



**HAL**  
open science

## **An extension of the FURIA classification algorithm to low quality data through fuzzy rankings and its application to the early diagnosis of dyslexia**

Ana Palacios, Luciano Sánchez, Ines Couso, Sébastien Destercke

### ► **To cite this version:**

Ana Palacios, Luciano Sánchez, Ines Couso, Sébastien Destercke. An extension of the FURIA classification algorithm to low quality data through fuzzy rankings and its application to the early diagnosis of dyslexia. *Neurocomputing*, 2016, 176, pp.60-71. 10.1016/j.neucom.2014.11.088 . hal-01254306

**HAL Id: hal-01254306**

**<https://hal.science/hal-01254306>**

Submitted on 12 Jan 2016

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

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

# An extension of the FURIA classification algorithm to low quality data through fuzzy rankings and its application to the early diagnosis of dyslexia

Ana Palacios

*Computer Science Department. Granada University. 33204 Gijon. Spain*

Luciano Sánchez

*Computer Science Department. Oviedo University. 33204 Gijon. Spain*

Inés Couso

*Statistics Department. Oviedo University. 33204 Gijon. Spain*

Sébastien Destercke

*UMR CNRS 6599 Heudysiac. France*

---

## Abstract

An early detection and reeducation of dyslexic children is critical for their integration in the classroom. Parents and instructors can help the psychologist to detect potential cases of dyslexia before the children's writing age. Artificial intelligence tools can also assist in this task. Dyslexia symptoms are detected with tests whose results may be vague or ambiguous, thus machine learning techniques for low quality data are advised. In particular, in this paper it is suggested that a new extension to vague datasets of the classification algorithm FURIA (Fuzzy Unordered Rule Induction Algorithm) has advantages over other approaches in both the computational effort during the learning stage and the linguistic quality of the induced classification rules. The new approach is benchmarked with different test problems and compared to other artificial intelligence tools for dyslexia diagnosis in the literature.

*Keywords:* Dyslexia, Low Quality Data, Fuzzy Rule Based Classifiers, FURIA

---

## 1. Introduction

The term “dyslexia” was coined in 1887 by R. Berlin [2], who stated that the condition belongs to the group of aphasia and is closely related to wordblindness. The

---

*Email addresses:* palacios@decsai.ugr.es (Ana Palacios), luciano@uniovi.es (Luciano Sánchez), couso@uniovi.es (Inés Couso), sdestercke@gmail.com (Sébastien Destercke)

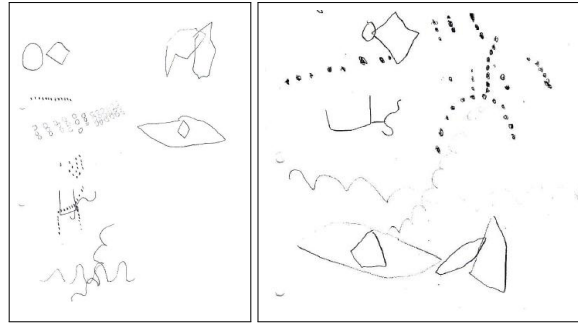


Figure 1: Left: Bender's test of one child assigned to the class "no dyslexia". Right: The same test solved by a child for whom the expert could not decide between the classes "no dyslexia" and "control and revision".

term implies that the condition has as its characteristic a difficulty with reading, and express that the cause is a physical disease of the brain, with suspected localizations in the left hemisphere. S. Orton [20] also suggested a faulty patterning of the brain function, and believed that the phenomena he observed in children's reading lay a basic state of ambiguous hemispheric dominance, physiological in nature. Notwithstanding this, the modern concept of dyslexia is not restricted to a neurological condition and various theories exist that concern minimal cerebral dysfunction, hereditary involvement, developmental lag or emotional causation [34]. According to [30], dyslexia is a common learning disorder that manifests itself as a difficulty for the distinction and memorization of letters, lack of order and rhythm in the placement and poor structuring of sentences, affecting both reading and writing. Dyslexia may also be defined as the learning difficulties of people whose IQ is normal and do not have physical or psychological problems that may explain these difficulties.

An early prediction of reading problems is needed for a proper therapy [19]. Reading difficulties and bad understanding lead to poor school performance, low self-esteem, and attitudes and behaviors that may affect the classroom. There are, however, different types of syndromes related to dyslexia, such as hyperactivity, attention deficit disorder or dysgraphia. They all share a certain degree of similarity, but differ in the acquisition of certain processes, such as identification, recognition or understanding. Failures or deficit in each of these processes produce a variety of different problems. Identifying the specific learning disorder is important, because each syndrome has different reeducation techniques and a different evolution in time. A predictive screening model is customarily used that is based on a large pool of tests (see Table 1). As an example, in Figure 1 two examples of the children's answers to the visual-motor coordination test "Bender" [1] are given. The psychologist scores these tests following a list of "if-then" rules measuring how well the child draws what he sees. There is a high degree of subjectivity in this evaluation, and it is possible that two experts assign different scores to the same drawing.

It is not mandatory in Spain that children attend school before the age of six, thus the role of the parents is crucial to detect cases before the children's writing age.

Category	Test	Description
Verbal comprehension	BAPAE[7] BADIG[36] BOEHM[3]	Vocabulary Verbal orders Basic concepts
Logic reasoning	RAVEN [29] BADIG[36] ABC[10]	Color Figures Actions and details
Memory	Digit WISC-R[35] BADIG[36] ABC[10]	Verbal-additive memory Visual memory Auditive memory
Level of maturation	ABC[10]	Combination of different tests
Sensory-motor skills	BENDER[1] ABD[11] BADIG[36] BAPAE[7] STAMBACK[31] HARRIS/HPL[13] ABC[10] GOODENOUGHT[12]	Visual-motor coordination Motor Coordination Perception of shapes Spatial relations, shapes, orientation Auditive perception, rhythm Laterality Spelling Spatial orientation, body scheme
Attention	Toulouse[32] ABC[10]	Attention and fatigability Attention and fatigability
Reading-writing	TALE[33]	Analysis of reading and writing

Table 1: Most often used tests in Spanish schools for predictive screening of dyslexia.

Preschool instructors can also help with the task. In both cases, artificial intelligence (AI) techniques are valuable, as shown in [22, 24]. Parents and instructors can efficiently use AI methods for helping the psychologist, and there exist software tools that implement methods for detecting potentially affected infants [26, 27].

Machine learning techniques for low quality data are the most appropriate for this activity, as dyslexia symptoms are detected with tests whose results may be vague or ambiguous. Apart from this requirement, it is also desired that the AI method of choice elicits a Knowledge Base (KB) with a high degree of linguistic understandability thus a psychologist can endorse it. Furthermore, the number of children used to learn the KB is still small in this study but is expected to grow when the tool is publicited. A fast learning algorithm is needed that can cope with large datasets in the future.

Past algorithms were Genetic Cooperative-Competitive Learning (GCCL) [22] and Boosting of individual fuzzy rules [24] (see Appendix B). Both of them have advantages and drawbacks. On the one hand, GCCL uses the standard fuzzy logic inference. The evolved rules are understandable and the inference process is intuitive. Unfortunately, GCCL routinely produces KBs comprising hundreds of rules and the learning is very slow. On the other hand, Boosting is fast and the resulting KBs are accurate, but weighed rules and additive inference are used. This setup is counterintuitive for the expert, that cannot grasp the meaning of even the simplest KBs.

Because of these reasons, in this paper it is suggested to use the state-of-the-art

```

Procedure FURIA()
  Select a class and learn crisp classification rules discriminating
    this class from the others (call local procedure RuleSetForOneClass())
  Remove redundant antecedents
  Fuzzify rules maximizing the purity of the fuzzification of each attribute
  Compute confidence degrees for all rules considering the certainty factor
  Evaluate rules and apply rule stretching if there are uncovered examples
Local Procedure RuleSetForOneClass()
  While StoppingConditions() == false do
    Call method RuleGrowing()
    If StoppingConditions() == true then
      Delete the newly created rule
    End If
  End While
  Perform rule pruning.
End of local procedure
End of Procedure

```

Figure 2: Outer loop of the FURIA algorithm

classification algorithm “Fuzzy Unordered Rule Induction Algorithm” (FURIA), described in [16, 18] and extended to low quality data in [25]. This study is based on an extended version of this last reference. FURIA is expected to keep the good numerical properties of Boosting and at the same time to produce much smaller KBs than GCCL. The inference procedure demanded by FURIA-based KBs is not standard either, but it is much simpler than that of Boosting and deemed suitable for the task at hand.

This paper is organized as follows: Section 2 introduces the FURIA algorithm and remarks the parts that are more relevant for this proposal. In the same section, the changes effected to this algorithm are detailed, paying special attention to the new definition of the ranking between fuzzy intervals. In Section 3, numerical results are given. Concluding remarks and future work are discussed in Section 4.

## 2. An extension of the FURIA algorithm to low quality data

Fuzzy Unordered Rules Induction Algorithm (FURIA) [16, 18, 17] is a novel fuzzy rule-based classification method extending the classical RIPPER [5]. For the convenience of the reader, an algorithmic description of FURIA is included in Figures 2 and 3, where the parts that will be altered in the generalization are marked in boldface.

The algorithmic schema needs not to be altered in order to introduce vague data, however all expressions that depend on a count of the number of instances have to be rewritten, as generally speaking this count will become a fuzzy interval. A new fuzzy ranking, defined in Section 2.2, will be used to perform the comparisons, in particular:

1. Ranking (sorting) the instances according to the value of a fuzzy attribute.
2. Ranking the fuzzy values of the information gain.

```

Method RuleGrowing()
  Grow rule using an information gain measure to choose the best conjunct
  to be added into the rule antecedent.
  Stop adding conjuncts when the rule starts covering negative instances.
End of method

Method StoppingConditions()
  If there are not uncovered instances of the current class
  then StoppingConditions=true
  If rule error  $\geq 0.5$  then StoppingConditions=true
  If the description length of the ruleset is 64 bits greater than
  the smallest found then StoppingConditions=true
  StoppingConditions=false
End of method

```

Figure 3: Subroutines called in the main loop of FURIA

Apart from these, the following operations take fuzzy values:

1. Computing the rule purity, that quantifies the quality of the fuzzification procedure, depending on the number of partially covered examples.
2. Computing the certainty factor, that measures the confidence assigned to the piece of information described by the rule.
3. The rule stretching procedure, that is used to simplify the antecedents for improving generalization, depending on the number of examples covered by the rule.

Since the algorithm is not being altered, this section is organized into two parts. First, in Subsection 2.1, “Notation”, a listing is provided with the definitions of those parts of the original FURIA algorithm that will involve computing with fuzzy intervals in the extended version. Second, in Subsection 2.2 each of these parts is redefined. In addition to this, in the same section 2.2 the logical operator “higher than” between fuzzy intervals (that becomes a fuzzy ranking) is introduced. It is remarked that the rankings found in the literature are not coherent with the statistical interpretation of a fuzzy set used in this study and therefore a new ranking is being proposed.

### 2.1. Notation

- **Training set:** The training set is  $D \subset \mathbb{R}^d$  whose instances are vectors

$$x = (x_1, \dots, x_d) \in D. \quad (1)$$

- **Antecedent:** Each antecedent of a FURIA fuzzy classification rule is a multivariate trapezoidal fuzzy set whose membership is

$$I^F(x) = \bigwedge_{i=1, \dots, d} I_i^F(x_i) \quad (2)$$

and its core is the interval  $I = I_1 \times \dots \times I_d$ , where the indicator function of  $I_i$ ,  $i = 1, \dots, d$  is

$$I_i(x_i) = \begin{cases} 1 & \text{if } I_i^F(x_i) = 1 \\ 0 & \text{else.} \end{cases} \quad (3)$$

- **Information gain:** This criterion measures the improvement of a rule with respect to the default for the target class and is used as a stopping condition in the rule growing procedure. Let  $I$  be the core of the antecedent of the rule at hand, and let  $l$  be the target class. Then, the number of positive examples for the fuzzy classification rule  $r$  is

$$p_r = \#\{x \in I \mid \text{class}(x) = l\} \quad (4)$$

and the number of negative examples for that rule is

$$n_r = \#\{x \in I \mid \text{class}(x) \neq l\}. \quad (5)$$

The total number of positive and negative examples in the dataset are named  $p$  and  $n$ , respectively. Then, the information gain is defined as follows [17]:

$$\text{IG}_r = p_r \times \left( \log_2 \left( \frac{p_r + 1}{p_r + n_r + 1} \right) - \log_2 \left( \frac{p + 1}{p + n + 1} \right) \right). \quad (6)$$

- **Pruning:** Rules comprise  $q$  antecedents  $\langle a_1, \dots, a_q \rangle$  combined with the AND operator. The order of the antecedents reflects their importance thus pruning a rule consists of selecting a sublist  $\langle a_1, \dots, a_i \rangle$ , with  $i \leq q$ . In order to find a suitable value for  $i$ , the following rule-value metric is computed first [17]:

$$V_r = \frac{p_r + 1}{p_r + n_r + 2} \quad (7)$$

Let the number of positive covered and negative uncovered examples of the rule, when pruned at the  $i$ -th antecedent, respectively be  $P_i$  and  $N_i$ :

$$P_i = \#\{x \mid x \text{ is covered by } \langle a_1, \dots, a_i \rangle \wedge \text{class}(x) = l\} \quad (8)$$

$$N_i = \#\{x \mid x \text{ is not covered by } \langle a_1, \dots, a_i \rangle \wedge \text{class}(x) \neq l\}. \quad (9)$$

and let be defined the value [17]

$$\text{worth}_i = \frac{P_i + N_i}{p + n}. \quad (10)$$

This value measures how likely is each antecedent to be pruned. If

$$\max_{i=1, \dots, q} \text{worth}_i > V_r, \quad (11)$$

then the term where the value of “ $\text{worth}_i$ ” is maximum is selected for pruning.

- **Purity:** This value measures the quality of the fuzzification procedure and it is used for determining the support of the fuzzy sets defining the rule antecedents. Let  $D^i$  be the subset of the training data that follows:

$$D^i = \{(x_1, \dots, x_d) \mid x_j \in \text{supp}(I_j^F) \text{ for all } j \neq i\}. \quad (12)$$

$D^i$  is partitioned into positive and negative instances,  $D_+^i$  and  $D_-^i$ . Given the values

$$p_i = \sum_{x \in D_+^i} I_i^F(x_i) \quad (13)$$

$$n_i = \sum_{x \in D_-^i} I_i^F(x_i), \quad (14)$$

the purity of the fuzzification of the  $i$ -th attribute is [17]:

$$\text{pur}_i = \frac{p_i}{p_i + n_i} \quad (15)$$

- **Certainty factor:** The certainty factor CF of a rule  $\langle I^F, l \rangle$ , for a training set  $D_T$ , is [17]:

$$\text{CF} = \frac{2 \cdot \frac{\sum_{x \in D_T, \text{class}(x)=l} p(x)}{\sum_{x \in D_T} p(x)} + \sum_{x \in D_T, \text{class}(x)=l} I^F(x)}{2 + \sum_{x \in D_T} I^F(x)} \quad (16)$$

where  $p(x)$  is the weight of instance  $x$ .

- **Rule stretching:** Rule stretching (or generalization) deals with uncovered examples (those classified by the default rule in RIPPER). The generalization procedure consists of making (preferably minimal) simplifications of the antecedents of the rules until the query instance is covered. The instance is then classified by the rule with the highest evaluation, according to the value [17]

$$\text{STR} = \text{CF} \cdot \frac{d+1}{m+2} \cdot I^F(x) \quad (17)$$

where  $k$  is the size of the generalized antecedent and  $m$  is the size of the entire antecedent before applying this procedure. Notice that,  $\frac{d+1}{m+2}$  aims at discarding heavily pruned rules. If no stretched rule is able to cover the given example  $x_i$ , it is assigned a class based on the *a priori* distribution.



## 2.2. Extended definitions

In this study, fuzzy memberships are used for two purposes. On the one hand, fuzzy rule-based systems will be used for building a classifier that inputs the scores of the tests and predicts the learning disorder of the child. The antecedents of the rules in this classifier are fuzzy sets, whose membership functions are learned from data. These can be given a linguistic meaning under the premises of the fuzzy logic theory. On the other hand, the uncertainty in the scores of the tests will be represented by means of fuzzy intervals, but these will be assigned an statistical meaning.

Following [6], the statistical view of a fuzzy set as a nested random set is adopted for this second purpose. Fuzzy intervals  $\tilde{X}$  and  $\tilde{Y}$  on quantities  $X, Y \in \mathbb{R}^2$  are expressed by their equivalent possibility distributions  $\pi_X$  and  $\pi_Y$ . Given a possibility distribution  $\pi : \mathbb{R} \rightarrow [0, 1]$ , the possibility measure over any set  $A \subseteq \mathbb{R}$  is

$$\Pi(A) = \sup_{x \in A} \pi(x). \quad (18)$$

Let  $\mathcal{P}(\Pi)$  be the set of probabilities bounded by  $\Pi$ , such that

$$\mathcal{P}(\Pi) = \{P \in \mathbb{P}_{\mathbb{R}} | P(A) \leq \Pi(A)\} \quad (19)$$

with  $\mathbb{P}_{\mathbb{R}}$  the set of all probabilities on the real line. Within this view, a fuzzy set  $\pi$  can also be described by an equivalent random set  $\Gamma_{\pi}$  that maps the uniform distribution  $\mathcal{U}_{[0,1]}$  to the alpha-cuts, i.e.  $\Gamma_{\pi} : [0, 1] \rightarrow \mathbb{R}$  with  $\Gamma_{\pi}(\alpha) = \pi_{\alpha}$ . Therefore,

$$\Pi(A) = P_{\mathcal{U}_{[0,1]}}(\{\alpha | \Gamma_{\pi}(\alpha) \cap A \neq \emptyset\})$$

Similarly, lower and upper expectation bounds  $\underline{\mathbb{E}}_{\pi}(X)$  and  $\overline{\mathbb{E}}_{\pi}(X)$  reached within the set  $\mathcal{P}(\Pi)$  can be associated to a fuzzy set  $\pi_X$ . Such bounds are simply expressed as

$$\underline{\mathbb{E}}_{\pi}(X) = \int_{\mathbb{R}} x d\overline{F}_{\pi}(x) = \int_0^1 \underline{\pi}_{\alpha} d\alpha \quad (20)$$

and

$$\overline{\mathbb{E}}_{\pi}(X) = \int_{\mathbb{R}} x dF_{\pi}(x) = \int_0^1 \overline{\pi}_{\alpha} d\alpha. \quad (21)$$

Also note that, given a function  $f$  on  $\mathbb{R}$ , the lower and upper expectations read  $\underline{\mathbb{E}}_{\pi}(f) = \int_0^1 \inf_{x \in \pi_{\alpha}} f(x) d\alpha$  and  $\overline{\mathbb{E}}_{\pi}(f) = \int_0^1 \sup_{x \in \pi_{\alpha}} f(x) d\alpha$ , and the upper/lower expectations of the indicator function  $\mathbf{1}_{(A)}$  of an event  $A$  ( $\mathbf{1}_{(A)}(x) = 1$  if  $x \in A$ , zero otherwise) correspond to the upper/lower probabilities of  $A$ .

Under this statistical view of a fuzzy set, the concepts introduced in the preceding definitions are repurposed for vague data in the remaining of this section.

- **Logical operator “Higher than”.** There are multiple points in the extended algorithm where two fuzzy intervals must be compared. The extension of the logical operator “higher than” to vague data proposed in this section can be considered as a fuzzy version of the statistical preference criterion stating that  $X > Y$  if  $P(X > Y) > 0.5$ , providing a total ordering between fuzzy sets.

Given two random variables  $X$  and  $Y$  with probability distributions  $P_X$  and  $P_Y$ , a common way to assess whether  $X$  is higher or lower than  $Y$  is to use statistical preferences, that is to compute  $P_{X,Y}(X > Y)$  under independence and to declare that  $X > Y$  if  $P(X > Y) > 0.5$ . Such preferences have the advantage to provide a total ordering over random variables (as opposed to stochastic dominance), and to not be too sensible to the numerical values of  $X$  and  $Y$  (as opposed to expected values).

For any pair  $(x, y)$ , let us consider the comparison function  $I_{>}(x, y)$  such that

$$I_{>}(x, y) = \begin{cases} 1 & \text{if } x > y \\ 0 & \text{else.} \end{cases} \quad (22)$$

The probability  $P_{X,Y}(X > Y)$  is then the expectation of  $I_{>}(x, y)$  under  $P_{X,Y}$ . It is straightforward, using Equations 20 and 21, to extend the value  $P_{X,Y}(X > Y)$  to the case of non-interactivity, namely we have

$$\underline{P}(X > Y) = \int_{0,1} \min_{\substack{x \in \pi_{X,\alpha} \\ y \in \pi_{Y,\alpha}}} f_{>}(x, y) d\alpha = \int_{0,1} \mathbf{1}_{(\underline{x}_{X,\alpha} > \bar{y}_{Y,\alpha})} d\alpha \quad (23)$$

and

$$\bar{P}(X > Y) = \int_{0,1} \max_{\substack{x \in \pi_{X,\alpha} \\ y \in \pi_{Y,\alpha}}} f_{>}(x, y) d\alpha = \int_{0,1} \mathbf{1}_{(\bar{x}_{X,\alpha} > \underline{y}_{Y,\alpha})} d\alpha \quad (24)$$

where  $\mathbf{1}_{(A)}$  is the indicator function of some event  $A$ . Note that, if possibilities  $\pi_X, \pi_Y$  are equivalent to intervals  $[\underline{x}, \bar{x}]$  and  $[\underline{y}, \bar{y}]$  ( $\pi_X, \pi_Y \in \{0, 1\}$ ), then  $\underline{P}(X > Y) = 1$  if  $\underline{x} > \bar{y}$ , 0 else hence coming back to check interval dominance, while  $\bar{P}(X > Y) = 1$  if  $\bar{x} > \underline{y}$ , hence coming back to check for a weak ordering between  $[\underline{x}, \bar{x}]$  and  $[\underline{y}, \bar{y}]$ .

It is then natural to extend condition  $P(X > Y) > 0.5$  to condition  $\underline{P}(X > Y) > 0.5$  to declare  $\tilde{X} > \tilde{Y}$ . Note that in our case this condition only provides a partial ordering, as we may have  $\underline{P}(X > Y) \leq 0.5$  and  $\underline{P}(Y > X) \leq 0.5$ , since only the duality  $\underline{P}(X > Y) = 1 - \bar{P}(Y \leq X)$  holds. Additional properties of this ranking, along with a refinement that provides a total order, are detailed in Appendix A.

- **Low Quality Dataset:** Let  $D_T = \{(\tilde{X}_1, Z_1), \dots, (\tilde{X}_n, Z_n)\}$  be a set of vague data, where  $n$  is the number of instances,  $\tilde{X}_i = (\tilde{X}_{i1}, \dots, \tilde{X}_{id})$ ,  $\tilde{X}_{ij} \in \mathcal{F}(\mathcal{R})$  and  $Z_i \subset C = \{c_1, \dots, c_m\}$  for  $i = 1, \dots, n$ ,  $j = 1, \dots, d$ .
- **Number of instances of a given class:** The number of instances  $\bar{f}_{c_j}$  of class  $c_j$  and the relative frequencies of classes  $\bar{r}_{c_j}$  are:

$$\bar{f}_{c_j} = \sum_{i=1}^n \bar{\delta}_{c_j, Z_i} \quad (25)$$

where

$$\bar{\delta}_{a,A} = \{\delta_{a,b} : b \in A\} = \begin{cases} 1 & \{a\} = A, \\ 0 & a \notin A, \\ [0, 1] & \text{else,} \end{cases} \quad (26)$$

and

$$\overline{fr}_{c_j} = \frac{\sum_{i=1}^n \overline{\delta}_{c_j, Z_i} p_i}{\sum_{i=1}^n p_i} = \frac{\sum_{i=1}^n \overline{P}_{c_j}^i}{\sum_{i=1}^n p_i} \quad (27)$$

where  $p_i$  are the weights of the instances and

$$\overline{P}_{c_j}^i = \begin{cases} p_i & c_j = Z_i, \\ 0 & c_j \notin Z_i, \\ [0, p_i] & \text{else.} \end{cases} \quad (28)$$

- **Class being processed:** The class being processed  $c_j$  is defined by the default number of correct classifications:

$$\overline{\text{defAcRT}}_{c_j} = \frac{1 + \overline{\text{defAccu}}_{c_j}}{1 + \sum_{i=1}^n p_i} \quad (29)$$

where

$$\overline{\text{defAccu}}_{c_j} = \sum_{i=1}^n \overline{\delta}_{c_j, Z_i} p_i = \sum_{i=1}^n \overline{P}_{c_j}^i. \quad (30)$$

The crisp stopping criterion when creating rules of class  $c_j$  was  $1 - \overline{\text{defAcRT}}_{c_j} > \text{Threshold}$ , whose extension is

$$\begin{aligned} P(\overline{\text{defAcRT}}_{c_j} > \text{Threshold}) &> 0.5 = \\ P([a_{c_j}, b_{c_j}] > \text{Threshold}) &> 0.5 = \\ = P([a_{c_j} - \text{Threshold}, b_{c_j} - \text{Threshold}] > 0) &> 0.5 \end{aligned} \quad (31)$$

- **Information gain:** The numbers of positive and negative examples are imprecise. Let  $I$  be the core of the antecedent of the rule at hand, and let  $c_l$  be the target class. Then, the information gain is a fuzzy number whose  $\alpha$ -cuts are defined as follows:

$$[\overline{IG}_r]_\alpha = \sum_{[\tilde{x}_i]_\alpha \subseteq I} \overline{P}_{c_l}^i \cdot (\log_2(\overline{\text{fstAccuRate}}_{c_l}) - \log_2(\overline{\text{defAcRT}}_{c_l})) \quad (32)$$

where

$$\overline{\text{fstAccuRate}}_{c_l} = \frac{1 + \overline{\text{fstAccu}}_{c_l}}{1 + \overline{\text{Coverfst}}_{c_l}} \quad (33)$$

$$\overline{\text{fstAccu}}_{c_l} = \sum_{[\tilde{x}_i]_\alpha \subseteq I} \overline{P}_{c_l}^i \quad (34)$$

$$\overline{\text{Coverfst}}_{c_l} = \sum_{[\tilde{x}_i]_\alpha \subseteq I} p_i \quad (35)$$

- **Rule pruning:** This consists in finding the position in the antecedent list of the rule  $\langle a_1, \dots, a_q \rangle$  with  $i \leq q$  where the rule must be split, according to the following criteria:

1. The value defined in Eq. 7 is extended as follows:

$$\bar{V}(r) = \frac{1 + \overline{\text{defAccu}}_{c_j}}{2 + \sum_{i=1}^n p_i} \quad (36)$$

2. For each antecedent, the number of positive covered instances  $\tilde{\text{Pos}}_{a_m}$  and negative uncovered instances  $\tilde{\text{Neg}}_{a_m}$  are:

$$[\tilde{\text{Pos}}_{a_m}]_\alpha = \sum_{[\tilde{X}_i]_\alpha \text{ is covered by } \langle a_1, \dots, a_m \rangle} \bar{P}_{c_j}^i \quad (37)$$

$$[\tilde{\text{Neg}}_{a_m}]_\alpha = \sum_{[\tilde{X}_i]_\alpha \text{ is not covered by } \langle a_1, \dots, a_m \rangle} \bar{P}_{c_j}^i \quad (38)$$

where

$$\bar{P}_{c_j}^i = \begin{cases} p_i & (\{c_j\} \neq Z_i \text{ and } \#Z_i = 1) \text{ or } (c_j \notin Z_i), \\ [0, p_i] & c_j \in Z_i \text{ and } c_j \neq Z_i. \end{cases} \quad (39)$$

3. The net worth of each antecedent is

$$\tilde{\text{worth}}_{a_m} = \frac{\tilde{\text{Pos}}_{a_m} + \tilde{\text{Neg}}_{a_m}}{\sum_{i=1}^n p_i}, \quad (40)$$

and this last value is used to decide the splitting position, as shown below.

4. If  $\tilde{\text{worth}}_{a_m}$ , with  $m = 1, \dots, q$ , precedes  $\bar{V}(r)$ , the splitting point is the  $m$ -th antecedent.

- **Purity:** Let  $D_\alpha^i$  be the following subset of the training data:

$$D_\alpha^i = \{k \mid [\tilde{X}_{kj}]_\alpha \subseteq \text{supp}(I_j^F) \text{ for all } j \neq i\}. \quad (41)$$

Each set  $D_\alpha^i$  is partitioned into positive and negative instances,  $D_{\alpha+}^i$  and  $D_{\alpha-}^i$ . Given the values

$$p_{i\alpha} = \sum_{k \in D_{\alpha+}^i} \sup\{I_i^F(x) \mid x \in [\tilde{X}_{ki}]_\alpha\} \quad (42)$$

$$n_{i\alpha} = \sum_{k \in D_{\alpha-}^i} \sup\{I_i^F(x) \mid x \in [\tilde{X}_{ki}]_\alpha\} \quad (43)$$

the purity of the fuzzification of the  $i$ -th attribute is:

$$\text{pur}_i = \int_0^1 \frac{p_{i\alpha}}{p_{i\alpha} + n_{i\alpha}} d\alpha \quad (44)$$

Table 2: Summary descriptions of the LQD benchmarks.

Dataset	Ex.	Atts.	Classes	%Classes
B200mlI [23]	19	4	2	((0.47,0.73],[0.26,0.52])
B200mlP [23]	19	5	2	((0.47,0.73],[0.26,0.52])
Long [23]	25	4	2	([36,64],[36,64])
BLong [23]	25	4	2	([36,64],[36,64])
100mlI [23]	52	4	2	((0.44,0.63],[0.36,0.55])
100mlP [23]	52	4	2	((0.44,0.63],[0.36,0.55])
B100mlI [23]	52	4	2	((0.44,0.63],[0.36,0.55])
B100mlP [23]	52	4	2	((0.44,0.63],[0.36,0.55])
Ice7-6 [4]	42	7	3	((0.47,0.54],[0.19,0.30],[0.21,0.26])
Ice7-4 [4]				
Ice8-7 [4]	42	8	3	((0.47,0.54],[0.19,0.30],[0.21,0.26])
Ice8-5 [4]				
Ice-shedding [4]	42	7	2	((0.47,0.54],[0.46,0.53])
Car [8]	33	8	4	(0.30,0.242,0.242,0.212)
Water_4 [14, 28]	316	4	2	(0.705,0.294)

- **Certainty factor:** The certainty factor of a rule  $\langle I^F, l \rangle$  is a fuzzy set, defined by its level cuts

$$[\tilde{C}F]_\alpha = \frac{2 \cdot \frac{\sum_{i=1, \dots, n} \bar{P}_{c_j}^i}{\text{class}(\tilde{X}_i)=l} + \sum_{i=1, \dots, n} \bar{P}_{c_j}^i \cdot \sup\{I^F(x) \mid x \in [\tilde{X}_i]_\alpha\}}{\sum_{i=1}^n p_i} + \sum_{i=1}^n p_i \cdot \sup\{I^F(x) \mid x \in [\tilde{X}_i]_\alpha\}}{2 + \sum_{i=1}^n p_i \cdot \sup\{I^F(x) \mid x \in [\tilde{X}_i]_\alpha\}} \quad (45)$$

- **Rule stretching:** The extension of Eq. 17 is straightforward:

$$[S\tilde{T}R]_\alpha = [\tilde{C}F]_\alpha \cdot \frac{d+1}{m+2} \cdot \sup\{I^F(x) \mid x \in [\tilde{X}_i]_\alpha\} \quad (46)$$

### 3. Numerical results

In this section the extension of FURIA to LQD is assessed first with benchmark problems. The second part of this section describes a case study of dyslexia diagnosis in schools of Asturias (Spain).

#### 3.1. Assessment of FURIA-LQD with benchmark problems

The datasets “Athleticism at Oviedo University” [23], “Ice adhesion strength” [4], “Car” [8], and “Barcelona’s water distribution” [14, 28] are used to compare the proposed method to other approaches. The main characteristics of these datasets are summarized in Table 2, where “Ex.” represents the number of examples, “Att.” is the number of attributes, “Classes” is the number of classes, and “%Classes” is the fraction of patterns of each class. All these datasets are available in the repository <https://ccia35.edv.uniovi.es/datasets>. All experiments were repeated 100

Table 3: Behaviour of GCCL-LQD [22], Boost-LQD [24] and the new proposal (FURIA-LQD) in several datasets.

	<b>GCCL-LQD</b>	<b>Boost-LQD</b>	<b>FURIA-LQD</b>
	$Acc_{Tst}$	$Acc_{Tst}$	$Acc_{Tst}$
<b>Athleticism at Oviedo University</b>			
100mlP	[0.640,0.824]	[0.642,0.820]	[0.601,0.780]
100mlI	[0.622,0.824]	[0.624,0.830]	[0.592,0.793]
B100mlP	[0.651,0.840]	[0.650,0.839]	[0.609,0.798]
B100mlI	[0.631,0.828]	[0.644,0.842]	[0.568,0.766]
B200mlP	[0.520,0.738]	[0.594,0.812]	[0.590,0.808]
B200mlI	[0.527,0.768]	[0.585,0.829]	[0.564,0.809]
Long	[0.410,0.679]	[0.492,0.760]	[0.528,0.746]
BLong	[0.375,0.674]	[0.470,0.770]	[0.464,0.764]
<b>Ice adhesion strength</b>			
Ice7-6	[0.512,0.596]	[0.548,0.631]	[0.526,0.606]
Ice7-4	[0.522,0.597]	[0.572,0.651]	[0.528,0.602]
Ice8-7	[0.525,0.597]	[0.554,0.628]	[0.545,0.621]
Ice8-5	[0.507,0.566]	[0.550,0.618]	[0.534,0.608]
Ice-shedding	[0.550,0.619]	[0.639,0.708]	[0.713,0.782]
<b>Car</b>			
Car	[0.389,0.389]	[0.436,0.436]	[0.608,0.608]
<b>Barcelona's water distribution</b>			
Water_4	[0.713,0.713]	[0.602,0.602]	[0.651,0.651]
<b>Mean</b>	[0.527,0.670]	[0.559,0.702]	[0.566,0.709]

times from bootstrap resamples of the training set. The test set comprises “out of the bag” elements. Each test partition is repeated 1000 times for different random crisp selections. GCCL-LQD [22] and Boosting-LQD [24] were configured with a population of size 100, crossover probability 0.9, mutation probability 0.1, 200 generations, 5 labels/variable, uniform fuzzy partitions.

In Table 3, the expected test errors of each combination of algorithm and dataset are provided. Each of these values is an interval, computed by means of Eqs. (20) and (21), i.e. the expectation of the fuzzy test error whose corresponding possibility distribution is  $\pi$ , is the interval

$$[\mathbb{E}_{\pi}, \bar{\mathbb{E}}_{\pi}] = \left[ \int_0^1 \underline{\pi}_{\alpha} d\alpha, \int_0^1 \bar{\pi}_{\alpha} d\alpha \right]. \quad (47)$$

It is shown that the improvement over previous approaches for the most complex problem (Car) is remarkable. The performance of the new classifier improves that of Boosting in most of the problems with a high ratio between the number of features and instances (Ice-shedding, Car, Water), however FURIA does not seem to improve the accuracy of Boosting for Ice-7, Ice-8 and the low-dimensional problem “Athleticism”. These differences are not statistically relevant, thus it can be assumed that the accuracies of FURIA and Adaboost are roughly the same. However,

- FURIA is much faster than Boosting. The combined learning + validation time of FURIA was about 12 times faster than Adaboost or GCCL for the datasets mentioned in this section (5 minutes vs. 1 hour in some cases, see Figure 4). FURIA is the alternative of choice when computational resources are limited.
- The number of rules of the classifiers produced by FURIA are also much lower. The highest number of rules produced by this algorithm for the studied datasets

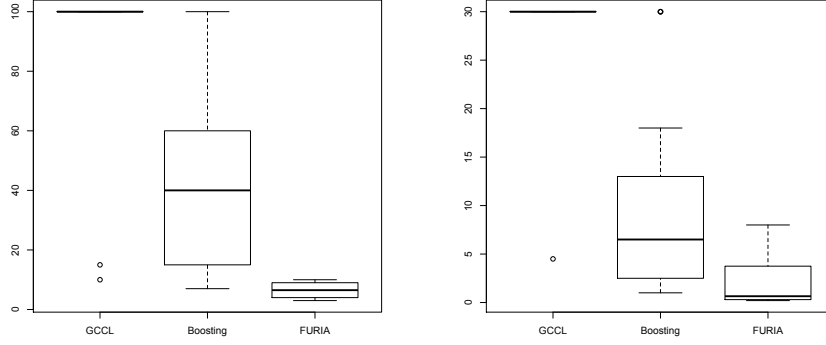


Figure 4: Left: Number of rules produced by GCCL, Boosting and FURIA for the datasets in Table 3. Right: Learning time (minutes) of the three algorithms.

Table 4: Null hypothesis: The expected error is the same than the average.

Conditions	GCCL-LQD	BOOST-LQD	FURIA-LQD
$[q-a, q+a]$	[0.527,0.670]	[0.559,0.702]	[0.566,0.709]
$F_{(q-a)}^* \in$	[0,0.025]	[0,0.83]	[0,0.93]
$F_{(q+a)}^* \in$	[0.025,1]	[0.85,1]	[0.94,1]
	INCONCLUSIVE	INCONCLUSIVE	INCONCLUSIVE

was of 15, while boosting and GCCL obtain knowledge bases comprising hundreds of rules for the largest problems (see Figure 4). That is to say, the linguistic quality of the results of FURIA is much better. It is also remarked that the number of labels for each variable must be determined by trial and error in Boosting and GCCL, but FURIA determines this parameter automatically.

The differences in linguistic quality need not to be studied with statistical tests because FURIA-generated knowledge bases were uniformly smaller for all executions of the algorithm. The statistical relevance of the differences in accuracy is assessed with bootstrap tests for LQD, following the experimental design proposed in [21]. The null hypothesis of this test is that the average number of misclassifications for each dataset does not depend on the algorithm. In Table 4 is shown that the mentioned advantages of FURIA over Boosting for datasets with high ratio between the number of features and instances are compensated by the results for datasets with a lower dimensionality thus the differences are not significant for a 95% confidence level.

### 3.2. Case study: diagnostic of dyslexia

The answers to the tests in Figure 1 for 65 schoolchildren between 5 and 8 were collected in this two-year long experiment. Each child has been individually diagnosed by an expert psychologist and labelled with one or more of the terms “no dyslexia”, “control and revision”, “dyslexic” and “other disorders” (inattention, hyperactivity, etc.). In

addition to this, parents or instructors evaluated on their own the results of the tests, without further instructions from the psychologist (“non-expert” datasets). These two control sets are compared with the purpose of determining whether there are significant differences between the outputs of two classifiers trained with expert and non-expert data. The results are shown in Table 5. FURIA-LQD is significantly better (see Figure 5) specially for “expert” datasets (dyslexic-11-01 and dyslexic-11-12). The statistical analysis of these results are shown in Tables 6 and 7, that follow the experimental design suggested in [15] and [21]. Observe that FURIA-LQD is the best algorithm for all datasets.

Table 5: Success rate of GCCL-LQD [22], Boost-LQD [24] and FURIA-LQD in the dyslexia case study.

	<b>GCCL-LQD</b>	<b>Boost-LQD</b>	<b>FURIA-LQD</b>
	<i>Acc<sub>Test</sub></i>	<i>Acc<sub>Test</sub></i>	<i>Acc<sub>Test</sub></i>
<b>Datasets non-expert</b>			
Dyslexic-12-12	[0.356,0.522]	[0.351,0.497]	[0.472,0.644]
Dyslexic-12-01	[0.469,0.599]	[0.414,0.508]	[0.530,0.679]
<b>Dataset expert</b>			
Dyslexic-11-01	[0.427,0.555]	[0.433,0.553]	[0.576,0.729]
Dyslexic-11-12	[0.310,0.472]	[0.401,0.536]	[0.546,0.738]
<b>Dyslexic mean</b>	[0.390,0.537]	[0.4,0.524]	[0.531,0.698]

Table 6: Left: Rankings of the three considered algorithms. Right: Mean rankings for Friedman’s test

Dataset	GCCL-LQD.R	Boost-LQD.R	FURIA-LQD.R
Dyslexic-12-12	2	3	1
Dyslexic-12-01	2	3	1
Dyslexic-11-01	3	2	1
Dyslexic-11-12	3	2	1

Algorithm	Ranking
FURIA-LQD	1
GCCL-BOOST	2.5
BOOST-LQD	2.5

The differences between the success rate of the classifiers trained with “expert” and “non expert” datasets were not statistically relevant. This is a good result, because implies that an automated screening procedure is possible that does not require of the presence of the expert, but the success rate is nevertheless too low. Neither GCCL-LQD [22] nor Boost-LQD [24] are able to surpass a 60% success rate. FURIA-LQD is under 68%. Notwithstanding this, the purpose of the screening system is not to diagnose dyslexia but to anticipate possible disorders; positive classifications will be subjected to further tests. Indeed, this problem could have been regarded as an imbalanced classification problem, where misclassifying a dyslexic child is assigned a higher cost than the opposite error, but this would require a separate extension of the FURIA algorithm to cost-based LQD classification. A simpler approach is possible that takes advantage

Table 7: Wilcoxon’s test for comparing FURIA-LQD with GCCL-LQD and Boost-LQD

Comparison	p-value	Hypothesis
FURIA-LQD vs GCCL-LQD	1.5e-05	Reject
FURIA-LQD vs. Boost-LQD	1.5e-05	Reject



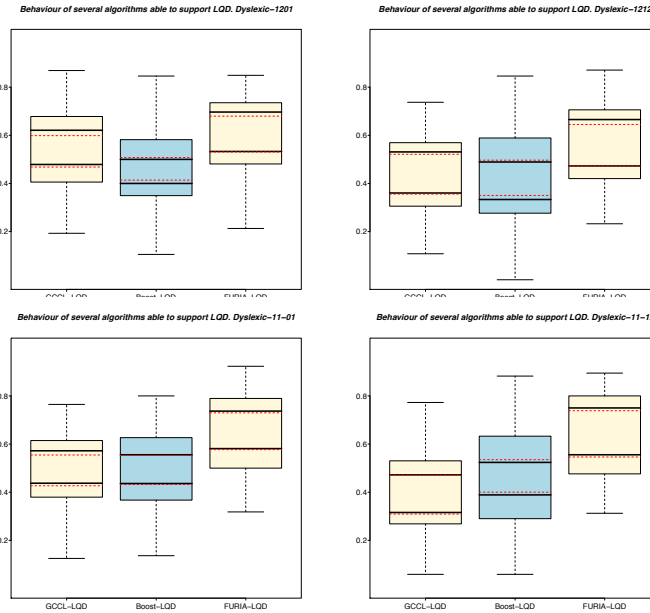


Figure 5: Boxplots showing the statistical differences between the algorithms.

of the capability of FURIA-LQD of learning from multi-labelled datasets.

This simpler approach consists in defining a new class of type “unknown”, that is added to the set of labels of all the instances of the training set where the classifier committed an error. For instance, if a child in the training set was labelled “dyslexia” and “other disorders” and the classification system outputs “no dyslexia” for this training instance, it is relabeled with the three terms “dyslexia”, “other disorders” and “unknown”. The learning is repeated over the extended training set, and this learning/relabeling process is iterated until none of the instances of the training set is incorrectly classified, i.e. the output of the classifier is a subset of the extended set of labels for all elements of the training data. With this simple change the accuracy of the “inexpert” classifier is improved until  $[0.765, 0.903]$  and the failure rate drops to  $[0.082, 0.214]$ , with a 48% of not classified instances (i.e. children that are assigned the class “unknown” in the test stage). In future works, a second battery of tests will be prepared for children classified as “unknown” thus a second classifier can be connected in cascade to the system proposed in this paper.

In Figure 8 some of the linguistic rules learned for the non-expert classifier are shown. Observe that the structure of the rules is simple and in many cases only one or two linguistic terms are needed. Each rule compares the vague result assigned by parents or instructors to the result of a test (“Actions and Details”, “Figures”, “Auditive memory”, etc.) and concludes that the learning problem is “Dyslexic”, “Control” or “Unknown”. Each knowledge base combines a small number of these rules (between 5

Table 8: Example of rules in the knowledge base

If Actions and Details is [3,15] and Figures is [1,17] and Auditive memory is [1,13] then DYSLEXIC
If Figures is [1,11] then CONTROL
If Figures is [17,28] then NON DYSLEXIC
If Vocabulary is 1 then DYSLEXIC
If Actions and Details is [3.7,5] then UNKNOWN
If Auditive Perception is [5.5,7.6] and Order Comprehension is ND or [6,8] and Spatial Orientation is ND or [6.9,8] then UNKNOWN

and 10), thus the resulting classifier can be easily understood by the psychologist, that can endorse it or, on the contrary, point out defects in its structure that could cause an incorrect assessment.

#### 4. Concluding remarks and future work

In this work a screening method for the early detection of dyslexia is proposed. The method is based on a classifier that inputs the subjective evaluation of different tests made by parents or teachers and produces a diagnostic or rejects the data. Success rates over 90% are possible with less than 50% rejection. In future works, specific test and classifiers will be designed that are connected in cascade with the proposed system and diagnose the rejected cases.

From a methodological point of view, the definition of the algorithm “FURIA” has been extended to LQD in this contribution. First results seem to show that this algorithm is preferred over boosting or GCCL when computing resources are limited. The linguistic quality of the outcome is also better. However, the accuracy of Boosting can still be higher for some datasets. In future works, further comparisons should be made that also involve the learning time. It is expected that FURIA improves over the alternatives in scenarios with a limited time for evolving a knowledge base, and the results obtained so far seem to confirm this. Lastly, the linguistic quality has been studied on the basis that a small number of rules is better, however the scattered fuzzy partitions produced by FURIA might not always be regarded as “human understandable” by most metrics of linguistic quality, that could be included in the analysis.

#### Acknowledgements

This work was supported by the Spanish Ministerio de Economía y Competitividad under Project TIN2011-24302, including funding from the European Regional Development Fund.

#### Appendix A. Theoretical properties of the proposed ranking operator

The partial ordering proposed in Section 2.2 generalises interval dominance criterion, thus it will be denoted it  $>_{ID}$ .

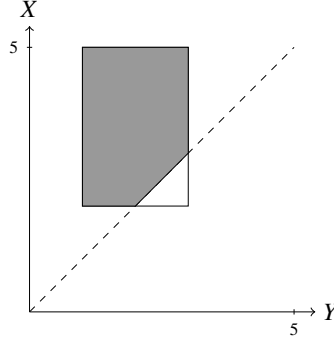


Figure A.6:  $S(X, Y)_\alpha$  illustration

**Proposition 1.** *The ordering  $>_{ID}$  is transitive, that is given three fuzzy sets  $\tilde{X}, \tilde{Y}, \tilde{Z}$ ,*

$$\tilde{X} >_{ID} \tilde{Y} \wedge \tilde{Y} >_{ID} \tilde{Z} \Rightarrow \tilde{X} >_{ID} \tilde{Z}$$

*Proof.* First note that, as  $\alpha$ -cuts are nested,  $\mathbf{1}_{(\underline{x}_{X,\alpha} > \bar{\pi}_{Y,\alpha})}$  is non-decreasing in  $\alpha$ . That is

- if  $\mathbf{1}_{(\underline{x}_{X,\alpha} > \bar{\pi}_{Y,\alpha})} = 1$ , then for any  $\beta > \alpha$ ,  $\mathbf{1}_{(\underline{x}_{X,\beta} > \bar{\pi}_{Y,\beta})} = 1$ ;
- if  $\mathbf{1}_{(\underline{x}_{X,\alpha} > \bar{\pi}_{Y,\alpha})} = 0$ , then for any  $\beta < \alpha$ ,  $\mathbf{1}_{(\underline{x}_{X,\beta} > \bar{\pi}_{Y,\beta})} = 0$ .

This means that  $\underline{P}(X > Y) > 0.5$  implies that there exists a value  $\alpha_1 < 0.5$  such that  $\mathbf{1}_{(\underline{x}_{X,\alpha_1} > \bar{\pi}_{Y,\alpha_1})} = 1$ . Similarly,  $\underline{P}(Y > Z) > 0.5$  implies that there exists a value  $\alpha_2 < 0.5$  such that  $\mathbf{1}_{(\underline{x}_{Y,\alpha_2} > \bar{\pi}_{Z,\alpha_2})} = 1$ . Let us now consider the value  $\alpha_3 = \max(\alpha_1, \alpha_2) < 0.5$ . For this value, we have  $\mathbf{1}_{(\underline{x}_{X,\alpha_3} > \bar{\pi}_{Z,\alpha_3})} = 1$ , since

$$\underline{x}_{X,\alpha_3} > \bar{\pi}_{Y,\alpha_3} > \underline{x}_{Y,\alpha_3} > \bar{\pi}_{Z,\alpha_3}$$

□

Also note that the values  $\underline{P}(X > Y)$  and  $\bar{P}(X > Y)$  coincide respectively with the Equations

$$NSD(X, Y) = 1 - \sup\{\alpha \mid \underline{x}_{X,\alpha} \leq \bar{\pi}_{Y,\alpha}\} = 1 - PD(Y, X) \quad (\text{A.1})$$

$$PD(X, Y) = \sup\{\alpha \mid \bar{\pi}_{X,\alpha} \geq \underline{x}_{Y,\alpha}\} \quad (\text{A.2})$$

proposed by Dubois and Prade [9] in another context, and where *NSD* (Necessary strict dominance) is the strongest requirement, while *PD* (possibilistic dominance) is the loosest.

In many applications, it is desirable to obtain not a partial ordering (as is  $>_{ID}$ ) but a complete ranking over the possible alternatives, forbidding incomparabilities (note

that this is the choice of most known proposals of fuzzy ranking). In practice, such a complete ranking should refine the partial ranking obtained by using the cautious approach obtained by first principles.

In practice, this means that we should derive a score  $S(X, Y) \in [\underline{P}(X > Y), \bar{P}(X > Y)]$  with an associated ordering  $>_S$  such that  $\tilde{X} >_S \tilde{Y}$  if  $S(X, Y) > 0.5$ , so that

- $X >_{ID} Y \Rightarrow X >_S Y$ ;
- $X \not>_{ID} Y \Rightarrow X \not>_S Y$ ;

The solution we propose is to compute an  $\alpha$ -cut wise statistical preference and then to integrate it over all  $\alpha$ -cuts. That is, we propose to associate to each  $\alpha$ -cut  $\pi_{X,\alpha}$  and  $\pi_{Y,\alpha}$  two uniform probabilities  $P_{X,\alpha}$  and  $P_{Y,\alpha}$ , and then to compute  $S(X, Y)_\alpha := P_\alpha(X > Y)$  by considering that  $P_{X,\alpha}$  and  $P_{Y,\alpha}$  are independent. Figure A.6 illustrates the value  $S(X, Y)_\alpha$  when  $\pi_{X,\alpha} = [2, 5]$  and  $\pi_{Y,\alpha} = [1, 3]$ . We can then compute

$$S(X, Y) = \int_0^1 S(X, Y)_\alpha d\alpha, \quad (\text{A.3})$$

and we can check that it is consistent with  $>_{ID}$ .

**Proposition 2.** *Given two fuzzy sets  $X, Y$ , then  $S(X, Y) \in [\underline{P}(X > Y), \bar{P}(X > Y)]$*

*Proof.* First, note that

- if  $\underline{\pi}_{X,\alpha} > \bar{\pi}_{Y,\alpha}$ , then  $S(X, Y)_\alpha = 1$  and
- if  $\bar{\pi}_{X,\alpha} > \underline{\pi}_{Y,\alpha}$ , then  $S(X, Y)_\alpha = 0$  and
- $S(X, Y)_\alpha \in (0, 1)$  otherwise

This means that, for an  $\alpha$ -cut,

$$S(X, Y)_\alpha \in \left[ \min_{\substack{x \in \pi_{X,\alpha} \\ y \in \pi_{Y,\alpha}}} f_{>}(x, y), \max_{\substack{x \in \pi_{X,\alpha} \\ y \in \pi_{Y,\alpha}}} f_{>}(x, y) \right]$$

Hence,  $\int_0^1 S(X, Y)_\alpha d\alpha \in [\underline{P}(X > Y), \bar{P}(X > Y)]$  □

To our knowledge, the score  $S(X, Y)$  given by Eq. A.3 has never been proposed as a way to rank fuzzy sets, however we have shown here that it can be considered as a specific extension of classical statistical ranking.

## Appendix B. Pseudocode of GCCL and Boosting algorithms

The pseudocode of the GCCL [22] and Boosting algorithms [24] for LQD is reproduced in this appendix for the convenience of the reader.

### 1. GCCL: Outline

```

function GFS
1   Initialize population
2   for iter in {1,...,Iterations}
3     for sub in {1,...,subPop}
4       Select parents
5       Crossover and mutation
6       assignConsequent(offspring)
7     end for sub
8     Replace the worst subPop individuals
9     assignFitness(population,dataset)
10    end for iter
11  Purge unused rules
return population

```

## 2. GCCL: Assignment of consequents

```

function assignConsequent(rule)
1   for c in {1,...,Nc}
2     grade = 0
3     compExample = 0
4     for example in {1,...,N}
5        $\tilde{m} = \text{fuzMembership}(\text{Antecedent}, \text{example}, c)$ 
6       grade = grade  $\oplus$   $\tilde{m}$ 
7       if ( $\sup \{x : \tilde{m}(x) > 0\} > 0$ ) then
8         compExample = compExample + 1
9       end if
10    end for example
11    weight[c] = grade  $\otimes$  compExample
12  end for c
13  mostFrequent = {1,...,Nc}
14  for c in {1,...,Nc}
15    for c1 in {c+1,...,Nc}
16      if (weight[c] dominates weight[c1]) then
17        mostFrequent = mostFrequent - {c1}
18      end if
19    end for c1
20  end for c
21  Consequent = select(mostFrequent)
22  CF[rule] = computeConfidenceOfConsequent
return rule

```

## 3. GCCL: Assignment of fitness to the rules in the KB

```

function assignFitness(population,dataset)
1   for example in {1,...,N}
2     setWinnerRule =  $\emptyset$ 
3     for r in {1,...,M}
4       dominated = FALSE
5        $r.\tilde{m} = \text{fuzMembership}(\text{Antecedent}[r], \text{example}) * \text{CF}[r]$ 

```

```

6   for sRule in setWinnerRule
7     if (sRule dominates r) then
8       dominated = TRUE
9     end if
10    end for sRule
11    if (not dominated and r.m̃ > 0) then
12      for sRule in setWinnerRule
13        if (r.m̃ dominates sRule) then
14          setWinnerRule = setWinnerRule - { sRule }
15        end if
16      end for sRule
17      setWinnerRule = setWinnerRule ∪ { r }
18    end if
19  end for r
20  if (setWinnerRule == ∅) then
21    setWinnerRule = setWinnerRule ∪ { rule_freq_class }
22  setOfCons = ∅
23  for sRule in setWinnerRule
24    setOfCons = setOfCons ∪ { consequent(sRule) }
25  end for sRule
26  deltaFit = 0
27  if ({class(example)} == setOfCons and
28     size(setOfCons) == 1) then
29    deltaFit = {1}
30  else
31    if ({class(example)} ∩ setOfCons ≠ ∅) then
32      deltaFit = {0, 1}
33    end if
34  end if
35  Select winnerRule ∈ setWinnerRule
36  fitness[winnerRule] = fitness[winnerRule] ⊕ deltaFit
37 end for example
38 return fitness

```

#### 4. Boosting of fuzzy rules for LQD and multi-class problems.

##### Input data:

- 1 Low quality training set  $(\tilde{\mathcal{X}}_1, \tilde{y}_1), \dots, (\tilde{\mathcal{X}}_m, \tilde{y}_m)$ ,  
 $\tilde{\mathcal{X}}_i \in \mathcal{F}(R^n)$  and  $\tilde{y}_i \in \mathcal{F}(\{1, \dots, C\})$
- 2 Number of fuzzy rules  $H \subset \{1, \dots, N\}$
- 3 Number of partitions  $E$

##### Local Variables:

- 1  $\tilde{w} \in \mathcal{F}(R^m)$ : weights of the instances
- 2  $\alpha \in R^{H \times C}$ : votes of the weak hypotheses
- 3  $y \subseteq \{-1, 1\}^m$ : consequents of the weak hypotheses
- 4  $s \in [0, 1]^{H \times C}$ : confidence of the consequents of rules
- 5  $\mathcal{A} \in \mathcal{F}([0, 1]^{m \times n \times E})$ : fuzzy membership

```

begin
1  for k=1 until C
2    for i=1 until m
3      Initialize  $\tilde{w}_i$ :  $\tilde{w}_i = 1/m$ 
4    end-for
5  repeat H/C times
6    Evolutionary Generation of a new fuzzy rule(k):
7      “ $R_j$ : if x is  $A^j$  then  $c_k$ ”  $\rightarrow$   $\widehat{\text{fitness}}_{R_j}$ 
8      Calculate the number of votes of the rule:
9      “ $R_j$ : if x is  $A^j$  then  $c_k = \alpha_k^j$ ”
10     Update the weights of the instances:  $\tilde{w}_i$ 
11   end-repeat
12 end-for
13 Convert votes into confidences:  $s_k^j$ 
14 Emit rules: for all j, if any  $s_k^j \neq 0$ 
15  $\overline{\text{Error}}_{\text{train}}$ : Evaluate the training set  $\leftarrow$  Inference from LQD
16  $\overline{\text{Error}}_{\text{test}}$ : Evaluate the test files
end

```

5. Boosting: Evolutionary Generation of a new fuzzy rule:

```

Evolutionary Generation of a new fuzzy rule(k):
1  begin
2    Initialize population
3     $\widehat{\text{fitness}} \leftarrow$  ImpreciseFitness(population,k)
4  for iter=1 until Iterations
5    Select parents  $\leftarrow$  Tournament (precedence operators)
6    Crossover and mutation
7    Replace the worst individuals  $\leftarrow$  Precedence operators
8     $\widehat{\text{fitness}} \leftarrow$  ImpreciseFitness(population,k)
9  end-for
10  $R_j$  minimizes  $\widehat{\text{fitness}} \leftarrow$  Precedence operators
11 return  $R_j$ 
12 end

```

6. Boosting: Fitness function for the evolutionary generation of a new fuzzy rule.

**ImpreciseFitness (population,k):**

```

1 begin
2   for j=1 until population
3      $\widetilde{fit}_{R_j} = 0$ 
4     for i=1 to m,  $\alpha \in [0, 1]$ 
5       if  $([\bar{y}_i]_\alpha = k) y_i = 1$ 
6       if  $([\bar{y}_i]_\alpha \cap k = \emptyset) y_i = -1$ 
7       if  $([\bar{y}_i]_\alpha \neq k \text{ and } [\bar{y}_i]_\alpha \cap k \neq \emptyset) y_i = \{1, -1\}$ 
8        $[\widetilde{fit}_{R_j}]_\alpha = [\widetilde{fit}_{R_j}]_\alpha \oplus [\bar{w}_j]_\alpha$ 
           $\otimes \{\exp(\{-(y_i) \otimes A^j(x)\} | x \in [X_i]_\alpha)\}$ 
9     end-for
10  end-for
11 end

```

**Bibliography**

- [1] L. Bender, Test Guestáltico Visomotor, Buenos Aires: Paidós, 1982 (in Spanish).
- [2] R. Berlin, Eine besondere Art von Wortblindhert (Dyslexie). Monograh, Verlag von J. F. Bergmann, Wiesbaden, 1887 (in German).
- [3] A.E. Boehm, BOEHM: Test de Boehm de conceptos basicos, Madrid:TEA, 1967 (in Spanish).
- [4] E. Brouwers, A. Peterson, J. L. Palacios, L. Centolanza, Ice Adhesion Strength Measurements for Rotor Blade Edge Materials, 67th Annual Forum Proceedings, American Helicopter Society, Virginia Beach, VA, 2011.
- [5] W. Cohen, Fast effective rule induction, In Prieditis A., Russel S. (eds), Proceeding of the 12th International Conference on Machine Learning (ICML), 115-123 (1995).
- [6] I. Couso, L. Sánchez, Higher order models for fuzzy random variables, Fuzzy Sets and Systems 159, 237-258, 2008.
- [7] M. V. De la Cruz, Bateria de aptitudes para el aprendizaje escolar, Publicaciones de psicología aplicada, Madrid: TEA, 1989 (in Spanish).
- [8] F. A. T. De Carvalho, R. M. Souza, M. Chavent, Y. Lechevallier, Adaptive Hausdorff distances and dynamic clustering of symbolic interval data, Pattern Recognition 27, 167-179, 2006.
- [9] D. Dubois, H. Prade. Ranking fuzzy numbers in the setting of possibility theory. Information Sciences, 30(3): 183-224, 1983.
- [10] L. Filho, Test A.B.C., Barcelona: Paidós, 1937 (in Spanish).
- [11] M. Frostig, Test de desarrollo de la percepción visual, Madrid: TEA, 1978 (in Spanish).
- [12] F. L. Goodenought, Test de la figura humana, Barcelona: Paidós, 1991 (in Spanish).
- [13] A.J. Harris, Test de dominancia lateral, Madrid: TEA, 1978 (in Spanish).



- [14] L. Hedjazi, J. Aguilar-Martin, M. V. Le Lann, Similarity-margin based feature selection for symbolic interval data, *Pattern Recognition Letters* 32, 578-585, 2011.
- [15] F. Herrera et al., *Statistical Inference in Computational Intelligence and Data Mining*, Thematic Public Websites of the Research Group SCI2S. Granada University, Spain. URL: <http://sci2s.ugr.es/sicidm/>
- [16] J.C. Hühn and E. Hüllermeier, FURIA: an algorithm for unordered fuzzy rule induction, *Data Mining and Knowledge Discovery* 19, 293-319, 2009.
- [17] J. C. Hühn, E. Hüllermeier, FURIA: Fuzzy Unordered Rule Induction Algorithm, URL:<http://www.uni-marburg.de/fb12/kebi/research/software/furia> (2009).
- [18] J.C. Hühn and E. Hüllermeier, An analysis of the FURIA algorithm for fuzzy rule induction, In *Advances in Machine Learning I*, 32-344, 2010.
- [19] J. J. Jansky. Early Prediction of Reading Problems. *Bulletin of the Orton Society*, 23. pp 78-89. 1973.
- [20] Orton, S. *Reading, Writing and Speech problems in children*. New York: W. W. Norton. 1937.
- [21] J. Otero, L. Sánchez, I. Couso, A. Palacios, Bootstrap analysis of multiple repetitions of experiments using an interval value multiple comparison procedure, *Journal of Computer and System Sciences*, doi:10.1016/j.jcss.2013.03.009, Accepted.
- [22] A. Palacios, L. Sánchez, and I. Couso, Diagnosis of dyslexia with low quality data with genetic fuzzy systems, *International Journal on Approximate Reasoning* 51, 993-1009, 2010.
- [23] A. Palacios, L. Sánchez, I. Couso, Future performance modelling in athleticism with low quality data-based GFSs, *Journal of Multivalued Logic and Soft Computing* 17(2-3), 207-228, 2011.
- [24] A. Palacios, L. Sánchez, and I. Couso, Boosting of fuzzy rules with low quality data, *Journal of Multiple-Valued Logic and Soft Computing*, 19(5-6), 591-619, 2012.
- [25] A. Palacios, L. Sánchez, and I. Couso, An extension of the FURIA classification algorithm to low quality data, *HAIS 2013*.
- [26] A. Palacios, L. Sánchez, I. Couso. Supervising classrooms comprising children with dyslexia and other learning problems with graphical exploratory analysis for fuzzy data: Presentation of the software tool and case study. *FUZZ-IEEE 2014*, 2133-2140, 2014.
- [27] A. Palacios, L. Sánchez. Una herramienta software para la supervisión de grupos de niños con dislexia y otros problemas de aprendizaje. *Jornadas Universidad y Discapacidad*. Oviedo. 2014.
- [28] J. Quevedo *et al.*, Validation and reconstruction of flow meter data in the Barcelona water distribution network, *J. Control Eng. Practice* 18, 640-651, 2010.
- [29] J.C. Raven, *El test de matrices progresivas*, Barcelona: Paidós, 1962 (in Spanish).

- [30] R.M. Rivas and P. Fernández, *Dislexia, disortografía y disgrafía*, Pirámide, 1994 (in Spanish).
- [31] M. Stamback, *Tono y psicomotricidad*, Madrid:Pablo del Río, 1978 (in Spanish).
- [32] E. Toulouse and H. Pieron. *Prueba perceptiva y de atención*, 1904 (in Spanish).
- [33] J. Toro and M. Cervera, *TALE Test de análisis de la lectoescritura*, Madrid: Pablo del Río, 1980 (in Spanish).
- [34] R. Wagner. Rudolf Berlin: Originator of the Term Dyslexia. *Bulletin of the Orton Society*, 23. pp 57-64. 1973.
- [35] D. Wechsler, *WISC: Escala de inteligencia de Wechsler para niños revisada*, Madrid: TEA, 1993 (in Spanish).
- [36] C. Yuste, *BADYG-E: Bateria de aptitud diferencial y general*, Madrid:CEPE, 1985 (in Spanish).