm as conjunction operator, together with the Penalised Certainty Factor (Ishibuchi and Yamamoto, 2005) approach for the rule weight and Fuzzy Reasoning Method of the winning rule. This configuration will be employed for all the FRBCSs used in this work, including Chi et al.'s method, Ishibuchi et al.'s approach and E-Algorithm.

After several trials, we selected the following values for the parameters in the learning method for building HFRBCSs:

- Rule Generation:
  - $\delta, n(t+1)$-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

In order to reduce the effect of imbalance, we will employ the SMOTE preprocessing method (Chawla et al, 2002) for our experiments, consider only the

1-nearest neighbour to generate the synthetic samples (using the euclidean distance), and we balance the training data to the 50% class distribution. The E-Algorithm is always applied without preprocessing.

For Ishibuchi et al.'s rule generation method and E-Algorithm, only rules with three or less antecedent attributes are generated. Furthermore we have restricted the number of fuzzy rules in the RB to 30 for each class, using as selection measure the product of support and confidence. This configuration is the one indicated by the authors in (Ishibuchi and Yamamoto, 2005; Xu et al, 2007).

Table 2 Summary Description for imbalanced data-sets.

Data-set	**#Ex.**	**#Atts.**	**Class(min.,maj.)**	**%Class(min.; maj.)**	**IR**
Data-sets with Low Imbalance (IR 1.5 to 9)					
Glass1	214	9	(build-win-non_float-proc;remainder)	(35.51, 64.49)	1.82
Ecoli0vs1	220	7	(im;cp)	(35.00, 65.00)	1.86
Wisconsin	683	9	(malignant;benign)	(35.00, 65.00)	1.86
Pima	768	8	(tested-positive;tested-negative)	(34.84, 66.16)	1.90
Iris0	150	4	(Iris-Setosa;remainder)	(33.33, 66.67)	2.00
Glass0	214	9	(build-win-float-proc;remainder)	(32.71, 67.29)	2.06
Yeast1	1484	8	(nuc;remainder)	(28.91, 71.09)	2.46
Vehicle1	846	18	(Saab;remainder)	(28.37, 71.63)	2.52
Vehicle2	846	18	(Bus;remainder)	(28.37, 71.63)	2.52
Vehicle3	846	18	(Opel;remainder)	(28.37, 71.63)	2.52
Haberman	306	3	(Die;Survive)	(27.42, 73.58)	2.68
Glass0123vs456	214	9	(non-windowglass;remainder)	(23.83, 76.17)	3.19
Vehicle0	846	18	(Van;remainder)	(23.64, 76.36)	3.23
Ecoli1	336	7	(im;remainder)	(22.92, 77.08)	3.36
New-thyroid2	215	5	(hypo;remainder)	(16.89, 83.11)	4.92
New-thyroid1	215	5	(hyper;remainder)	(16.28, 83.72)	5.14
Ecoli2	336	7	(pp;remainder)	(15.48, 84.52)	5.46
Segment0	2308	19	(brickface;remainder)	(14.26, 85.74)	6.01
Glass6	214	9	(headlamps;remainder)	(13.55, 86.45)	6.38
Yeast3	1484	8	(me3;remainder)	(10.98, 89.02)	8.11
Ecoli3	336	7	(imU;remainder)	(10.88, 89.12)	8.19
Page-blocks0	5472	10	(remainder;text)	(10.23, 89.77)	8.77
Data-sets with High Imbalance (IR higher than 9)					
Yeast2vs4	514	8	(cyt;me2)	(9.92, 90.08)	9.08
Yeast05679vs4	528	8	(me2;mit,me3,exc,vac,erl)	(9.66, 90.34)	9.35

Table 2 (*Continued*)

Vowel0	988	13	(hid;remainder)	(9.01, 90.99)	10.10
Glass016vs2	192	9	(ve-win-float-proc;build-win-float-proc,build-win-non_float-proc,headlamps)	(8.89, 91.11)	10.29
Glass2	214	9	(Ve-win-float-proc;remainder)	(8.78, 91.22)	10.39
Ecoli4	336	7	(om;remainder)	(6.74, 93.26)	13.84
Yeast1vs7	459	8	(vac;nuc)	(6.72, 93.28)	13.87
Shuttle0vs4	1829	9	(RadFlow;Bypass)	(6.72, 93.28)	13.87
Glass4	214	9	(containers;remainder)	(6.07, 93.93)	15.47
Page-blocks13vs2	472	10	(graphic;horiz.line,picture)	(5.93, 94.07)	15.85
Abalone9vs18	731	8	(18;9)	(5.65, 94.25)	16.68
Glass016vs5	184	9	(tableware;build-win-float-proc,build-win-non_float-proc,headlamps)	(4.89, 95.11)	19.44
Shuttle2vs4	129	9	(FpvOpen;Bypass)	(4.65, 95.35)	20.5
Yeast1458vs7	693	8	(vac;nuc,me2,me3,pox)	(4.33, 95.67)	22.10
Glass5	214	9	(tableware;remainder)	(4.20, 95.80)	22.81
Yeast2vs8	482	8	(pox;cyt)	(4.15, 95.85)	23.10
Yeast4	1484	8	(me2;remainder)	(3.43, 96.57)	28.41
Yeast1289vs7	947	8	(vac;nuc,cyt,pox,erl)	(3.17, 96.83)	30.56
Yeast5	1484	8	(me1;remainder)	(2.96, 97.04)	32.78
Ecoli0137vs26	281	7	(pp,imL;cp,im,imU,imS)	(2.49, 97.51)	39.15
Yeast6	1484	8	(exc;remainder)	(2.49, 97.51)	39.15
Abalone19	4174	8	(19;remainder)	(0.77, 99.23)	128.87

The Imbalance Ratio (IR) (Orriols-Puig et al, 2009), defined as the ratio of the number of instances of the majority class and the minority class, is used as a threshold to categorise the different imbalanced scenarios: data-sets with a *low imbalance* when the instances of the positive class are between 10 and 40% of the total instances (IR between 1.5 and 9) and data-sets with a *high imbalance* where there are no more than 10% of positive instances in the whole data-set compared to the negative ones (IR higher than 9).

Specifically, we have considered forty-four data sets from UCI repository (Asuncion and Newman, 2007) with different IR. Table 2 summarises the data employed in this study and shows, for each data set, the number of examples (#Ex.), number of attributes (#Atts.), class name of each class (minority and majority), class attribute distribution and IR. This table is ordered by the IR, from low to highly imbalanced data-sets.

In order to develop the study, we use a five fold cross validation approach, that is, five partitions for training and test sets, 80% for training and 20% for test, where the five test data-sets form the whole set. For each data-set we consider the average results of the five partitions.

5.2 Statistical Tests for Evaluation

In this paper, we use the hypothesis testing techniques to provide statistical support to the analysis of the results (García et al, 2009; Sheskin, 2006). 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, causing the statistical analysis to lose credibility with these parametric tests (Demšar, 2006).

We will use the Wilcoxon signed-rank test (Wilcoxon, 1945) as a non-parametric statistical procedure for performing pairwise comparisons between two algorithms. For multiple comparisons we use the Iman-Davenport test (Sheskin, 2006) to detect statistical differences among a group of results, and the Holm post-hoc test (Holm, 1979) in order to find which algorithms are distinctive among a $1 \times n$ comparison.

The post-hoc procedure allows us to know whether a hypothesis of comparison of means could be rejected at a specified level of significance α. However, it is very interesting to compute the p-value associated to each comparison, which represents the lowest level of significance of a hypothesis that results in a rejection. In this manner, we can know whether two algorithms are significantly different and how different they are.

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.

These tests are suggested in the studies presented in (Demšar, 2006; García and Herrera, 2008; García et al, 2009, 2010), where their use in the field of Machine Learning is highly recommended. For a wider description of the use of these tests, any interested reader can find additional information on the Website http://sci2s.ugr.es/sicidm/, together with the software for their application.

5.3 Experimental Study

In this part of the study we will focus on determining whether the HFRBCS is robust in the framework of imbalanced data-sets and if it improves the performance of other FRBCSs approaches and the well known C4.5 algorithm. Following this idea, Table 3 shows the results for the test partitions for each FRBCS method with its associated standard deviation. Specifically, by columns we include the Chi et al.'s method with 3 and 5 labels (Chi-3 and Chi-5), the Ishibuchi et al.'s method (Ishibuchi05), the E-Algorithm and the HFRBCS. Additionally, we include the results for the C4.5 decision tree. This table is divided by the IR, on the one hand data-sets with low imbalance and, on the other hand, data-sets with high imbalance. The best global result for test is stressed in **boldface** in each case.

Table 3 Detailed table of results for FRBCSs in imbalanced data-sets. Only test results are shown.

Data-set	Chi-3	Chi-5	Ishibuchi05	E-Algorithm	HFRBCS	C4.5
Data-Sets with Low Imbalance						
Glass1	64.90 ± 6.91	64.91 ± 6.87	59.29 ± 10.33	0.00 ± 0.00	73.66 ± 4.66	75.11 ± 3.74
Ecoli0vs1	92.27 ± 5.93	95.56 ± 5.15	96.70 ± 2.40	95.25 ± 4.75	93.63 ± 6.45	97.95 ± 2.20
Wisconsin	88.91 ± 2.13	43.58 ± 5.86	95.78 ± 1.38	96.01 ± 1.55	88.24 ± 1.63	95.44 ± 2.01
Pima	66.80 ± 5.93	66.78 ± 2.28	71.10 ± 4.45	55.01 ± 4.64	68.72 ± 5.26	71.26 ± 4.05
Iris0	100.0 ± 0.00	98.97 ± 2.29	100.0 ± 0.00	100.0 ± 0.00	100.0 ± 0.00	98.97 ± 2.29
Glass0	64.06 ± 3.51	63.69 ± 1.80	69.39 ± 7.70	0.00 ± 0.00	76.57 ± 8.05	78.14 ± 2.21
Yeast1	67.69 ± 1.91	69.66 ± 1.52	51.41 ± 12.18	0.00 ± 0.00	71.71 ± 2.39	70.86 ± 2.95
Vehicle1	70.92 ± 4.34	71.88 ± 1.25	64.89 ± 4.37	3.09 ± 6.90	71.76 ± 2.64	69.28 ± 3.41
Vehicle2	85.54 ± 3.36	87.19 ± 3.04	67.82 ± 4.95	43.83 ± 13.17	90.61 ± 2.17	94.85 ± 1.68
Vehicle3	69.22 ± 4.89	63.13 ± 1.95	63.12 ± 4.06	0.00 ± 0.00	66.80 ± 3.34	74.34 ± 1.08
Haberman	58.91 ± 6.03	60.40 ± 2.40	62.65 ± 2.84	4.94 ± 11.06	57.08 ± 4.09	61.32 ± 3.85
Glass0123vs456	85.83 ± 3.04	85.94 ± 1.66	88.56 ± 5.18	82.09 ± 6.96	88.37 ± 3.97	90.13 ± 3.17
Vehicle0	86.41 ± 3.06	84.93 ± 1.61	75.94 ± 1.42	39.07 ± 16.49	88.92 ± 1.96	91.10 ± 2.70
Ecoli1	85.28 ± 9.77	86.05 ± 8.57	85.71 ± 2.86	77.81 ± 7.90	84.18 ± 12.69	76.10 ± 9.58
New-Thyroid2	89.81 ± 10.77	96.34 ± 6.65	94.21 ± 4.23	88.57 ± 3.82	99.72 ± 0.63	96.51 ± 4.87
New-Thyroid1	87.44 ± 8.11	95.38 ± 8.80	89.02 ± 13.52	88.52 ± 8.79	98.58 ± 2.48	97.98 ± 3.79
Ecoli2	88.01 ± 5.45	87.64 ± 4.96	87.00 ± 4.43	70.35 ± 15.36	87.62 ± 8.24	91.60 ± 4.86
Segment0	94.99 ± 0.45	95.88 ± 1.21	42.47 ± 2.79	95.33 ± 1.14	97.51 ± 1.11	99.26 ± 0.61
Glass6	83.87 ± 9.82	78.13 ± 7.78	86.27 ± 8.19	90.23 ± 3.77	86.95 ± 10.84	83.00 ± 9.05
Yeast3	90.13 ± 4.09	89.33 ± 3.30	77.06 ± 17.73	81.99 ± 2.28	90.41 ± 2.34	88.50 ± 3.66
Ecoli3	87.58 ± 4.08	91.61 ± 4.95	85.39 ± 3.70	78.54 ± 8.68	90.81 ± 4.43	88.77 ± 7.65
Page-Blocks0	79.91 ± 4.29	87.25 ± 1.94	32.16 ± 9.61	64.51 ± 2.79	91.40 ± 0.67	94.84 ± 1.52
Mean	81.29 ± 4.90	80.19 ± 3.90	74.81 ± 5.83	57.05 ± 5.46	84.69 ± 4.09	**85.70 ± 3.68**
Data-Sets with High Imbalance						
Yeast2vs4	86.80 ± 5.53	86.39 ± 7.35	70.85 ± 23.45	80.92 ± 9.09	89.32 ± 4.18	85.09 ± 10.15
Yeast05679vs4	78.91 ± 5.99	75.99 ± 6.39	79.49 ± 9.54	59.99 ± 16.44	73.18 ± 7.47	74.88 ± 10.88
Vowel0	98.37 ± 0.61	97.87 ± 1.84	89.03 ± 6.63	89.63 ± 6.09	98.82 ± 1.62	94.74 ± 5.22
Glass016vs2	40.84 ± 7.62	56.17 ± 5.16	41.18 ± 15.40	0.00 ± 0.00	58.37 ± 20.04	48.91 ± 29.44
Glass2	47.67 ± 10.16	49.24 ± 8.19	43.55 ± 15.70	9.87 ± 22.07	54.84 ± 20.57	33.86 ± 32.29
Ecoli4	91.27 ± 7.43	92.11 ± 8.35	86.92 ± 8.65	92.43 ± 8.24	93.02 ± 8.17	81.28 ± 11.67

Table 3 (*Continued*)

Yeast1vs7	80.05 ± 6.43	63.02 ± 12.62	53.15 ± 10.35	27.55 ± 26.06	70.74 ± 12.40	67.73 ± 2.28
Shuttle0vs4	99.12 ± 1.15	98.71 ± 1.18	99.16 ± 1.15	98.40 ± 1.26	99.12 ± 1.15	99.97 ± 0.07
Glass4	84.96 ± 13.80	81.75 ± 11.24	78.27 ± 17.70	83.38 ± 19.89	70.39 ± 40.49	83.71 ± 10.78
Page-Blocks13vs4	91.92 ± 4.76	92.93 ± 9.48	94.53 ± 4.88	94.12 ± 10.33	98.64 ± 0.65	99.55 ± 0.47
Abalone9-18	63.93 ± 11.00	66.47 ± 10.67	65.78 ± 9.23	32.29 ± 20.61	67.56 ± 14.01	53.19 ± 8.25
Glass016vs5	71.48 ± 40.17	75.59 ± 42.27	88.77 ± 2.48	65.14 ± 39.41	77.96 ± 43.61	72.08 ± 42.33
Shuttle2vs4	89.99 ± 8.61	78.34 ± 43.87	99.17 ± 1.13	100.0 ± 0.00	97.49 ± 2.71	99.15 ± 1.90
Yeast1458vs7	62.40 ± 4.55	58.76 ± 8.57	40.80 ± 16.58	0.00 ± 0.00	62.49 ± 6.26	41.19 ± 6.06
Glass5	81.56 ± 12.65	64.33 ± 38.40	89.96 ± 2.43	50.61 ± 47.17	68.73 ± 39.56	86.70 ± 15.44
Yeast2vs8	72.75 ± 14.99	78.76 ± 8.60	72.83 ± 14.97	72.83 ± 14.97	72.47 ± 15.10	78.23 ± 13.05
Yeast4	82.99 ± 3.10	83.07 ± 2.58	71.36 ± 23.29	32.16 ± 20.59	82.64 ± 2.29	65.00 ± 8.94
Yeast1289vs7	76.12 ± 7.24	69.26 ± 4.57	48.55 ± 16.86	50.00 ± 13.62	69.37 ± 4.37	64.13 ± 9.00
Yeast5	93.41 ± 5.35	93.64 ± 2.70	94.94 ± 0.38	88.17 ± 7.04	94.20 ± 2.59	92.04 ± 4.99
Ecoli0137vs26	71.04 ± 41.38	49.57 ± 46.41	71.31 ± 41.65	73.65 ± 43.09	71.48 ± 41.80	71.21 ± 41.31
Yeast6	87.50 ± 10.55	87.73 ± 9.32	88.42 ± 6.06	51.72 ± 13.76	84.92 ± 12.88	80.38 ± 15.47
Abalone19	62.96 ± 8.27	66.71 ± 10.21	66.09 ± 9.40	0.00 ± 0.00	70.19 ± 8.56	15.58 ± 21.36
Mean	78.00 ± 10.51	75.75 ± 13.63	74.28 ± 11.72	56.95 ± 15.44	**78.45 ± 14.11**	72.21 ± 13.70
All Data-Sets						
Mean	79.65 ± 7.71	77.97 ± 8.77	74.55 ± 8.78	57.00 ± 10.45	**81.67 ± 9.10**	78.95 ± 8.69

This study is divided into two parts. First, we will analyze the results globally for all imbalanced data-sets and then, we will study the two imbalance scenarios defined in this paper. Furthermore, our aim is to test the HFRBCS against the FRBCSs approaches and C4.5 separately.

Global Analysis of the Hierarchical Fuzzy Rule Based Classification System

First of all, we will study the performance of the HFRBCS with the remaining FRBCSs approaches. In order to compare the results, we will use a multiple comparison test to find the best approach in this case, considering the results in the test partitions (GM_{Tst}). The results of Iman-Davenport tests informs us of the rejection of the null hypothesis of equality of means (p-value near to zero), telling us of the existence of significant differences among the observed results in all data-sets. Next, Table 4 shows the rankings of the 5 algorithms considered.

Now, we apply a Holm test to compare the best ranking method (HFRBCS) with the remaining fuzzy methods. The result of this test is shown in Table 5, 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 the best ranking algorithm or not.

Table 4 Rankings obtained through a Friedman test for FRBCSs in all imbalance data-sets.

Algorithm	Ranking
HFRBCS	2.09091
Chi-5	2.77273
Chi-3	3.00000
Ishibuchi05	3.02273
E-Algorithm	4.11364

Table 5 Holm test Table for FRBCSs in all imbalanced data-sets. HFRBCS is the control method.

i	algorithm	z	p	α/i	Hypothesis
4	E-Algorithm	6.00038	1.96858E-9	0.01250	Rejected for HFRBCS
3	Ishibuchi05	2.76422	0.00576	0.01667	Rejected for HFRBCS
2	Chi-3	2.69680	0.00700	0.02500	Rejected for HFRBCS
1	Chi-5	2.02260	0.04311	0.05000	Rejected for HFRBCS

Therefore, analyzing the results presented in Table 3 and the statistical study shown in Tables 4 and 5 we conclude that our model is a solid FRBCS approach to deal with imbalanced data-sets, as it has shown to be the best performing algorithm when comparing with the remaining fuzzy rule learning methods applied in this study.

Finally, we use a Wilcoxon test for the comparison with the C4.5 algorithm, which is shown in Table 6. We can observe that our proposal achieves a higher ranking, but this is not enough to reject the null hypothesis. We may conclude that both approaches have a similar performance when treating all imbalanced data-sets as a whole, without taking into account the IR.

Table 6 Wilcoxon test to compare the HFRBCS against C4.5 in all imbalanced data-sets. R+ corresponds to HFRBCS and R- to C4.5.

Comparison	R^+	R^-	Hypothesis (α =0.05)	p-value
HFRBCS vs. C4.5	589	401	Not Rejected	0.273

Analysis of the Hierarchical Fuzzy Rule Based Classification System According to the Imbalance Ratio

In the final part of our study, we will analyze the behaviour of our hierarchical approach in each imbalanced scenario. Table 7 shows, by columns, the geometric mean in training and test of the different algorithms considered, for the two types of data-sets, that is, low and high imbalance (IR lower than 9 and higher than 9 respectively). The last column corresponds to the global results. Reader can refer to Table 3, presented in the previous part of this study, where we show the detailed results in each data-set.

Table 7 Results table for FRBCSs for the different degrees of imbalance

Algorithm	Low Imbalance		High Imbalance		All Data-Sets	
	GM_{Tr}	GM_{Tst}	GM_{Tr}	GM_{Tst}	GM_{Tr}	GM_{Tst}
Chi-3	85.50 ± 1.28	81.29 ± 4.90	83.64 ± 2.43	78.00 ± 10.51	84.57 ± 1.86	79.65 ± 7.71
Chi-5	91.31 ± 0.69	80.19 ± 3.90	89.04 ± 1.32	75.75 ± 13.63	90.17 ± 1.01	77.97 ± 8.77
Ishibuchi05	75.45 ± 3.04	74.81 ± 5.83	76.90 ± 6.35	74.28 ± 11.72	76.17 ± 4.70	74.55 ± 8.78
E-Algorithm	58.33 ± 4.09	57.05 ± 5.46	65.72 ± 5.06	56.95 ± 15.44	62.02 ± 4.57	57.00 ± 10.45
HFRBCS	94.30 ± 0.80	84.69 ± 4.09	93.35 ± 1.30	**78.45 ± 14.11**	93.82 ± 1.05	**81.57 ± 9.10**
C4.5	94.95 ± 0.87	**85.70 ± 3.68**	95.81 ± 1.77	72.21 ± 13.70	95.38 ± 1.32	78.95 ± 8.69

The main conclusion extracted from this table is that our HFRBCS is very robust in both imbalanced scenarios considered, as it obtains very competitive results independently of the IR. Next, we will analyze the results in each case, for data-sets with low and high imbalance. As we did in the previous section, we will compare the HFRBCS with the FRBCSs and with the C4.5 decision tree separately.

- **Data-sets with low imbalance:** This study is shown in Tables 9 and 10. First, we check for statistical differences an Iman-Davenport tests obtaining a p-value of 2.98974E-5. Table 9 shows the ranking for the algorithms and Table 8 contains a Holm test, which shows that the HFRBCS is better in performance than the remaining FRBCS unless the Chi et al.'s method with 5 labels.

Table 8 Rankings obtained through a Friedman test for FRBCSs in data-sets with low imbalance.

Algorithm	Ranking
HFRBCS	1.97727
Chi-5	2.63636
Chi-3	3.06818
Ishibuchi05	3.11364
E-Algorithm	4.20454

Table 9 Holm test table for FRBCSs in data-sets with low imbalance. HFRBCS is the control method.

i	algorithm	z	p	α/i	Hypothesis
4	E-Algorithm	4.67197	2.98329E-6	0.01250	Rejected for HFRBCS
3	Ishibuchi05	2.38366	0.01714	0.01667	Rejected for HFRBCS
2	Chi-3	2.28831	0.02212	0.02500	Rejected for HFRBCS
1	Chi-5	1.38252	0.16681	0.05000	Not Rejected

Now, we will compare the performance achieved by our proposal with C4.5 in low imbalanced data-sets by means of a Wilcoxon test, which is shown in Table 10. Furthermore, we compare the HFRBCS with the Chi et al.'s approach with 5 labels in order to check if we find differences between both algorithms.

Table 10 Wilcoxon test to compare the HFRBCS against Chi-5 and C4.5 in data-set with low imbalance. R+ corresponds to HFRBCS and R- to Chi-5 and C4.5 in each case.

Comparison	R^+	R^-	Hypothesis (α =0.05)	p-value
HFRBCS vs. Chi-5	219	34	Rejected for HFRBCS	0.003
HFRBCS vs. C4.5	84	169	Not Rejected	0.168

The main conclusion after this study is that the HFRBCS is better than the rest of the FRBCS methods. It outperforms the base Chi LRG-method, the Ishibuchi et al.'s approach and the E-Algorithm. When compared with C4.5, there are no statistical differences in this imbalance scenario.

- **Data-sets with high imbalance**: This part of the study is very important, since it includes the data-sets with a higher degree of imbalance. In this manner, we can analyze how the imbalance actually affects the different methods employed in this study. For this purpose, we use a Iman-Daverport test in order to find statistical differences obtaining a p-value of 0.00330. Next, Table 11 shows the ranking for the FRBCS algorithms, in which our HFRBCS proposal is the first one. Finally, we perform a Holm test, which is shown in Table 12, where we can only conclude that the HFRBCS is better than the E-Algorithm in data-sets with high imbalance.

Table 11 Rankings obtained through a Friedman test for FRBCSs in data-sets with high imbalance.

Algorithm	Ranking
HFRBCS	2.20454
Chi-5	2.90909
Chi-3	2.93182
Ishibuchi05	2.93182
E-Algorithm	4.02273

Table 12 Holm test table for FRBCSs in data-sets with high imbalance. HFRBCS is the control method.

i	algorithm	z	p	α/i	Hypothesis
4	E-Algorithm	3.81385	1.36818E-4	0.01250	Rejected for HFRBCS
3	Ishibuchi05	1.52554	0.12712	0.01667	Not Rejected
2	Chi-3	1.52554	0.12712	0.02500	Not Rejected
1	Chi-5	1.47787	0.13944	0.05000	Not Rejected

A Wilcoxon test (Table 13) will help us to make a pairwise comparison between our proposal and the remaining algorithms, including C4.5 in this case. Now, we detect differences between the HFRBCS and the Chi et al.'s method with 5 labels per variable, but it remains statistically similar to the Ishibuchi et al.'s algorithm and the Chi et al.'s method with 3 labels. Nevertheless, watching the results for the comparison with C4.5 we see that the null hypothesis is rejected in favour of our HFRBCS proposal.

Table 13 Wilcoxon test to compare the HFRBCS against the remaining FRBCS approaches and C4.5 in data-set with high imbalance. R+ corresponds to HFRBCS and R- to the remaining algorithms in each case.

Comparison	R^+	R^-	Hypothesis (α =0.05)	p-value
HFRBCS vs. C4.5	192	61	Rejected for HFRBCS	0.033

According to these results, we must emphasise the good behaviour achieved in highly imbalanced data-sets by the all fuzzy models studied here, particularly for our proposal. Furthermore, we can determine that it is very competitive, since it outperforms C4.5 algorithm in this type of data-sets, with a p-value of 0.033.

6 Concluding Remarks

In this work, we have presented a wide overview on hierarchical linguistic fuzzy systems, describing the main features of this type of systems, the learning methodology proposed to build such a model and the extensions developed in the literature.

The main aim of this hierarchical approach is to obtain a good balance among different granularity levels, applying a higher granularity in the areas where the RB has a bad performance in order to obtain a better coverage of that area of the space of solutions, and a lower granularity that provides a good generalization.

Regarding the interpretability-accuracy tradeoff of this methodology, we have stated that it is not always true that a set of rules with a higher granularity level performs a more accurate modeling of a problem than another with a lower one. The relationship between accuracy and interpretability does not only depend on granularity and specificity, but also on other factors, for example, rule weights, flexible rule consequents, and moreover, compasity of and cooperation policies between the rules.

As a case of study, we have shown the improvement obtained by applying this methodology in the framework of imbalanced data-sets. Specifically, the classification accuracy of the base FRBCS is enhanced in the overlapping areas between the minority and majority classes by combining both the fine and thick granularity. In the experimental study, we have shown statistically that our proposal performs better than well known FRBCSs approaches and that clearly outperforms the C4.5 decision tree, generally for all data-sets and particularly in data-sets with high imbalance.

Acknowledgment

This work had been supported by the Spanish Ministry of Science and Technology under Projects TIN2008-06681-C06-01/02, and the Andalusian Research Plan TIC-3928.

References

Alcalá, R., Cano, J.R., Cordón, O., Herrera, F., Villar, P., Zwir, I.: Linguistic modeling with hierarchical systems of weighted linguistic rules. Int. J. Approx. Reason. 32(2–3), 187–215 (2003)

Alcalá, R., Alcalá-Fdez, J., Herrera, F.: A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection. IEEE T. Fuzzy Syst. 15(4), 616–635 (2007)

Asuncion, A., Newman, D.: UCI machine learning repository (2007), http://www.ics.uci.edu/~mlearn/MLRepository.html

Barandela, R., Sánchez, J.S., García, V., Rangel, E.: Strategies for learning in class imbalance problems. Pattern Recogn. 36(3), 849–851 (2003)

Bastian, A.: How to handle the flexibility of linguistic variables with applications. Int. J. Uncertain Fuzz 2(4), 463–484 (1994)

Batista, G., Prati, R.C., Monard, M.C.: A study of the behaviour of several methods for balancing machine learning training data. SIGKDD Explor. 6(1), 20–29 (2004)

Bonissone, P.P., Decker, K.: Selecting uncertainty calculi and granularity: An experiment in trading-off precision and complexity. In: Kanal, L.N., Lemmer, J.F. (eds.) UAI, pp. 217–248. Elsevier, Amsterdam (1985)

Carse, B., Fogarty, T.C., Munro, A.: Evolving fuzzy rule based controllers using genetic algorithms. Fuzzy Set Syst. 80, 273–294 (1996)

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: Synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357 (2002)

Chawla, N.V., Japkowicz, N., Kolcz, A.: Editorial: special issue on learning from imbalanced data sets. SIGKDD Explor. 6(1), 1–6 (2004)

Chi, Z., Yan, H., Pham, T.: Fuzzy algorithms with applications to image processing and pattern recognition. World Scientific, Singapore (1996)

Cordón, O., Herrera, F.: A proposal for improving the accuracy of linguistic modeling. IEEE T. Fuzzy Syst. 8(4), 335–344 (2000)

Cordón, O., del Jesus, M.J., Herrera, F.: A proposal on reasoning methods in fuzzy rule-based classification systems. Int. J. Approx Reason 20(1), 21–45 (1999)

Cordón, O., Herrera, F., Hoffmann, F., Magdalena, L.: Genetic Fuzzy Systems. In: Evolutionary tuning and learning of fuzzy knowledge bases. World Scientific, Singapore (2001a)

Cordón, O., Herrera, F., Villar, P.: Generating the knowledge base of a fuzzy rule-based system by the genetic learning of data base. IEEE T. Fuzzy Syst. 9(4), 667–674 (2001b)

Cordón, O., Herrera, F., Zwir, I.: Linguistic modeling by hierarchical systems of linguistic rules. IEEE T. Fuzzy Syst. 10(1), 2–20 (2002)

Cordón, O., Herrera, F., Zwir, I.: A hierarchical knowledge-based environment for linguistic modeling: Models and iterative methodology. Fuzzy Set Syst. 138(2), 307–341 (2003)

Demšar, J.: Statistical comparisons of classifiers over multiple data sets. J. Mach. Learn. Res. 7, 1–30 (2006)

Domingos, P.: Metacost: a general method for making classifiers cost sensitive. In: Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining, pp. 155–164 (1999)

Estabrooks, A., Jo, T., Japkowicz, N.: A multiple resampling method for learning from imbalanced data sets. Comput. Intell. 20(1), 18–36 (2004)

Fernández, A., García, S., del Jesus, M.J., Herrera, F.: A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets. Fuzzy Set Syst. 159(18), 2378–2398 (2008)

Fernández, A., del Jesus, M., Herrera, F.: Hierarchical fuzzy rule based classification system with genetic rule selection for imbalanced data-set. Int. J. Approx. Reason. 50, 561–577 (2009)

Gacto, M.J., Alcalá, R., Herrera, F.: Adaptation and application of multi-objective evolutionary algorithms for rule reduction and parameter tuning of fuzzy rule-based systems. Soft Comput. 13(5), 419–436 (2009)

García, S., Herrera, F.: An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons. J. Mach. Learn. Res. 9, 2677–2694 (2008)

García, S., Fernández, A., Luengo, J., Herrera, F.: A study of statistical techniques and performance measures for genetics–based machine learning: Accuracy and interpretability. Soft Comput. 13(10), 959–977 (2009)

García, S., Fernández, A., Luengo, J., Herrera, F.: Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. Inform. Sciences 180, 2044–2064 (2010)

García, V., Mollineda, R., Sánchez, J.S.: On the k-NN performance in a challenging scenario of imbalance and overlapping. Pattern Anal. Appl. 11(3–4), 269–280 (2008)

He, H., Garcia, E.A.: Learning from imbalanced data. IEEE T. Knowl. Data En. 21(9), 1263–1284 (2009)

Herrera, F., Martínez, L.: A model based on linguistic 2-tuples for dealing with multigranular hierarchical linguistic contexts in multi-expert decision making. IEEE T. Syst. Man. CY B 31(2), 227–234 (2001)

Herrera, F., Herrera-Viedma, E., Martínez, L.: A fusion approach for managing multigranularity linguistic terms sets in decision making. Fuzzy Set Syst. 114(1), 43–58 (2000)

Holm, S.: A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6, 65–70 (1979)

Ishibuchi, H., Nakashima, T.: Effect of rule weights in fuzzy rule-based classification systems. IEEE T. Fuzzy Syst. 9(4), 506–515 (2001)

Ishibuchi, H., Yamamoto, T.: Rule weight specification in fuzzy rule-based classification systems. IEEE T. Fuzzy Syst. 13, 428–435 (2005)

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

Ishibuchi, H., Nozaki, K., Yamamoto, N., Tanaka, H.: Selecting fuzzy if-then rules for classification problems using genetic algorithms. IEEE T. Fuzzy Syst. 9(3), 260–270 (1995)

Ishibuchi, H., Nakashima, T., Nii, M.: Classification and modeling with linguistic information granules: Advanced approaches to linguistic Data Mining. Springer, Heidelberg (2004)

Kuncheva, L.: Fuzzy classifier design. Springer, Berlin (2000)

Magdalena, L., Monasterio-Huelin, F.: A fuzzy controller with learning through the evolution of its knowledge base. Int. J. Approx. Reason. 16(3), 335–358 (1997)

Michalewicz, Z.: Genetic algorithms + data structures = evolution programs. Springer, Heidelberg (1996)

Orriols-Puig, A., Bernadó-Mansilla, E.: Evolutionary rule-based systems for imbalanced data-sets. Soft Comput. 13(3), 213–225 (2009)

Orriols-Puig, A., Bernadó-Mansilla, E., Goldberg, D.E., Sastry, K., Lanzi, P.L.: Facetwise analysis of XCS for problems with class imbalances. IEEE T. Evolut. Comput. 13, 260–283 (2009)

Palm, R., Driankov, D., Hellendoorn, H.: Model based fuzzy control. Springer, Berlin (1997)

Pedrycz, W.: Fuzzy modelling: paradigms and practice. Kluwer Academic Press, Dordrecht (1996)

Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo (1993)

Peña Reyes, C.A., Sipper, M.: Fuzzy CoCo: a cooperative coevolutionary approach to fuzzy modeling. IEEE T. Fuzzy Syst. 9(5), 727–737 (2001)

Sanz, J., Fernández, A., Bustince, H., Herrera, F.: Improving the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets and genetic amplitude tuning. Inform. Sciences 180(19), 3674–3685 (2010)

Sheskin, D.: Handbook of parametric and nonparametric statistical procedures, 2nd edn. Chapman & Hall/CRC (2006)

Su, C.T., Hsiao, Y.H.: An evaluation of the robustness of MTS for imbalanced data. IEEE T. Knowl. Data En. 19(10), 1321–1332 (2007)

Su, C.T., Chen, L.S., Yih, Y.: Knowledge acquisition through information granulation for imbalanced data. Expert Syst. Appl. 31, 531–541 (2006)

Sun, Y., Kamel, M.S., Wong, A.K., Wang, Y.: Cost-sensitive boosting for classification of imbalanced data. Pattern Recogn. 40, 3358–3378 (2007)

Sun, Y., Wong, A.K.C., Kamel, M.S.: Classification of imbalanced data: A review. Int. J. Pattern Recogn. 23(4), 687–719 (2009)

Thrift, P.: Fuzzy logic synthesis with genetic algorithms. In: 4th International Conference on Genetic Algorithms (ICGA 1991), pp. 509–513. Morgan Kaufmann, San Mateo (1991)

Wang, L.X., Mendel, J.M.: Generating fuzzy rules by learning from examples. IEEE T. Syst. Man CYB 25(2), 353–361 (1992)

Weiss, G., Provost, F.: Learning when training data are costly: The effect of class distribution on tree induction. J.Artif. Intell. Res. 19, 315–354 (2003)

Wilcoxon, F.: Individual comparisons by ranking methods. Biometrics 1, 80–83 (1945)

Wu, G., Chang, E.: KBA: Kernel boundary alignment considering imbalanced data distribution. IEEE T. Knowl. Data En. 17(6), 786–795 (2005)

Xu, L., Chow, M.Y., Taylor, L.S.: Power distribution fault cause identification with imbalanced data using the data mining-based fuzzy classification e-algorithm. IEEE T. Power Syst. 22(1), 164–171 (2007)

Yager, R.R., Filev, D.: Essentials of Fuzzy Modeling and Control. John Wiley & Sons, Chichester (1994)

Yang, Q., Wu, X.: 10 challenging problems in data mining research. Int. J. Inf. Tech. Decis. 5(4), 597–604 (2006)

Yen, J., Wang, L., Gillespie, C.W.: Improving the interpretability of TSK fuzzy models by combining global learning and local learning. IEEE T. Fuzzy Syst. 6(4), 530–537 (1998)

Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning. Inform. Sciences 8(3), 199–249 (1975)

Zhou, Z., Liu, X.: Training cost-sensitive neural networks with methods addressing the class imbalance problem. IEEE T. Knowl Data En. 18(1), 63–77 (2006)