Multi-objective based FRBSs for trade-off improvement in Accuracy and Interpretability: a rule relevance point of view.[☆]

M. I. Rey[a], M. Galende[b], M. J. Fuente[c], G. I. Sainz-Palmero[c,*]

[a] *INDOMAUT S.L., Pol. Ind. San Cristóbal, 47012 Valladolid, SPAIN.*
[b] *CARTIF Centro Tecnológico. 47151 Boecillo (Valladolid), SPAIN.*
[c] *Dpt. of Systems Engineering and Control. School of Industrial Engineering. University of Valladolid. 47011 Valladolid, SPAIN.*

Abstract

Fuzzy rule-based systems (FRBSs) are a common alternative for applying fuzzy logic in different areas and real-world problems. The schemes and algorithms used to generate these types of systems imply that their performance can be analyzed from different points of view, not only model accuracy. Any model, including fuzzy models, needs to be sufficiently accurate, but other perspectives, such as interpretability, are also possible for the FRBSs. Thus, the Accuracy-Interpretability trade-off arises as a challenge for fuzzy systems, as approaches are currently able to generate FRBSs with different trade-offs.

Here, rule Relevance is added to Accuracy and Interpretability for a better trade-off in FRBSs. These three factors are involved in this approach to make a rule selection using a multi-objective evolutionary algorithm.

The proposal has been tested and compared with nine datasets, two linguistic and two scatter fuzzy algorithms, four measures of interpretability and two rule relevance formulations. The results have been analyzed for different views of Interpretability, Accuracy and Relevance, and the statistical tests have shown that significant improvements have been achieved. On the other hand, the Relevance-based role of fuzzy rules has been checked, and it has been shown that low Relevance rules have a relevant role for trade-off,

[☆] This work has been partially supported by the Spanish Ministry of Economy and Competitiveness through the Project no. DPI2015-67341-C2-2-R

[*] Corresponding author:
Email address: gresai@eii.uva.es (G. I. Sainz-Palmero)

while some rules with high Relevance must sometimes be removed to reach an adequate trade-off.

Keywords:
Fuzzy Rule-Based Systems, Interpretability, Rule Relevance, Orthogonal Transformations

1. Introduction

Fuzzy rule-based systems (FRBSs) are a common approach for applying fuzzy logic in many areas of activity, both in research areas and for solving real-world problems. In fact, the FRBS based approaches are used by scientists and practitioners for modeling, control, decision making, etc (Kacprzyk and Pedrycz, 2015; Konar, 2005; Karray and De Silva, 2004). In order to generate these FRBSs, many different algorithms and approaches are available in the scientific literature (Alcalá-Fdez and Alonso, 2016; Magdalena, 2015), each considering their performance from different points of view. On this point, linguistic and precise approaches appear as two major ways to address FRBS generation (Fernández et al., 2015; Casillas et al., 2003a,b).

FRBS performance has been an open and debatable issue for a long time in research domains. Accuracy is a basic goal for any model, including FRBS models. However, other views of the performance are possible, especially when fuzzy logic is involved; views such as Interpretability, which is intrinsically connected with some of the foundations of fuzzy logic, for instance, the capability to express, represent and understand knowledge in linguistic terms as humans do.

This type of performance can be essential, or even compulsory, in some application areas, as well as in theoretical developments, in which the behavior of the FRBS, when modeling a reality, must be clearly defined and understood. So, to obtain FRBSs with an adequate Interpretability is an interesting goal to reach. Yet, on the other hand, FRBSs must be accurate enough because an insufficiently accurate model is useless.

Keeping all this in mind, the challenge to obtain a performance based on several views, such as Accuracy and Interpretability, arises. The goal is to reach an Accuracy and Interpretability trade-off or balance between these views which are considered mutually contradictory, since an improvement on one side could imply a worsening of the performance from the other point of view (Alonso et al., 2015; Casillas et al., 2003a).

This Accuracy-Interpretability trade-off is an open issue and many approaches are available in the literature that focus on different views, such as genetic fuzzy systems or rule selection (Fernández et al., 2015; Fazzolari et al., 2013a; Cordón, 2011), considering several metrics regarding complexity or semantics, (Gacto et al., 2011; Alonso et al., 2009; Mencar and Fanelli, 2008; Zhou and Gan, 2008), etc.

In this context, our proposal deals with the FRBS Accuracy-Interpretability trade-off addressed by a multi-objective evolutionary-based rule selection, choosing the most adequate and significant rules according to the metrics of Accuracy, Interpretability and Relevance. In most cases, Relevance has been used to reduce complexity in the FRBS (Zhou et al., 2009; Setnes, 2003), but here, rule Relevance is as important a factor as accuracy or interpretability for the trade-off. In Rey et al. (2012), early ideas connected with this proposal were introduced.

The rest of the paper is organized as follows: Section 2 introduces the idea of Relevance in different fields. Section 3 describes the main works connected with the Accuracy-Interpretability trade-off challenge, considering different points of view and approaches, and also describes rule relevance as a metric to be considered in this analysis. Section 4 introduces the theoretical principles to be taken into account in this work, that is, FRBS modeling algorithms in Section 4.1, orthogonal transformations in Section 4.2, and MOEAs in Section 4.3. The proposal of this work is explained in Section 5, and the experimentation is described in Section 6, including the methodology in Section 6.1 and the most valuable results in Section 6.2. Finally, in Sections 7 and 8, the main conclusions of this proposal are set out.

2. Relevance

Relevance is an idea managed and understood as "something that is or is considered as worthwhile". In our daily activities, this concept is present from music or book recommendations to any type of decision-making. A relevant idea, concept, issue, etc. acquires its meaning when it is compared to others. The idea of Relevance has been used in different areas of human knowledge, from philosophy (Keynes, 2013), psychology (Horn and Ward, 2008), pattern recognition (Devroye et al., 2013), or feature selection and extraction (Liu and Motoda, 2012), to machine learning (Yu and Liu, 2004). The challenge is how to evaluate this relevance.

In information sciences, several ideas of Relevance are managed in order to make a selection from competing sources, geared toward maximizing results and/or minimizing effort in dealing with results. Here, *System or algorithmic relevance* is managed as the relation between a query and the information objects in a system, as retrieved or as failed to be retrieved by a given procedure or algorithm (Hjørland, 2010). In other linked fields such as Information Retrieval (IR), the concept of Relevance is separated into two major classes : (1) objective or system-based relevance; and (2) subjective or human (user)-based relevance, corresponding to the system-driven and the cognitive user-oriented approaches.

This is connected with *Relevance Feedback*, a human computer interaction technique to capture and re-use the knowledge of a user. It has been extensively used in text-based document retrieval systems (Okabe and Yamada, 2005), interactive content-based image retrieval systems (Kundu et al., 2015), or fingerprint identification systems (Kwan et al., 2015).

Focusing on the rules of an FRBS, there are some metrics about Relevance: the *Probabilistic Test*, based on the available learning data, in which a fuzzy IF-THEN statement represents a locally relevant aspect of the dependency of the input and output variables; a Relevance factor, based on the computation of confidence intervals, is computed for each fuzzy rule in the FRBSs in order to subsequently select the most relevant ones (Krone and Taeger, 2001). *Ratios between membership*, resulting in three definitions of Relevance: Relevance of a rule on a region, Relevance of a rule of the fuzzy system, and Relevance of the fuzzy system (Salgado, 2008). The *Relevance Vector Learning Mechanism*, where relevant fuzzy rules are obtained using the Relevance Vector Machine (RVM), acquires relevance vectors and weights by maximizing a marginal likelihood (Kim et al., 2006).

On the other hand, we have Relevance as a factor to improve the Interpretability of the rules, involved in Pedrycz (2003), quantified in terms of the data covered by the antecedent and conclusion, or Relevance as an automatic method of rule reduction, where the relevance of a rule or rule base involves accuracy, statistical significance, and clearness (Mikut et al., 2005). In this field, the Relevance of rules based on the *Orthogonal Transformation based Ranking* to make a rule selection is carried out by Setnes (2003), and based on the variability of each rule, using *Forward Stepwise* and *Backward Elimination* procedures in Zhou et al. (2009). This view is considered in this work and is used with the Accuracy-Interpretability trade-off in the following section.

3. Accuracy-Interpretability Trade-Off

FRBSs are a very popular approach for modeling in different areas as solutions to a wide range of problems (Magdalena, 2015; Konar, 2005; Karray and De Silva, 2004). A major dilemma for these FRBS based models is the evaluation of their performance. On the one hand, Accuracy is essential for any type of model, fuzzy or not fuzzy, that is targeted by Precise Fuzzy Modeling. On the other hand, fuzzy logic permits the evaluation of these FRBSs from other points of view, such as the Interpretability that is targeted by Linguistic Fuzzy Modeling. This double view has usually been considered as contradictory, bringing about a major challenge for FRBSs: the Accuracy-Interpretability trade-off, which means to find a balance or compromise between the necessary accuracy and the desired interpretability. Nowadays, the way to reach this trade-off is an area of discussion and debate concerning the different approaches (Ishibuchi and Nojima, 2015; Casillas et al., 2003a,b).

A review of the literature focused on this challenge shows that Accuracy is a well-defined performance based on the model *error*. However, the definition of Interpretability, and thus the way to measure this concept going from complexity to semantic issues, has been an open issue for a long time (Alonso et al., 2015; Gacto et al., 2011; Alonso et al., 2009; Zhou and Gan, 2008; Mencar and Fanelli, 2008). Several classifications and taxonomies about Interpretability, its definitions, views, measurements and methodologies to reach the trade-off, are available in such specialized literature as Zhou and Gan (2008), in which two levels of interpretability are established according to the FRBS components: Low level for fuzzy set issues and High level for fuzzy rules. In Alonso et al. (2009), the previous taxonomy about interpretability is generalized. Another point of view was introduced in Mencar and Fanelli (2008), where the constraints to be applied to obtain interpretable FRBSs are defined. Finally, in Gacto et al. (2011), a taxonomy based on complexity and semantic issues for the rule and fuzzy set is introduced, reviewing most measures based on complexity and semantic issues in this scheme.

These taxonomies, and other works, show there is a wide set of indexes or measurements about interpretability, each representing its own view regarding Interpretability. First of all there is the *number of rules*, perhaps the most popular measurement (Fazzolari et al., 2013b; Márquez et al., 2012; Alonso and Magdalena, 2011; Gacto et al., 2010; Casillas et al., 2009; Mikut et al., 2005; Ishibuchi et al., 1997). Then we have the *number of conditions in* *the antecedent* (Nguyena et al., 2015; Antonelli et al., 2011; Pulkkinen et al., 2008; Ishibuchi et al., 2001), the *number of variables* (Alonso and Magdalena, 2011; Mikut et al., 2005) or the *number of membership functions* (Guillaume and Charnomordic, 2003; Roubos and Setnes, 2001), where the idea of complexity is managed. Secondly, we have those involving semantic issues such as: the *distinguishability* of the fuzzy sets (Oliveira, 1999), *consistence* and *similarity* of the fuzzy rule base (Alonso and Magdalena, 2011; Pulkkinen et al., 2008), the *number of rules simultaneously fired* (Pancho et al., 2013; Márquez et al., 2012), etc.; or those defining more complex indexes to obtain some semantic restrictions, such as *GM3M* (Gacto et al., 2010), *RBC* (Alonso and Magdalena, 2011), *Integrity I* (Antonelli et al., 2011), *Transparency* (Pulkkinen et al., 2008), *Cointension* (Mencar et al., 2011) or *RMI* (Galende et al., 2014).

All these accuracy and interpretability indexes permit the FRBS Accuracy and Interpretability to be evaluated, but how can the trade-off be reached? This question, again, has had different approaches: rule generation, rule selection and tuning of rule fuzzy sets from existing FRBSs have been some options. The most popular way to carry out these tasks has been addressed as a genetic based optimization problem, but other options have also been used.

A general compilation of methods based on genetic algorithms can be found in Fernández et al. (2015). MOEA based rule generation from data is in Nguyena et al. (2015); Antonelli et al. (2011); Casillas et al. (2009) and Ishibuchi and Nojima (2007). Decision trees and pruning strategies are used in Mikut et al. (2005) for rule generation. Meanwhile, this goal is guided by user preferences in Guillaume and Charnomordic (2003), decision trees and MOEA in Pulkkinen et al. (2008), and true tables are involved in Mencar et al. (2011).

Another view is based on an existing, improved FRBS: approaches based on rule selection and MOEA are developed in Galende et al. (2012); Márquez et al. (2012); Pulkkinen et al. (2008); Ishibuchi and Nojima (2007); Ishibuchi et al. (2001) and Ishibuchi et al. (1997). A GA tuning based approach can be found in Roubos and Setnes (2001), while MOEA based rule selection and tuning is used by Galende et al. (2014); Fazzolari et al. (2013b); Alcalá et al. (2011) and Gacto et al. (2010). This view is used in this approach.

Another view is found in Pancho et al. (2013), where the FRBS is analyzed as a social network based on *goodness* and *relative coverage of the rules*, where these rules and their links are pruned and graphically represented.

A further approach to the Accuracy-Interpretability trade-off, involving reduction of complexity by rule selection, is based on the Relevance of FRBS and the ranking of its rules. The concept of Relevance is very popular with other domains, such as Feature Selection and Extraction (Liu and Motoda, 2012), and it is usually based on orthogonal transformations. In the FRBS domain, approaches using orthogonal transformations are based on obtained values from fired rule matrices, such as: *eigen values*, *R-values*, *variances*, *c-values*, α*-values*, w_1*-values*, w_2*-values* and *L-values* for each rule, which allows a ranking of rules to be generated. So here, the challenge is to define which rules should be preserved and which should be removed, based on this ranking.

This idea of Relevance, as commented in the previous section, has been used in different ways, such as in Zhou et al. (2009); Destercke et al. (2007); Setnes (2003) and Yen and Wang (1999) to reduce the complexity of the FRBSs by rule selection, thus reaching a better interpretability. The target is to preserve the most relevant rules, selecting *relevant rules* or removing *non relevant rules*. Specifically, a relevance ranking based sequential (*forward stepwise* and *backward elimination*) selection of rules is used in Zhou et al. (2009) to make groups of rules. All of the relevant rules are taken into account and some of the lower relevant rules may be considered. Meanwhile, in Setnes (2003), a "gap" is searched through the relevance value of the rules to select the rules to be preserved; while in Zhou and Gan (2007), a threshold for relevance is defined and then the selection is carried out. On the other hand, in Alonso and Magdalena (2011), a rule ranking based on an interpretability index is used; while in Rey et al. (2012), a MOEA approach based on accuracy and interpretability, which includes Relevance, is carried out. The latter was an early approach of this current work.

The role of (fuzzy) rules with low relevance has also been discussed in the specialized literature (Zhou et al., 2009; Setnes, 2003). In general, in most cases, this type of rule has been considered as the right candidate to be removed when a reduction of the complexity of FRBS (and other views of interpretability) was involved. In this way, rules with low relevance do not seem very interesting under the almost ordinary view of rule relevance and Accuracy-Interpretability trade-off. On the other hand, rules with high relevance seem to be the right rules to be preserved through any selection of rules. In this way, the idea of looking for a *Gap* to make rule selection has been one of the most popular approaches.

In this context, this proposal to address the Accuracy-Interpretability trade-off challenge is based on the use of Relevance, Accuracy and Interpretability to carry out a rule selection in a multi-objective evolutionary way. Rule relevance is evaluated by orthogonal transformations as in Zhou et al. (2009), Destercke et al. (2007), etc., which permits a better evaluation of relevance because it does not depend on parameters/threshold values or iterative estimations. Here, rule relevance allows the most meaningful rules for each FRBS based model to be considered, so that the model contains the most adequate variability regarding each input/output space in comparison with other approaches that only consider accuracy and interpretability for the trade-off. The rule selection based on these three objectives and an evolutionary strategy permits a more efficient selection of rules, avoiding the difficulties of finding a "gap" between relevant and non-relevant rules, and the failure to consider the latter, in Setnes (2003), or regarding the inefficiencies of the sequential selection of the rules with high and low relevance, such as in Zhou et al. (2009).

On the other hand, the proposal allows the effect of Relevance on the Accuracy-Interpretability trade-off and the role of rules to be checked according to their own Relevance.

4. Theoretical Issues

4.1. Fuzzy Rule-Based Systems

A fuzzy rule-based system (FRBS) can be seen as a knowledge base that includes a rule base with information described by IF-THEN fuzzy rules and a data base with the correspondence of the fuzzy values, an inference engine containing a fuzzification interface, an inference system, and a defuzzification interface (Fernández et al., 2015). Based on these components, FRBS can be classified under different views, perhaps the most usual way being based on the rule type of their knowledge base, then the FRBS can be classified as (Herrera, 2008):

Scatter FRBS, rule antecedents and consequents are defined by fuzzy sets, each with their own semantics.

Linguistic FRBS, usually contains ($Mamdani$) rules: rule antecedents and consequents are defined by fuzzy sets associated with linguistic terms. These fuzzy sets share the semantics for all rules.

TSK FRBS (*Takagi-Sugeno-Kang*), rule antecedents are defined using fuzzy sets, and rule consequents are functions of the rule antecedents.

These different types of FRBS rules can be generated using different fuzzy modeling algorithms. Fuzzy neural networks and fuzzy genetic systems are the major approaches to generate FRBSs, so in this work, two fuzzy neural networks are used: FasArt and NefProx, and two fuzzy genetic systems: S-IRL and L-IRL. Two of them are scatter-based algorithms: FasArt and S-IRL, and two linguistic-based algorithms: NefProx and L-IRL.

FasArt (*Fuzzy Adaptive System ART based*) is a Neuro-Fuzzy system based on the Adaptative Resonance Theory (ART) that it is able to generate scatter systems. It is characterized by *single point* fuzzification, *product* inference, and *average of fuzzy set center* defuzzification (Cano Izquierdo et al., 2001).

S-IRL (*Scatter Iterative Rule Learning*), guided by a genetic algorithm, is able to generate scatter systems. It is characterized by *center of gravity weighted by the matching* defuzzification and *minimum t-norm* as implication and conjunctive operators (Cordón and Herrera, 2001) [1].

NefProx (*Neuro-Fuzzy Function Approximation*) is a Neuro-Fuzzy algorithm based on supervised learning able to generate linguistic systems. It is characterized by *max-min* inference and *mean of maximum* defuzzification (Nauck and Kruse, 1999) [2].

L-IRL (*Linguistic Iterative Rule Learning*) that, following a similar strategy to S-IRL, is able to generate linguistic systems (Cordón and Herrera, 1997) [3].

All these algorithms generate FRBSs with different Accuracy-Interpretability trade-offs, according to their nature.

[1] S-IRL available in the KEEL software tool as MOGUL-IRLHC-R
[2] NefProx available at `http://fuzzy.cs.uni-magdeburg.de/nefprox/`
[3] L-IRL available in the KEEL software tool as MOGUL-IRLSC-R

4.2. Orthogonal Transformations

Orthogonal transformations (Golub and Van Loan, 2012) are one of the most useful and powerful tools of numerical linear algebra, and they are used in many areas such as control, signal processing, feature selection, etc. The target of these transformations is to discover the intrinsic dimensionality of data, and this is done through matrix decompositions involving orthogonal matrices. Thus, given a matrix M, it can be decomposed using different orthogonal transformations as follows:

SVD (*Singular Value Decomposition*) $M = U\Sigma V^T$, where U and V are orthogonal matrices and Σ is a non-negative diagonal matrix with the *singular values* $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_M \geq 0$ in decreasing order as diagonal.

P-QR (*Pivored QR Decomposition*) $M\Pi = QR$, where Π is a permutation matrix, Q is an orthogonal matrix and R is an upper triangular matrix whose diagonal values are called *R-values*.

OLS (*Orthogonal Least Square*) $M = WA$, where W is an orthogonal matrix and A is an upper-triangular matrix with unity diagonal values.

Then, in this case, an FRBS can be formulated as a linear regression problem according to (Setnes, 2003; Yen and Wang, 1999):

$$y = P\theta + e \tag{1}$$

where $y = [y_1, y_2, \cdots, y_N]^T$ are the system outputs, $\theta = [c_1, c_2, \cdots, c_M]^T$ are the consequents of the M rules, $e = [e_1, e_2, \cdots, e_N]^T$ are the approximation errors and $P = [p_1, p_2, \cdots, p_M] \in R^{N\times M}$ contains the firing strength of all M rules for the N inputs, with $p_i = [p_{i1}, p_{i2}, \cdots, p_{iN}]^T$. Also, $x = [x_1, \cdots, x_N]^T$ are the inputs and $\{A_{i1}, \cdots, A_{iN}\}$ the fuzzy sets defined for the antecedent, so p_i can be expressed as:

$$p_i(x) = \frac{\prod_{j=1}^{N} A_{ij}(x_j)}{\sum_{k=1}^{M}\prod_{j=1}^{N} A_{kj}(x_j)} \tag{2}$$

These orthogonal decompositions applied on this firing strength matrix P permit the relevance of the rules of an FRBS to be evaluated (Zhou et al., 2009; Destercke et al., 2007; Setnes, 2003; Yen and Wang, 1999):

- Using SVD, the most relevant rules are those associated with higher *singular values*.

- Applying P-QR, the most active and least redundant rules are those whose *R-values* are higher.
- OLS, higher values of the explained output variance $[xVar]_i = \frac{g_i^2 w_i^T w_i}{y^t y}$ ($g = A\theta$) are assigned to relevant rules.

The OLS approach takes into account not only the rule antecedents but also the rule consequents in comparison with the SVD and P-QR approaches.

4.3. Multi-Objective Evolutionary Algorithms

Multi-Objective Evolutionary Algorithms (MOEAs) are computational algorithms inspired by genetic foundations to find solution sets to problems subject to several objectives, to be simultaneously optimized. Their strategy is to evolve from candidate solutions towards a best non-dominated solutions set of the *Pareto Front*. In this proposal, a well-known MOEA has been taken into account: SPEA2 (*Strength Pareto Evolutionary Algorithm 2*) (Zitzler et al., 2001). This MOEA ensures an effective balance between exploitation and exploration in the search space. In this algorithm, each individual is evaluated taking into account the number of individuals it dominates and the number of individuals by which it is dominated, a nearest neighbor density estimation technique is used and an enhanced truncation method is applied.

5. Accuracy-Interpretability-Relevance Trade-Off: A Proposal

The target of this proposal is to improve the well-known Accuracy-Interpretability trade-off for FRBSs, while preserving the most relevant rules for each FRBS under the trade-off view. This goal is carried out by a multi-objective optimization-based rule selection involving: Accuracy, Interpretability and Relevance concepts. This proposal is not dependent on the type of FRBS (scatter or linguistic) and is carried out in two stages (Fig. 1):

1. Generation of an FRBS based on data and/or expert knowledge.
2. Improvement of the Accuracy-Interpretability trade-off of this FRBS through a selection of its rules based on MOEA, involving rule accuracy, interpretability and relevance issues.

The involvement of such issues as accuracy, interpretability and relevance implies their definition and formulation in order to be evaluated as objectives through a MOEA based rule selection. The rule selection of this proposal has three goals:

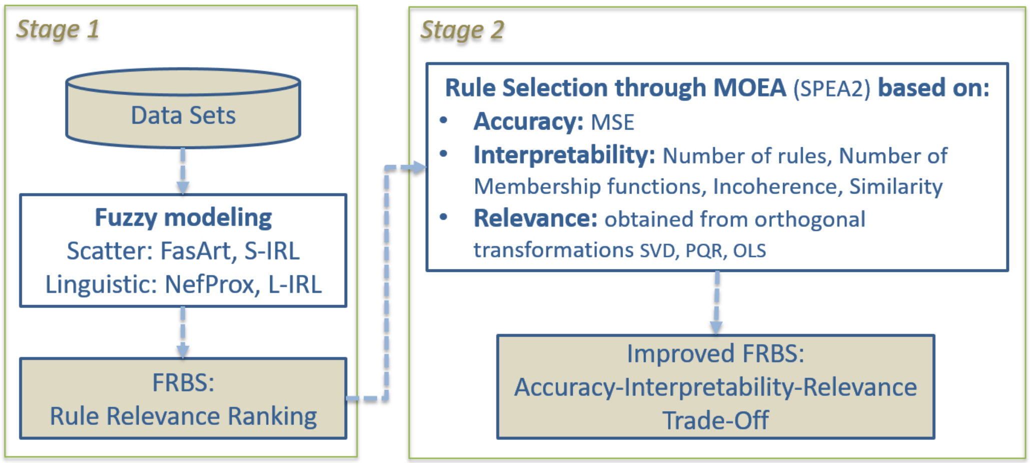

Figure 1: Improvement of FRBS Accuracy-Interpretability trade-off based on rule selection involving accuracy, interpretability and relevance concepts.

✓ *Maximize* Accuracy.

✓ *Maximize* Interpretability.

✓ *Maximize* Relevance.

Once the objectives have been established, the indexes or metrics to measure them have to be defined:

1. **Accuracy** of an FRBS is usually measured by its Mean Squared Error (MSE) (Eq. 3), where $|N|$ is the size of the dataset, $F(x_i)$ is the FRBS output when the input is the i-th sample, and y_i is the known desired output. MSE must be minimized in order to maximize accuracy.

$$MSE = \frac{1}{|N|}\sum_{i=1}^{|N|}(F(x_i) - y_i)^2 \tag{3}$$

2. **Interpretability** of the FRBS can be evaluated in different ways, as described in Section 3. Here, the indexes have been selected according to the taxonomy proposed by Gacto et al. (2011). In all cases, these indexes must be minimized in order to maximize interpretability:
   (a) The *number of rules* measured by Eq. 4 as an index of complexity-based interpretability at the rule base level.

$$NR = Number\ of\ rules \tag{4}$$

(b) The *number of membership functions* measured by Eq. 5 as an index of complexity-based interpretability at the fuzzy partition level.

$$NMF = Number\ of\ membership\ functions \tag{5}$$

(c) The *incoherence* of the rule base evaluated by Eq. 6 as an index of semantic-based interpretability at the rule base level, considering the consistency of a rule base as the absence of contradictory rules (same antecedents but different consequents).

$$\begin{aligned} &Incoherence = Inc = \tfrac{|(S_{kA}(R_i,R_j)>(1-\beta_I)\ AND\ S_{kC}(R_i,R_j)<\beta_I)|}{(RuleNumber-1)!} \\ &\forall 1 \leq i < j \leq RuleNumber \\ &\forall 1 \leq kA \leq AntecedentNumber \\ &\forall 1 \leq kC \leq ConsequentNumber \end{aligned} \tag{6}$$

A threshold for incoherence, β_I, is defined to evaluate the averaged non similarity of the consequents (S_{kC}) and averaged similarity of the antecedents (S_{kA}). $|\cdot|$ is the cardinality of the set, and similarity of the fuzzy sets (A, B) is measured as in Setnes et al. (1998).

(d) The *distinguishability* of the fuzzy sets evaluated by Eq. 7 as an index of semantic-based interpretability at the fuzzy partition level. This distinguishability of the fuzzy sets gets worse when membership functions are more similar for each fuzzy partition.

$$\begin{aligned} &Similarity = Sim = F_{kA}(F_{l,m}(S(MF_{kA.l}, MF_{kA.m}))) \\ &F \Rightarrow ArithmeticMean \\ &\forall 1 \leq l < m \leq MembershipFunctionNumber \\ &\forall 1 \leq kA \leq AntecedentNumber \end{aligned} \tag{7}$$

3. **Rule Relevance** of the FRBS, the idea is to estimate the relevance of each rule in order to rate the relevance of all the rules contained in the FRBS. Here, rule relevance is based on orthogonal transformations (*singular values* in SVD, *R-values* in P-QR and *variances* in OLS) to approach each rule's own variability regarding the input/output data. Using this local evaluation of each rule relevance, the relevance contained in the rule set of an FRBS is defined by Eq. 8, in such a way that $\sum_{i=1}^{n} Relevance_{Rule_i} = 1$ in the initial FRBS.

$$Relevance = Rel_{FS} = \frac{\sum_{i=1}^{n} Relevance_{Rule_i}}{n} \tag{8}$$

Two strategies have been taken into account to maximize the relevance preserved by the FRBS rules in the tradeoff:

(a) Preserving the most k relevant rules in the FRBS, minimizing Eq. 9. This strategy is based on the relevance of every rule:

$$Rel_{RH} = 1 - Rel_{FS} = 1 - \frac{\sum_{i=1}^{k} Relevance_{Rule_i}}{k} \tag{9}$$

(b) Removing the lesser j relevant rules in the FRBS, minimizing Eq. 10. This is formulated as:

$$Rel_{RL} = \sqrt[k]{\prod_{i=1}^{k} (1 - Relevance_{Rule_i})} \tag{10}$$

In order to reach the Accuracy-Interpretability-Relevance trade-off, a post-processing rule selection is performed with SPEA2. A multi-objective strategy for this challenge is an effective approach, since accuracy and interpretability are contradictory objectives. However, when semantic interpretability is managed by MOEAs, it does not perform as well as when complexity based interpretability is used.

6. Experimental Work

In order to evaluate this proposal, exhaustive experimental work has been carried out focusing on two targeted objectives:

1. To improve the Accuracy-Interpretability-Relevance trade-off for FRBSs. The Accuracy-Interpretability trade-off is the major and ordinary point for FRBS literature on this area, but here rule Relevance plays a key role for this procedure. The Accuracy-Relevance and Relevance-Interpretability views are also analyzed.
2. To check rule relevance for the achievement of a good Accuracy-Interpretability trade-off: Should rules with low relevance always be removed? Must rules with high relevance always be preserved?

In order to carry out this experimental work, checking its soundness, several alternatives for the FRBS model, rule relevance, interpretability indexes, MOEA and datasets have been used throughout the experiments:

- FasArt and S-IRL as scatter algorithms, and NefProx and L-IRL as linguistic algorithms, to obtain the initial FRBSs.
- SVD, P-QR and OLS as orthogonal transformations to evaluate rule relevance. The two rule relevance strategies: Rel_{RH} (Eq.9) and Rel_{RL} (Eq.10), have been used in experiments.
- NR, NMF, Sim and Inc as four different measurements of interpretability, according to Section 5.
- SPEA2 as MOEA approaches to guide the rule selection.
- Nine datasets describing real-world problems from the KEEL dataset repository (Alcalá-Fdez et al., 2009), whose main characteristics are summarized in Table 1.

Table 1: Datasets considered for experimental work

Datasets	**Name**	**Variables**	**Patterns**
Plastic Strength	PLA	3	1650
Quake	QUA	4	2178
Electrical Maintenance	ELE	5	1056
Abalone	ABA	9	4177
Stock prices	STP	10	950
Weather Izmir	WIZ	10	1461
Weather Ankara	WAN	10	1609
Mortgage	MOR	16	1049
Treasury	TRE	16	1049

Section 6.1 describes in detail the experimental methodologies for both major objectives (Accuracy-Interpretability-Relevance trade-off and the role of rule relevance); while Section 6.2 analyzes and discusses the results obtained when the algorithm SPEA2 is used.

6.1. Methodologies

Two methodologies have been developed for this experimental work according to the major objectives:

1. To improve the Accuracy-Interpretability trade-off for FRBSs, adding rule relevance as a factor to be taken into account.
2. To check the role of rule Relevance for reaching a good Accuracy-Interpretability trade-off, as well as other views: Accuracy-Relevance and Relevance-Interpretability trade-off.

6.1.1. FRBS Improvement: Accuracy, Interpretability and Relevance based on MOEA Rule Selection

Methodology 1 shows the experimental procedure used to carry out the experimental work, focusing on the improvement of the linguistic or scatter FRBSs, through a MOEA based rule selection, taking into account: the Accuracy, Interpretability and Relevance of the rule system. The results of applying this methodology are shown and discussed in Sections 6.2.1 and 6.2.2.

First, FRBSs are generated using the algorithms described in Subsection 4.1 and tuned according to the parameters shown in Table 2. A brief description of them can be seen in Section 4.1. All experiments are based on 5-*fold cross-validation*[4] and, in order to be able to use the Interpretability index Inc, the threshold for *incoherence* is tuned as $\beta_I = 0.2$.

Next, the rule selection based on SPEA2 is run six times. Traditional genetic operators have been selected and specific mechanisms to improve the search algorithm ability have been implemented. These are: a binary coding scheme for individuals, HUX as the crossover operator (Eshelman, 1991), classical mutation (Ishibuchi et al., 1997) with probability 0.2, a mechanism for incest prevention based on concepts of CHC (Eshelman, 1991), a restarting operator to maintain the best individuals in the external population, and the mating pool being progressively reduced. Other parameters are: a population size of 200, an external population size of 61 and $100,000$ evaluations (Alcalá et al., 2007).

Finally, the MOEA based approach permits the results to be analyzed under three different views: Accuracy-Interpretability, Accuracy-Relevance and

[4]The data partitions (5-fold) are available on the KEEL project website: `http://sci2s.ugr.es/keel/datasets.php`

Methodology 1 MOEA based FRBS rule selection based on accuracy, interpretability and relevance

```
for Algorithm Modeling = Scatter:Linguistic do
  for Orthogonal Transformation = SVD : P-QR : OLS do
    for Relevance = Rel_RH : Rel_RL do
      for Interpretability = NR : NMF : Inc : Sim do
        for Dataset = Dataset1 ··· Dataset9 do
          for CrossValidation = CV1 ··· CV5 do
            Generate initial FRBS
            Generate ordering rules according to the relevant rules obtained from
            orthogonal transformations
            for Run=1 a 6 do
              Run Genetic Algorithm SPEA2 for rule selection with three objec-
              tives:
                    Accuracy → min(MSE_tra)
                    Relevance → min(Rel_RL) or min(Rel_RH)
                    Interpretability → min(NR) or min(NMF) or min(Inc) or
              min(Sim)
            end for
          end for
          for View = Accuracy-Relevance : Accuracy-Interpretability : Relevance-
          Interpretability do
            Pareto Front Analysis {Best Acc, Median Acc-Inter, Best Inter}
          end for
        end for
        for View = Accuracy-Relevance : Accuracy-Interpretability : Relevance-
        Interpretability do
          Nonparametric Statistics Test {Best Acc, Median Acc-Inter, Best Inter}
        end for
      end for
    end for
  end for
end for
```

Interpretability-Relevance. Pareto Fronts are analyzed taking into account different representative points, or models, for each view: the best for each objective and the median one. This permits the users to make their own selection based on their preferences regarding the objectives. In any case, the users always have the Median solution as a compromise option, although other options are possible, such as in Ishibuchi and Nojima (2013). Other researchers have also used these kinds of projections for graphical representation and statistical analysis when three objectives are optimized together

Table 2: FRBS Tuning Parameters

#var	FasArt	NefProx	S-IRL and L-IRL
< 9	$\rho = 0.7$,$\gamma = 8$	$nLT = 5$	$nLT = 5$
≥ 9	$\rho = 0.7$,$\gamma = 6$	$nLT = 3$	$nLT = 3$
In any case			$\epsilon = 1.5$,$\omega = 0.05$,$K = 0.1$,$P = 61$
			$a = 0.35$,$b = 5$,$P_c = 0.6$,$P_m = 0.1$
			$Gen = 100$,$ES = 50$,$\alpha = 20\%$

ρ-vigilance parameter, γ-fuzzification rate, nLT-number of linguistic terms for linguistic partitions, ϵ-minimum covering degree, ω-covering for positive examples, K-percentage of negative examples, P-population size, a and b-crossover and mutation, P_c-crossover probability, P_m-mutation probability, Gen-for the number of generations, ES-evolutionary strategy applied until there is no improvement in ES generations over a percentage α of individuals of population

(Nguyena et al., 2015; Márquez et al., 2012; Galende et al., 2012; Antonelli et al., 2011; Gacto et al., 2010). For all cases, a comparison with other approaches and non-parametric statistical tests are applied in order to check the statistical significance of the results obtained.

6.1.2. Impact of Rule Relevance through the FRBSs

Methodology 2 shows the methodology of experimentation applied to evaluate the meaning of rule relevance and its distribution in the improved FRBSs by this multi-objective approach. The results of this study are shown and analyzed in Section 6.2.3.

The analysis involves every algorithm of Subsection 4.1 and all datasets described in Table 1. First, the range of rule relevance for each FRBS is divided into quarters labeled as: *Low Relevance, Low-Medium Relevance, High-Medium Relevance and High Relevance.* Then, the distribution of rules by quarters and the selected rules by quarters is analyzed to check their impact in reaching the Accuracy-Interpretability trade-off.

Methodology 2 Distribution of rules by relevance into improved FRBS

```
for Algorithm Modeling = Scatter:Linguistic do
  for Orthogonal Transformation = SVD : P-QR : OLS do
    for Relevance = Rel_RH : Rel_RL do
      for Interpretability = NR : NMF : Inc : Sim do
        for View = Accuracy-Relevance : Accuracy-Interpretability : Relevance-
        Interpretability do
          for Dataset = Dataset1 ··· Dataset9 do
            maxRel = max_{i=1}^{NR_ini} Relevance_{Rule_i}
            minRel = min_{i=1}^{NR_ini} Relevance_{Rule_i}
            width = (maxRel − minRel) / 4
            for c = 1 ··· 4 do
              Quarter^c = [minRel + (width ∗ (c − 1)), minRel + (width ∗ c)]
              InitialDistribution(%) = (NR^initial in Quarter^c) / NR^initial
              PreservedRules(%) = (NR^improved in Quarter^c) / (NR^initial in Quarter^c)
              ImprovedDistribution(%) = (NR^improved in Quarter^c) / NR^improved
            end for
          end for
          Mean distributions of selected rules are studied {Best Acc, Median Acc-
          Inter, Best Inter}
        end for
      end for
    end for
  end for
end for
```

6.2. Improving FRBS by MOEA Rule Selection: Results and Discussions

The results shown and discussed in this section are based on the methodologies previously commented in Sections 6.1.1 and 6.1.2, and shown in Fig. 1. The experimentation carried out has been very extensive and exhaustive, so it is not possible to present all the results for all the cases in detail. So, only the most contributable results are included in the following sections:

- Section 6.2.1 presents the results corresponding to the Accuracy-Interpretability view because this is the most usual for the scientific community. The results obtained by OLS, which take into account both rule antecedents and consequents, and measurement by relevance Rel_{RH}, are analyzed in detail. Then, results obtained with P-QR and SVD are summarized.

- Section 6.2.2 includes a global analysis for the other views: Accuracy-

Relevance and Relevance-Interpretability.

- Finally, in section 6.2.3, the role of the fuzzy rules is analyzed for the trade-off according to their relevance.

6.2.1. Accuracy-Interpretability Improvement

The results obtained in these cases are summarized in Tables 3, 4 and 5, which show average Δ values for the different interpretability indexes and all datasets, where the Δ value is the variation in % of improved FRBSs with respect to the initial one. The values shown in the tables are: indexes NR (number of rules), NMF (number of membership functions), Inc (incoherence) or Sim (similarity) involved in the Interpretability objective ($Inter$), MSE for training (E_{tra}) involved in the Accuracy objective and mean relevance of selected rules (Rel_{FS}) involved in the Relevance objective. Also the MSE for test (E_{tst}) and number of rules (NR) are also shown. All these values are shown for three representative systems in Pareto Front for the Accuracy-Interpretability view: most Interpretable (Best $Inter$), median Accuracy-Interpretability (Median Acc-$Inter$) and most Accurate (Best Acc). The grey color in the table indicates that the *Wilcoxon test*, performed with a confidence level of $\alpha = 0.1$, accepts that initial and improved models are similar.

Orthogonal Transformation: OLS.

The results obtained are based on the procedure shown in methodology 1. Table 3 shows the results obtained for each scatter and linguistic fuzzy model and interpretability measures when OLS is considered in Rel_{RH}.

Scatter Fuzzy Models:

1. **The most interpretable solutions**. The *Wilcoxon test* accepts that every measure has been improved. Interpretability improves in every case up to 70.20%, the *error* always being significantly improved for S-IRL cases and most FasArt cases up to 70.38%, while the *number of rules* is reduced for every case up to 68.28%. Relevance is always improved for FasArt cases and most S-IRL cases up to 63.36%.
   Summarizing, the rate of improvement is higher in S-IRL than FasArt for most indexes except relevance, but FasArt is generally more accurate. Regarding interpretability indexes, NR, NMF, Inc and Sim behave similarly in FasArt and S-IRL.
2. **The median solutions**. The *Wilcoxon test* accepts that most indexes have been improved, except in two cases of S-IRL, where they remain

Table 3: Average Δ Values (%) for Improved FRBS: Accuracy-Interpretability View, OLS, Rel_{RH}

FasArt

Inter	**Best *Inter***					**Median *Acc-Inter***					**Best *Acc***				
	Inter	E_{tra}	E_{tst}	NR	Rel_{FS}	*Inter*	E_{tra}	E_{tst}	NR	Rel_{FS}	*Inter*	E_{tra}	E_{tst}	NR	Rel_{FS}
NR	50.28	30.55	31.08	50.28	63.15	39.12	53.07	51.88	39.12	29.84	27.26	55.51	54.02	27.26	11.17
NMF	51.37	32.16	32.48	49.90	63.36	39.98	53.29	52.00	38.81	29.75	28.10	55.50	53.91	27.26	11.20
Inc	55.53	54.44	52.93	29.15	16.01	22.90	55.14	53.53	26.45	11.11	−0.76	55.47	53.87	26.07	9.97
Sim	27.61	22.52	25.15	46.54	48.65	24.17	48.81	47.99	40.48	31.86	17.91	54.99	53.64	34.77	26.12

S-IRL

Inter	**Best *Inter***					**Median *Acc-Inter***					**Best *Acc***				
	Inter	E_{tra}	E_{tst}	NR	Rel_{FS}	*Inter*	E_{tra}	E_{tst}	NR	Rel_{FS}	*Inter*	E_{tra}	E_{tst}	NR	Rel_{FS}
NR	68.28	63.96	62.45	68.28	46.74	58.31	70.65	69.43	58.31	18.38	49.50	71.85	70.77	49.50	1.09
NMF	70.20	63.54	62.26	68.26	46.24	60.10	70.44	69.22	58.45	18.62	50.84	71.85	70.77	49.46	1.02
Inc	55.56	71.48	70.38	47.75	3.73	11.79	71.59	70.52	47.37	3.04	−11.58	71.65	70.57	47.48	3.57
Sim	37.29	52.18	50.94	64.28	23.06	34.05	66.87	65.42	56.63	4.98	28.64	71.87	70.71	48.13	0.30

NefProx

Inter	**Best *Inter***					**Median *Acc-Inter***					**Best *Acc***				
	Inter	E_{tra}	E_{tst}	NR	Rel_{FS}	*Inter*	E_{tra}	E_{tst}	NR	Rel_{FS}	*Inter*	E_{tra}	E_{tst}	NR	Rel_{FS}
NR	86.80	31.14	30.85	86.80	580.30	77.44	54.48	53.50	77.44	182.42	63.24	61.23	60.44	63.24	101.83
NMF	75.84	42.30	42.48	74.29	188.06	67.45	59.02	58.27	66.36	58.44	55.22	61.53	60.86	54.54	53.97
Inc	65.96	60.02	59.27	57.15	61.59	56.93	60.76	60.06	55.04	47.74	39.58	61.34	60.60	53.10	57.47
Sim	43.05	34.62	34.85	72.88	63.48	38.85	52.86	52.05	68.25	36.18	30.47	60.74	60.17	59.74	84.18

L-IRL

Inter	**Best *Inter***					**Median *Acc-Inter***					**Best *Acc***				
	Inter	E_{tra}	E_{tst}	NR	Rel_{FS}	*Inter*	E_{tra}	E_{tst}	NR	Rel_{FS}	*Inter*	E_{tra}	E_{tst}	NR	Rel_{FS}
NR	83.19	40.58	40.06	83.19	202.50	75.82	63.95	63.18	75.82	124.92	54.14	71.60	71.02	54.14	70.24
NMF	85.06	42.89	42.11	83.41	205.97	77.39	64.22	63.41	75.90	127.11	61.92	70.57	69.97	60.72	78.46
Inc	99.82	68.49	67.67	65.11	111.45	81.84	69.98	69.13	62.29	85.39	40.36	70.59	69.76	60.77	79.76
Sim	44.50	34.67	33.00	81.62	108.57	39.27	61.10	60.19	74.94	4.91	31.50	70.53	69.85	59.99	59.11

the same. Interpretability has improved for most FasArt and S-IRL cases up to 60.10%, while the *error* is significantly reduced for every case up to 70.52%, and the *number of rules* is reduced for every case up to 58.45%. Relevance is improved for most FasArt and S-IRL cases up to 31.86%.

In general, the rate of improvement is higher for S-IRL than FasArt for most indexes except Relevance, but FasArt is more accurate. Regarding the interpretability indexes, NR, NMF, Inc and Sim behave similarly in FasArt and S-IRL with Interpretability and *error*, but not similarly in *number of rules* and Relevance.

3. **The most accurate solutions**.The *Wilcoxon test* accepts that most indexes have been improved except for one case of FasArt and two cases of S-IRL, where they remain the same. Interpretability has improved for every case of NR, NMF and Sim up to 50.84%, while *error* is significantly reduced for every case up to 70.77%, and the *number of*

rules is improved for every case up to 49.50%. Relevance is improved in most FasArt and S-IRL cases up to 26.12%.

So, the rate of improvement is higher in S-IRL than FasArt for most indexes except Relevance, but FasArt is more accurate. Regarding interpretability indexes, NR, NMF, Inc and Sim behave similarly in FasArt and S-IRL with Interpretability and *error*, but not similarly in *number of rules* and Relevance.

Linguistic Fuzzy Models:

1. **The most interpretable solutions**. The *Wilcoxon test* accepts an improvement for every case. Interpretability has improved for every case up to 99.82%, the *error* is significantly improved in most cases up to 67.67%, and the *number of rules* is reduced for every case up to 86.80%. Relevance is also improved for most cases up to 580.30%.
   So, L-IRL generally improves more than NefProx. Regarding the interpretability indexes, NR, NMF, Inc and Sim behave similarly in NefProx and L-IRL.
2. **The median solutions**. The *Wilcoxon test* accepts an improvement for every case. Interpretability has improved for every case up to 81.84%, the *error* is reduced for every case up to 69.13%, and the *number of rules* is significantly reduced for every case up to 77.44%. Relevance is always improved for L-IRL cases and most NefProx cases up to 182.42%.
   Regarding FRBS models, L-IRL generally improves more than NefProx. Regarding the interpretability indexes, NR, NMF, Inc and Sim, behave similarly in NefProx and L-IRL with Interpretability and *error*, but not similarly in *number of rules* and Relevance.
3. **The most accurate solutions**. The *Wilcoxon test* accepts that every analyzed measure has been improved. Interpretability has improved in every NefProx case, and every L-IRL case of $Inter = NR$, $Inter = NMF$ and $Inter = Sim$ up to 61.92%, while the *error* is significantly reduced for every case up to 71.02%, and the *number of rules* has improved for every case up to 63.24%. Relevance is improved in most cases up to 101.83%.
   Thus, L-IRL generally improves more than NefProx. Regarding the interpretability indexes, NR, NMF, Inc and Sim, are quite similar with *error* for NefProx and L-IRL, but for Interpretability, *number of rules* and Relevance, they are not similar.

Orthogonal Transformation: P-QR and SVD. To avoid a very long article for these orthogonal transformations, only average results for all datasets are shown. Table 4 for P-QR and Table 5 for SVD show the mean results for every (scatter and linguistic) fuzzy model and interpretability measure, considering Rel_{RH}. Both orthogonal transformations achieve improvements in Accuracy, Interpretability and Relevance for 98% of scatter models and 100% of linguistic ones. On average, interpretability indexes have been improved by up to 72% for scatter models and 100% for linguistic ones; the *error* has been reduced by 71% in all models; and Relevance has been improved by up to 33% and 199%, respectively. There are no significant differences between P-QR and SVD.

Table 4: Average Δ Values (%) for Improved FRBS: Accuracy-Interpretability View, P-QR, Rel_{RH}

FasArt

Inter	**Best *Inter***					**Median *Acc-Inter***					**Best *Acc***				
	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}
NR	48.75	38.53	37.03	48.75	32.86	37.86	53.63	52.17	37.86	19.65	26.86	55.46	53.83	26.86	10.20
NMF	50.20	38.09	37.49	48.78	32.78	38.95	53.70	52.25	37.81	19.60	27.68	55.46	53.83	26.86	10.20
Inc	55.56	54.34	52.96	27.50	11.96	27.99	55.04	53.51	25.03	10.13	3.30	55.39	53.78	25.03	9.79
Sim	26.88	26.71	27.98	45.59	32.35	23.42	49.35	48.29	39.43	29.34	17.12	54.97	53.63	32.33	24.50

S-IRL

Inter	**Best *Inter***					**Median *Acc-Inter***					**Best *Acc***				
	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}
NR	68.01	65.36	64.33	68.01	28.57	58.15	70.61	69.43	58.15	14.70	49.82	71.73	70.61	49.82	5.95
NMF	69.97	64.03	62.99	68.04	29.00	59.91	70.64	69.40	58.26	14.69	51.22	71.73	70.61	49.82	5.95
Inc	55.56	71.15	70.07	47.14	6.26	21.25	71.20	70.14	46.72	5.88	5.61	71.26	70.23	46.81	5.89
Sim	36.41	53.80	52.82	61.23	8.89	33.54	66.78	65.47	54.60	5.17	28.86	71.61	70.11	46.72	3.20

NefProx

Inter	**Best *Inter***					**Median *Acc-Inter***					**Best *Acc***				
	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}
NR	87.67	35.67	35.81	87.67	140.78	78.54	54.39	53.97	78.54	52.43	62.52	60.89	60.25	65.21	34.59
NMF	77.51	42.01	42.03	75.92	59.34	69.20	59.16	58.24	68.08	20.27	55.73	61.37	60.62	55.06	15.65
Inc	66.67	59.54	58.83	57.93	26.10	56.81	60.62	59.84	55.40	20.35	35.60	61.11	60.49	52.81	16.84
Sim	43.38	34.87	35.03	73.23	55.50	39.03	51.90	51.39	68.74	30.97	31.17	60.63	60.22	59.16	25.28

L-IRL

Inter	**Best *Inter***					**Median *Acc-Inter***					**Best *Acc***				
	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}
NR	83.43	41.72	41.26	83.43	134.04	75.42	64.52	63.74	75.42	97.45	54.20	71.53	70.97	54.20	51.93
NMF	85.16	42.46	42.83	83.52	140.29	77.28	63.63	63.10	75.79	98.69	62.32	70.52	69.98	61.11	62.50
Inc	99.63	68.40	67.78	65.55	76.39	77.04	69.80	69.17	62.55	65.82	15.77	70.64	70.04	59.85	60.28
Sim	44.10	35.84	34.92	81.22	101.83	39.43	60.93	60.22	75.02	81.44	30.75	70.50	69.79	60.38	56.94

In most cases, the higher percentages of Accuracy improvement are obtained when the Relevance is lower. The same conclusion is possible when *incoherence* and *similarity* are involved, due to the interpretability indexes for the selection of rules. On the other hand, this variation is higher for

Table 5: Average Δ Values (%) for Improved FRBS: Accuracy-Interpretability View, SVD, Rel_{RH}

FasArt

Inter	**Best *Inter***					**Median *Acc-Inter***					**Best *Acc***				
	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}
NR	51.80	21.88	23.10	51.80	19.49	40.40	52.35	50.74	40.40	9.26	26.46	55.44	53.94	26.46	3.27
NMF	53.38	21.77	22.92	51.87	19.52	41.55	52.37	50.82	40.35	9.10	27.36	55.43	53.91	26.53	3.03
Inc	55.44	54.21	52.96	26.70	5.94	25.39	55.02	53.51	24.82	4.14	3.49	55.39	53.85	24.53	4.01
Sim	26.46	12.85	15.08	43.07	26.93	23.01	45.77	44.98	36.19	17.89	16.44	54.90	53.63	29.04	12.72

S-IRL

Inter	**Best *Inter***					**Median *Acc-Inter***					**Best *Acc***				
	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}
NR	70.22	64.82	63.53	70.22	33.03	59.80	70.50	69.32	59.80	4.14	49.41	71.79	70.65	49.41	8.49
NMF	72.14	63.40	62.05	70.14	32.20	61.51	70.36	69.21	59.81	14.32	50.53	71.79	70.66	49.15	8.44
Inc	55.56	71.29	70.10	47.01	9.37	22.71	71.41	70.27	46.72	9.22	−7.82	71.47	70.09	46.86	9.10
Sim	37.44	52.37	51.84	62.31	24.52	33.99	67.11	65.97	54.90	13.52	28.78	71.76	70.68	46.82	7.70

NefProx

Inter	**Best *Inter***					**Median *Acc-Inter***					**Best *Acc***				
	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}
NR	84.58	34.90	35.33	84.58	199.01	76.22	53.50	52.92	76.22	77.13	62.57	60.56	59.79	62.57	43.14
NMF	79.53	40.54	40.91	77.50	85.99	70.47	58.72	57.96	69.06	25.06	55.30	61.41	60.65	54.63	23.49
Inc	66.07	59.31	58.50	55.54	39.10	56.88	60.34	59.38	53.65	32.87	39.31	60.99	60.08	51.11	24.05
Sim	43.21	32.95	33.07	73.35	6.19	38.96	51.22	50.70	68.79	47.57	31.25	60.49	60.11	59.60	45.55

L-IRL

Inter	**Best *Inter***					**Median *Acc-Inter***					**Best *Acc***				
	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}	$Inter$	E_{tra}	E_{tst}	NR	Rel_{FS}
NR	83.07	44.47	44.36	83.07	138.59	75.00	65.07	64.43	75.00	88.90	54.15	71.39	70.83	54.15	48.59
NMF	84.55	44.62	44.43	82.92	136.46	76.80	64.30	63.85	75.33	88.77	61.68	70.52	70.05	60.49	54.50
Inc	99.84	67.91	67.31	64.96	74.30	79.04	69.96	69.40	61.58	55.33	42.08	70.67	70.14	59.45	51.65
Sim	43.54	37.64	37.18	80.88	103.50	39.09	62.37	61.95	74.73	67.91	30.85	70.44	69.76	60.58	50.48

linguistic modeling than for scatter ones, and this suggests that low relevance rules are very important for Accuracy and Interpretability. Therefore, this implies paying attention to this issue for a better, or correct, Accuracy-Interpretability trade-off.

Finally, in order to compare the fuzzy models of this proposal, a comparison has been made using the following approaches: WM-R (Wang and Mendel, 1992) and GFS-RS-T (Cordón and Herrera, 1997; Ishibuchi et al., 1995), included in the KEEL software tool[5]; S_{SP2} $(MAXACC)$ obtained from Gacto et al. (2010) and ScatA[6], ScatB[7], LingA[8] and LingB[9] reported in

[5] `http://sci2s.ugr.es/keel/algorithms.php`

[6] FasArt Compact + S_{NSGAII}+$Inter_C$

[7] FasArt Complex + S_{NSGAII}+$Inter_C$

[8] NefProx Compact + S_{NSGAII}+$Inter_C$

[9] NefProx Complex + S_{NSGAII}+$Inter_C$

Table 6: Accuracy and Interpretability based comparison: Accuracy/Interpretability View, OLS, Rel_{RH}, $Inter = NR$

Datasets	FRBS model	Best *Inter*			Median *Acc-Inter*			Best *Acc*		
		E_{tst}	NR	$Inter$	E_{tst}	NR	$Inter$	E_{tst}	NR	$Inter$
PLA	FasArt	3.802	7.3	7.3	1.981	13.5	13.5	1.778	21.7	21.7
	S-IRL	2.773	29.1	29.1	2.332	40.1	40.1	2.249	52.6	52.6
	NefProx	3.808	6.2	6.2	3.225	10.2	10.2	2.632	13.5	13.5
	L-IRL	5.230	9.1	9.1	2.477	15.8	15.8	2.029	27.4	27.4
	LingA	4.391	10.6	10.6	3.572	13.0	13.0	3.222	17.0	17.0
	LingB	3.286	18.0	18.0	2.824	23.0	23.0	2.655	29.6	29.6
	ScatA	3.718	18.8	18.8	3.073	22.0	22.0	2.688	27.2	27.2
	ScatB	3.172	36.2	36.2	2.567	45.8	45.8	2.374	55.4	55.4
	WM-R	7.114	14.8	14.8	7.114	14.8	14.8	7.114	14.8	14.8
	GFS-RS-T	18.261	7.5	7.5	18.261	7.5	7.5	18.261	7.5	7.5
	S_{SP2}	4.832	12.0	12.0	4.832	12.0	12.0	4.832	12.0	12.0
WIZ	FasArt	4.502	49.2	49.2	4.131	52.5	52.5	4.141	55.7	55.7
	S-IRL	16.467	8.2	8.2	14.747	10.9	10.9	14.253	13.2	13.2
	NefProx	8.383	4.8	4.8	5.570	11.9	11.9	4.467	22.7	22.7
	L-IRL	16.950	8.6	8.6	7.851	13.3	13.3	5.319	21.7	21.7
	LingA	13.271	357.8	357.8	12.998	363.0	363.0	12.804	375.4	375.4
	LingB	19.099	373.6	373.6	18.829	380.6	380.6	18.860	392.8	392.8
	ScatA	17.571	116.0	116.0	17.401	124.4	124.4	16.701	134.5	134.5
	ScatB	23.251	358.0	358.0	23.182	365.9	365.9	23.067	380.5	380.5
	WM-R	14.736	104.8	104.8	14.736	104.8	104.8	14.736	104.8	104.8
	GFS-RS-T	7.346	51.2	51.2	7.346	51.2	51.2	7.346	51.2	51.2
	S_{SP2}	15.482	38.7	38.7	15.482	38.7	38.7	15.482	38.7	38.7
MOR	FasArt	0.583	10.1	10.1	0.235	14.3	14.3	0.198	18.6	18.6
	S-IRL	0.725	6.5	6.5	0.430	9.4	9.4	0.388	11.6	11.6
	NefProx	1.049	3.2	3.2	0.338	9.4	9.4	0.259	19.9	19.9
	L-IRL	0.845	4.7	4.7	0.585	6.9	6.9	0.455	10.0	10.0
	LingA	0.975	76.6	76.6	0.660	82.8	82.8	0.632	98.0	98.0
	LingB	0.415	171.6	171.6	0.382	179.0	179.0	0.377	191.8	191.8
	ScatA	1.266	36.8	36.8	1.216	38.8	38.8	1.178	41.4	41.4
	ScatB	0.447	64.1	64.1	0.381	69.3	69.3	0.373	79.5	79.5
	WM-R	1.946	77.6	77.6	1.946	77.6	77.6	1.946	77.6	77.6
	GFS-RS-T	0.750	24.3	24.3	0.750	24.3	24.3	0.750	24.3	24.3
	S_{SP2}	0.510	17.2	17.2	0.510	17.2	17.2	0.510	17.2	17.2

Galende et al. (2012). This comparison is based on three indexes: E_{tst}, NR and $Inter$. In order to avoid an overlong paper, only the comparison based on OLS and Rel_{RH} is shown; in the other cases, no significant differences are found. In the four latter works used for comparison, only the Best $Inter$, Median *Acc*-$Inter$ and Best *Acc* have been reported, while in the other cases, only Best *Acc* is available for comparison. In this latter case, it is compared with the Best $Inter$, Median *Acc*-$Inter$ and Best *Acc* of this proposal.

For the same reason, only a sample of the global comparison results are shown in Tables 6 and 7. These correspond to the interpretability indexes $Inter = NR$ and $Inter = Sim$, and the datasets: PLA, WIZ and MOR. Analyzing these results, it is possible to check that the fuzzy models reported in this approach are very competitive. In any case, to check the statistical impact of the comparison, some non-parametric tests for multiple comparison

Table 7: Accuracy and Interpretability based comparison: Accuracy/Interpretability View, OLS, Rel_{RH}, $Inter = Sim$

Datasets	FRBS model	Best ***Inter***			Median ***Acc-Inter***			Best ***Acc***		
		E_{tst}	NR	$Inter$	E_{tst}	NR	$Inter$	E_{tst}	NR	$Inter$
PLA	FasArt	3.966	10.2	0.121	2.262	13.3	0.138	1.784	19.1	0.165
	S-IRL	3.264	45.9	0.105	2.525	51.6	0.113	2.275	58.5	0.131
	NefProx	2.926	12.8	0.183	2.787	13.2	0.189	2.633	13.7	0.203
	L-IRL	6.279	8.8	0.161	3.146	13.7	0.193	2.037	25.8	0.234
	LingA	4.391	10.6	0.195	3.572	13.0	0.231	3.222	17.0	0.241
	LingB	3.286	18.0	0.164	2.824	23.0	0.172	2.655	29.6	0.186
	ScatA	3.718	18.8	0.151	3.073	22.0	0.176	2.688	27.2	0.254
	ScatB	3.172	36.2	0.132	2.567	45.8	0.143	2.374	55.4	0.167
	WM-R	7.114	14.8	0.244	7.114	14.8	0.244	7.114	14.8	0.244
	GFS-RS-T	18.261	7.5	0.234	18.261	7.5	0.234	18.261	7.5	0.234
	S_{SP2}	4.832	12.0	0.233	4.832	12.0	0.233	4.832	12.0	0.233
WIZ	FasArt	5.546	51.4	0.333	4.414	53.9	0.339	4.143	56.7	0.351
	S-IRL	20.777	9.5	0.393	16.772	11.1	0.401	14.215	13.4	0.414
	NefProx	12.864	9.1	0.510	7.358	12.0	0.542	4.343	22.5	0.610
	L-IRL	13.366	11.3	0.546	7.323	14.7	0.565	5.494	20.5	0.605
	LingA	13.271	357.8	0.557	12.998	363.0	0.559	12.804	375.4	0.561
	LingB	19.099	373.6	0.489	18.829	380.6	0.491	18.860	392.8	0.494
	ScatA	17.571	116.0	0.349	17.401	124.4	0.354	16.701	134.5	0.360
	ScatB	23.251	358.0	0.399	23.182	365.9	0.401	23.067	380.5	0.404
	WM-R	14.736	104.8	0.722	14.736	104.8	0.722	14.736	104.8	0.722
	GFS-RS-T	7.346	51.2	0.710	7.346	51.2	0.710	7.346	51.2	0.710
	S_{SP2}	15.482	38.7	0.716	15.482	38.7	0.716	15.482	38.7	0.716
MOR	FasArt	0.550	12.7	0.192	0.261	14.8	0.202	0.203	17.1	0.216
	S-IRL	1.299	5.8	0.282	0.578	8.6	0.312	0.389	11.5	0.339
	NefProx	0.949	6.5	0.340	0.451	10.2	0.391	0.264	17.2	0.471
	L-IRL	0.910	5.5	0.314	0.574	8.1	0.371	0.457	10.9	0.403
	LingA	0.975	76.6	0.403	0.660	82.8	0.405	0.632	98.0	0.409
	LingB	0.415	171.6	0.288	0.382	179.0	0.289	0.377	191.8	0.291
	ScatA	1.266	36.8	0.327	1.216	38.8	0.340	1.178	41.4	0.353
	ScatB	0.447	64.1	0.317	0.381	69.3	0.322	0.373	79.5	0.327
	WM-R	1.946	77.6	0.580	1.946	77.6	0.580	1.946	77.6	0.580
	GFS-RS-T	0.750	24.3	0.557	0.750	24.3	0.557	0.750	24.3	0.557
	S_{SP2}	0.510	17.2	0.553	0.510	17.2	0.553	0.510	17.2	0.553

have been done: Friedman's test, Iman and Davenport's test and Holm's *post-hoc* test.

To perform these tests, we used a confidence level $\alpha = 0.1$. Table 8 shows a summary for the different interpretability indexes, considering the Accuracy-Interpretability view, OLS and Rel_{RH}. The dark grey color identifies the model with the best Friedman's ranking, while the light grey color indicates that Holm's *post-hoc* test accepts that this model is similar to the winner. As shown in the table, the fuzzy models based on this proposal are the best in most cases, and equivalent to other approaches reported by other authors, not only in interpretability, but also in accuracy. Similar results have been obtained by the rest of the orthogonal transformations.

Table 8: Rankings obtained through Friedman's Test on the different measures, Iman-Davenport p-values and Holm's *post-hoc* similarities: Accuracy/Interpretability View, OLS, Rel_{RH}

Inter	**FRBS model**	**Best *Inter***			**Median *Acc-Inter***			**Best *Acc***		
		E_{tst}	NR	$Inter$	E_{tst}	NR	$Inter$	E_{tst}	NR	$Inter$
NR	FasArt	2.33	4.78	4.78	1.22	5.44	5.44	1.28	5.67	5.67
	S-IRL	5.22	4.11	4.11	5.11	3.72	3.72	5.39	3.11	3.11
	NefProx	6.33	1.22	1.22	4.89	2.00	2.00	4.61	2.78	2.78
	L-IRL	7.33	2.89	2.89	5.22	3.17	3.17	3.61	4.11	4.11
	LingA	6.50	7.44	7.44	7.22	7.89	7.89	7.33	8.00	8.00
	LingB	5.44	9.67	9.67	6.33	9.67	9.67	6.56	9.78	9.78
	ScatA	7.17	7.89	7.89	7.56	7.89	7.89	7.72	7.67	7.67
	ScatB	4.67	9.89	9.89	4.78	9.89	9.89	4.83	10.00	10.00
	WM-R	9.33	8.11	8.11	9.56	7.44	7.44	9.67	7.00	7.00
	GFS-RS-T	6.56	5.28	5.28	7.89	4.72	4.72	8.33	4.50	4.50
	S_{SP2}	5.11	4.72	4.72	6.22	4.17	4.17	6.67	3.39	3.39
	p-values	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NMF	FasArt	2.44	4.78	4.78	1.22	5.44	5.50	1.39	5.56	5.56
	S-IRL	5.17	3.94	3.94	5.06	3.44	3.44	5.33	3.11	3.11
	NefProx	5.50	2.11	1.94	4.67	3.28	3.22	4.33	3.67	3.67
	L-IRL	7.22	2.56	2.56	5.06	2.83	2.83	4.17	3.44	3.44
	LingA	6.72	7.33	7.33	7.22	7.78	7.78	7.22	8.00	8.00
	LingB	5.56	9.67	9.67	6.44	9.67	9.67	6.44	9.78	9.78
	ScatA	7.39	7.89	7.89	7.67	7.89	7.78	7.61	7.78	7.78
	ScatB	4.67	9.89	9.89	4.78	9.89	9.89	4.72	10.00	10.00
	WM-R	9.44	8.11	8.11	9.67	7.44	7.56	9.67	7.00	7.00
	GFS-RS-T	6.67	5.17	5.11	8.00	4.61	4.56	8.33	4.50	4.44
	S_{SP2}	5.22	4.56	4.78	6.22	3.72	3.78	6.78	3.17	3.22
	p-values	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Inc	FasArt	1.22	6.11	4.11	1.22	6.22	5.06	1.28	5.56	4.94
	S-IRL	4.72	3.33	4.11	5.06	3.33	3.89	5.22	3.33	3.83
	NefProx	4.22	3.83	4.44	4.44	3.83	6.00	4.44	3.72	6.11
	L-IRL	4.28	3.28	4.44	4.28	3.44	6.56	4.17	3.39	7.56
	LingA	7.28	6.78	7.22	7.33	7.67	6.56	7.22	8.00	6.11
	LingB	6.22	9.22	6.56	6.44	9.56	5.39	6.44	9.78	4.94
	ScatA	7.94	7.67	4.72	7.67	7.44	3.78	7.61	7.78	4.22
	ScatB	5.78	9.78	4.72	4.89	9.78	3.78	4.72	9.89	4.06
	WM-R	9.67	7.72	8.83	9.67	7.28	8.61	9.67	6.94	8.39
	GFS-RS-T	8.00	4.67	8.33	8.22	4.28	8.11	8.33	4.33	7.78
	S_{SP2}	6.67	3.61	8.50	6.78	3.17	8.28	6.89	3.28	8.06
	p-values	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.004
Sim	FasArt	2.61	5.33	1.33	1.11	5.67	1.33	1.28	5.22	1.44
	S-IRL	7.22	4.00	2.89	6.00	3.33	3.06	5.22	3.22	3.56
	NefProx	5.39	2.83	6.22	4.89	3.22	6.33	4.56	3.50	7.17
	L-IRL	7.56	1.94	5.56	5.89	2.44	6.33	4.17	3.67	7.00
	LingA	6.39	7.33	7.67	7.22	7.78	7.78	7.22	8.00	7.22
	LingB	5.00	9.56	4.56	6.11	9.78	3.89	6.44	9.78	3.78
	ScatA	6.83	7.78	3.67	7.44	7.78	3.44	7.61	7.78	3.67
	ScatB	4.33	9.78	4.22	4.33	9.78	4.17	4.72	9.89	3.44
	WM-R	9.33	8.00	10.83	9.56	7.67	10.72	9.67	7.00	10.61
	GFS-RS-T	6.44	4.94	9.83	7.56	4.61	9.72	8.33	4.39	9.44
	S_{SP2}	4.89	4.50	9.22	5.89	3.94	9.22	6.78	3.56	8.67
	p-values	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

6.2.2. Accuracy-Relevance and Relevance-Interpretability View

In addition to the Accuracy-Interpretability trade-off or view, the most popular one in this domain, this experimental work also studies the other two views: Accuracy-Relevance and Relevance-Interpretability. Table 9 shows the behavior of accuracy according to the average rule relevance (Accuracy-Relevance view), and Table 10 shows the behavior of interpretability according to the averaged rule relevance (Relevance-Interpretability view) for OLS, P-QR and SVD. These tables show: Interpretability ($Inter$), MSE for the test (E_{tst}), and the mean relevance of selected rules (Rel_{FS}).

All these values are shown for the three representative models in the Pareto Front: most Relevant (Best Rel), median Accuracy-Relevance (Median $Acc - Rel$) and most Accurate (Best Acc) for the Accuracy-Relevance view (Table 9). In the Relevance-Interpretability view (Table 10): most Interpretable (Best $Inter$), median Relevance-Interpretability (Median Rel-$Inter$) and most Relevant (Best Rel) are shown.

The average minimum (min) and maximum (max) percentages (%) of improvement for every orthogonal transformation and view are shown, considering the scatter and linguistic cases, and an average interpretability index (as averaging of NR, NMF, Inc, Sim) for each representative model.

Table 9: Minimum and Maximum Δ Values (%) for Improved FRBS: Accuracy-Relevance View, Rel_{RH}

Orthogonal Transformation		Best *Rel*			Median *Acc*-*Rel*			Best *Acc*		
Scatter Fuzzy Models		$Inter$	E_{tst}	Rel_{FS}	$Inter$	E_{tst}	Rel_{FS}	$Inter$	E_{tst}	Rel_{FS}
OLS	min (%)	18.35	29.92	58.36	17.89	49.47	25.64	17.91	53.64	9.97
	max (%)	68.97	64.50	106.14	59.17	69.14	64.3	50.84	70.77	26.12
P-QR	min (%)	18.88	41.20	31.25	17.67	50.48	17.45	17.12	53.63	5.89
	max (%)	68.05	63.96	47.43	59.46	68.96	37.93	51.22	70.61	24.50
SVD	min (%)	19.00	33.22	30.66	−31.14	46.31	17.89	16.44	53.63	4.01
	max (%)	68.87	63.45	64.84	59.39	68.85	37.71	50.53	70.68	12.72
Linguistic Fuzzy Models										
Orthogonal Transformation		**Best *Rel***			**Median *Acc*-*Rel***			**Best *Acc***		
		$Inter$	E_{tst}	Rel_{FS}	$Inter$	E_{tst}	Rel_{FS}	$Inter$	E_{tst}	Rel_{FS}
OLS	min (%)	23.12	30.37	336.95	18.24	51.32	205.77	30.47	60.17	53.97
	max (%)	84.48	46.73	805.11	73.85	64.33	407.43	61.92	71.02	101.83
P-QR	min (%)	26.95	32.96	136.27	29.38	51.29	79.52	15.77	60.22	15.65
	max (%)	82.32	45.11	221.93	73.28	63.81	156.57	62.52	70.97	62.50
SVD	min (%)	31.70	30.42	161.61	18.75	50.01	104.43	30.85	59.79	23.49
	max (%)	81.44	45.82	429.90	72.86	64.58	171.81	62.57	70.83	54.50

For both views, the *Wilcoxon test* accepts that every index (Interpretability, MSE and Relevance) has been improved in over 98% of cases, while

Table 10: Minimum and Maximum Δ Values (%) for Improved FRBS: Relevance-Interpretability View, Rel_{RH}

Scatter Fuzzy Models

Orthogonal Transformation		**Best *Inter***			**Median *Rel-Inter***			**Best *Rel***		
		$Inter$	E_{tst}	Rel_{FS}	$Inter$	E_{tst}	Rel_{FS}	$Inter$	E_{tst}	Rel_{FS}
OLS	min (%)	27.61	25.15	16.01	24.77	28.45	55.54	18.48	29.82	58.36
	max (%)	70.20	70.38	63.36	69.67	64.46	79.93	68.97	64.50	106.14
P-QR	min (%)	26.88	27.98	6.26	24.21	34.95	28.85	18.89	41.15	31.25
	max (%)	69.97	70.07	32.86	68.97	63.77	41.70	68.05	63.96	47.47
SVD	min (%)	26.46	22.92	5.94	24.07	16.56	30.01	19.01	33.22	30.66
	max (%)	72.14	70.10	33.03	70.88	63.39	47.95	68.87	63.45	64.84

Linguistic Fuzzy Models

Orthogonal Transformation		**Best *Inter***			**Median *Rel-Inter***			**Best *Rel***		
		$Inter$	E_{tst}	Rel_{FS}	$Inter$	E_{tst}	Rel_{FS}	$Inter$	E_{tst}	Rel_{FS}
OLS	min (%)	43.05	30.85	61.59	38.67	30.40	274.86	23.12	30.37	336.95
	max (%)	99.82	67.67	580.30	94.24	47.33	702.77	84.51	46.73	805.11
P-QR	min (%)	43.38	34.92	26.10	39.25	32.30	107.43	26.95	32.96	136.31
	max (%)	99.63	67.78	140.78	86.88	46.01	188.14	82.32	45.11	222.12
SVD	min (%)	43.21	33.07	39.10	39.11	32.92	151.04	31.70	30.42	161.61
	max (%)	99.84	67.31	199.01	83.57	47.05	250.74	81.44	45.82	429.9

for the remaining 2%, the performance of FRBSs is maintained. The results show similar behavior between OLS, P-QR and SVD for all cases. Analyzing each view separately:

- **Accuracy-Relevance view**: Interpretability is improved by between 29%-47% for scatter models and 44%-64% for linguistic ones. Accuracy improves by between 47%-62% for scatter FRBS and 37%-65% in linguistic FRBS, while Relevance improves by between 8%-76% for scatter and 41%-448% linguistic ones.

- **Relevance-Interpretability view**: Interpretability is improved by between 46% - 53% for scatter models and 58% - 73% for linguistic ones. Accuracy improves by between 46%-54% for scatter models and 37% - 45% for linguistic models; while Relevance improves by between 21% - 76% for scatter models and 92% - 448% for linguistic models.

6.2.3. Improved FRBS: Rules with Low & High Relevance

In this section, a relevance based analysis of the rules selected through the trade-off process is carried out, according to the scheme shown in Methodology 2. The analysis described below is focused on OLS and P-QR orthogonal transformations. The first one takes into account rule antecedents and consequents, while the second only considers rule antecedents. In both cases, the percentage of rules for each relevance quarter is calculated. Table 11 shows the rules of the initial models matched by quarters, according to their own relevance based on orthogonal transformations, and other views and results regarding the low & high relevance issue.

Table 11: Distribution (%) of rules classified by relevance quarters: OLS and P-QR, Accuracy-Interpretability View, initial models

Scatter Fuzzy Models (%)						
Orthogonal Transformation	**FRBS Model**	***Low Relevance***	***Medium-Low Relevance***	***Medium-High Relevance***	***High Relevance***
OLS	FasArt	85.70	8.49	2.81	2.99
	S-IRL	92.63	3.33	1.35	2.69
P-QR	FasArt	70.87	18.04	6.85	4.25
	S-IRL	72.71	19.39	4.75	3.15

Linguistic Fuzzy Models (%)					
Orthogonal Transformation	**FRBS Model**	***Low Relevance***	***Medium*-Low *Relevance***	***Medium-High Relevance***	***High Relevance***
OLS	NefProx	92.95	3.24	1.23	2.58
	L-IRL	94.41	2.74	0.89	1.97
P-QR	NefProx	85.11	8.98	3.14	2.77
	L-IRL	81.87	10.92	3.82	3.38

First of all, all these results show that most of the FRBS rules can be labeled as *Low Relevance* rules for both scatter models and linguistic ones. In this last case, the percentage of *Low Relevance* rules (around 88%) is even higher than the scatter case (around 80%). Thus, the idea of removing *Low Relevance* rules does not seem to be the best, and the idea of a "gap" as a criterion to define the border between rules to be preserved and rules to be removed, does not seem to be implementable. In general, to remove *Low Relevance* rules seems a highly questionable and debatable issue.

Average results for improved models using OLS are shown in tables 12 and 13. Tables 14 and 15 show the results for P-QR. Checking the lowest relevance rules of the improved models for the *Best Acc*, *Median Acc-Inter* and *Best Inter*: 30-50 % of their original *Low Relevance* rules have been preserved for most cases, which means from around 60% up to 90% of the rules of the improved models. In this context, these rules with *Low Relevance* become a serious factor, playing a relevant role in the Accuracy-Interpretability trade-off. On the other hand, rules with *High Relevance* are not always preserved on the way to obtaining this trade-off, a fact which is not discussed in the published works.

Table 12: Distribution (%) of rules classified by relevance quarters and interpretability indexes: Accuracy-Interpretability view, OLS, Rel_{RH}, improved scatter models. (a) Preserved rules (%) (b) Improved distribution (%)

FasArt

Best *Inter*

	Low Relevance				***Medium-Low Rel.***				***Medium-High Rel.***				***High Relevance***			
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	45.71	45.84	67.00	55.96	80.03	80.09	86.29	70.18	75.97	77.92	82.03	71.37	86.18	86.18	88.21	73.05
(b)	72.96	73.21	80.33	78.83	14.22	14.03	10.83	10.56	4.64	5.07	3.80	4.67	8.18	8.29	5.04	5.94

Median *Acc-Inter*

	Low Relevance				***Medium-Low Rel.***				***Medium-High Rel.***				***High Relevance***			
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	57.34	57.49	70.58	62.36	84.48	84.42	86.44	75.51	79.22	79.87	82.03	74.15	84.35	84.35	87.40	68.52
(b)	77.74	78.07	81.00	80.77	12.50	12.59	10.62	10.62	4.07	4.50	3.71	4.16	5.69	5.86	4.67	4.46

Best *Acc*

	Low Relevance				***Medium-Low Rel.***				***Medium-High Rel.***				***High Relevance***			
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	68.34	68.39	71.22	68.40	86.14	86.08	86.36	78.14	82.68	82.68	81.60	77.35	86.38	86.38	86.79	71.40
(b)	80.70	80.67	81.06	82.02	10.72	11.10	10.58	9.73	3.79	4.39	3.67	4.03	4.79	4.99	4.70	4.22

S-IRL

Best *Inter*

	Low Relevance				***Medium-Low Rel.***				***Medium-High Rel.***				***High Relevance***			
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	32.03	32.05	54.31	41.26	52.15	52.15	60.09	48.71	51.85	51.32	60.32	49.74	55.41	55.41	60.16	49.60
(b)	84.61	84.76	89.36	85.49	5.67	5.77	4.07	5.43	2.54	2.69	1.97	2.95	7.18	7.38	4.59	6.12

Median *Acc-Inter*

	Low Relevance				***Medium-Low Rel.***				***Medium-High Rel.***				***High Relevance***			
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	42.16	42.05	54.56	48.40	55.15	55.15	60.09	53.22	55.56	55.03	60.32	51.85	56.46	56.20	60.69	51.98
(b)	88.09	88.18	89.44	88.42	4.45	4.79	4.04	4.47	2.18	2.55	1.94	2.15	5.28	5.50	4.58	4.95

Best *Acc*

	Low Relevance				***Medium-Low Rel.***				***Medium-High Rel.***				***High Relevance***			
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	51.17	51.19	54.49	54.82	57.51	57.51	60.09	60.30	55.56	55.56	60.32	57.67	57.52	57.52	60.69	56.99
(b)	89.79	89.65	89.41	89.40	3.94	4.44	4.05	4.15	1.81	2.41	1.95	2.00	4.46	4.65	4.59	4.45

The situation with respect to these issues for linguistic and scatter FRBS models is discussed below.

- **Scatter Fuzzy Models**

  A first point about the rule selection carried out is the different selection (preservation) rates of rules between models: FasArt, a scatter approach, has a larger ratio of preserved rules ($50\% - 90\%$) than the rest of the approaches involved in this work: scatter or linguistic ones, and for all levels of rule relevance, as shown in tables 12(a) and 14(a).

  In general, this higher rate can mean fuzzy rules have a lower level of

Table 13: Distribution (%) of rules classified by relevance quarters and interpretability indexes: Accuracy-Interpretability view, OLS, Rel_{RH}, improved linguistic models. (a) Preserved rules (%) (b) Improved distribution (%)

NefProx

Best *Inter*

	Low Relevance				***Medium-Low Rel.***				***Medium-High Rel.***				***High Relevance***			
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	6.50	14.70	32.39	17.78	40.31	48.63	73.30	48.64	60.19	51.75	72.69	52.31	63.51	48.20	61.04	40.09
(b)	52.44	68.79	76.51	74.42	10.93	6.80	6.05	6.42	6.93	3.63	2.43	3.18	18.58	9.66	3.89	4.87

Median *Acc-Inter*

	Low Relevance				***Medium-Low Rel.***				***Medium-High Rel.***				***High Relevance***			
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	15.56	23.25	36.04	23.06	51.36	58.76	76.87	54.93	59.72	59.65	76.85	56.48	55.63	45.72	62.39	41.67
(b)	71.38	75.62	76.60	76.82	6.95	6.16	6.08	6.00	3.75	2.86	2.55	2.58	6.81	4.24	3.66	3.49

Best *Acc*

	Low Relevance				***Medium-Low Rel.***				***Medium-High Rel.***				***High Relevance***			
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	27.66	35.14	36.74	30.97	73.13	75.26	77.04	73.64	76.85	78.07	76.85	75.93	69.14	64.19	67.12	64.41
(b)	73.67	76.12	76.42	75.18	6.72	6.10	5.99	6.42	3.01	2.63	2.54	2.79	5.48	4.04	3.94	4.50

L-IRL

Best *Inter*

	Low Relevance				***Medium-Low Rel.***				***Medium-High Rel.***				***High Relevance***			
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	16.03	15.23	29.64	14.15	50.75	51.87	71.54	38.39	41.95	45.40	56.32	36.78	57.29	57.03	75.26	43.49
(b)	74.62	74.17	84.28	81.05	12.73	12.89	8.09	9.23	2.48	2.86	1.66	2.53	10.17	10.08	5.97	7.19

Median *Acc-Inter*

	Low Relevance				***Medium-Low Rel.***				***Medium-High Rel.***				***High Relevance***			
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	24.36	23.83	34.50	20.47	58.61	59.74	70.97	51.31	48.28	47.70	53.45	39.08	59.38	58.33	71.35	50.78
(b)	82.52	82.39	85.49	85.42	8.77	9.06	7.64	7.87	1.78	1.78	1.50	1.66	6.93	6.77	5.37	5.06

Best *Acc*

	Low Relevance				***Medium-Low Rel.***				***Medium-High Rel.***				***High Relevance***			
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	42.58	37.80	38.98	39.25	69.48	71.16	70.41	70.04	55.17	55.75	52.87	44.83	76.82	71.09	72.14	63.28
(b)	86.69	86.00	85.81	87.62	7.04	7.47	7.49	7.09	1.38	1.39	1.43	1.08	4.89	5.14	5.27	4.22

incoherence, a similar *number of membership functions*, and a higher contribution of rules to the model Accuracy. Another point is that there are rules with a *High Relevance* that are not considered for the improved models: around $15-30\%$ for the FasArt case, and $50-55\%$ for the S-IRL case. All these figures are quite similar for OLS, P-QR and SVD.

Tables 12(b), 13(b), 14(b) and 15(b) show that the improved models are made for a “huge” % of rules with *Low Relevance*, in comparison with the rest of the model rules, for every interpretability index and Best *Acc*, Median *Acc-Inter* or Best *Inter* cases. From around 50% up

Table 14: Distribution (%) of rules classified by relevance quarters and interpretability indexes: Accuracy-Interpretability view, P-QR, Rel_{RH}, improved scatter models. (a) Preserved rules (%) (b) Improved distribution (%)

FasArt

Best *Inter*

	Low Relevance				***Medium-Low Rel.***				***Medium-High Rel.***				***High Relevance***			
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	44.60	44.60	65.71	58.69	62.37	62.37	81.20	56.29	77.54	77.54	87.08	63.58	85.63	85.63	89.08	65.68
(b)	44.19	44.28	55.62	54.72	26.01	26.13	24.02	22.46	16.34	16.55	12.17	13.46	13.45	13.63	8.18	9.35

Median *Acc-Inter*

	Low Relevance				***Medium-Low Rel.***				***Medium-High Rel.***				***High Relevance***			
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	55.74	55.74	68.49	63.79	71.91	71.91	82.45	63.69	83.69	83.69	89.04	69.40	86.64	86.64	89.80	77.12
(b)	51.47	51.59	56.58	55.89	24.78	25.08	23.56	22.08	13.81	14.21	11.93	12.51	9.94	10.14	7.92	9.53

Best *Acc*

	Low Relevance				***Medium-Low Rel.***				***Medium-High Rel.***				***High Relevance***			
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	66.42	66.42	69.41	68.95	79.41	79.41	82.52	74.65	87.70	87.70	88.77	77.34	87.79	87.79	89.22	81.78
(b)	56.38	56.23	56.67	57.42	23.54	24.04	23.44	22.76	12.07	12.67	11.92	11.81	8.01	8.21	7.97	8.01

S-IRL

Best *Inter*

	Low Relevance				***Medium-Low Rel.***				***Medium-High Rel.***				***High Relevance***			
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	26.20	26.16	55.93	47.07	48.47	48.51	63.40	48.14	58.92	58.77	66.57	50.52	51.47	51.92	53.95	45.15
(b)	54.79	54.77	66.75	65.97	27.74	27.95	20.99	20.30	9.89	9.98	7.22	7.30	7.58	7.89	5.04	6.44

Median *Acc-Inter*

	Low Relevance				***Medium-Low Rel.***				***Medium-High Rel.***				***High Relevance***			
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	37.67	37.63	56.32	52.54	54.36	54.28	63.76	54.06	61.17	61.02	66.57	56.97	50.11	50.11	54.18	49.44
(b)	62.41	62.56	66.87	67.01	23.84	24.15	21.05	20.23	8.19	8.54	7.10	7.24	5.57	5.78	4.98	5.52

Best *Acc*

	Low Relevance				***Medium-Low Rel.***				***Medium-High Rel.***				***High Relevance***			
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	48.21	48.21	56.18	58.29	58.69	58.69	63.62	62.44	62.82	62.82	66.57	63.72	51.02	51.02	53.95	51.47
(b)	66.69	66.54	66.90	67.93	21.62	22.12	21.00	20.30	6.93	7.53	7.13	6.92	4.77	4.97	4.98	4.84

to 90% of the rules of these models are *Low Relevance*. This is a little higher for the S-IRL case. So, a very important number of these rules are preserved for the improved models.

In general, rules with a *High Relevance* are more significant for the Best *Inter* model, while rules with a *Low Relevance* are more significant for the Best *Acc* model. There are no significant differences for the interpretability index considered in each case.

- **Linguistic Fuzzy Models**

  In comparison with scatter models, the linguistic ones have a lower

Table 15: Distribution (%) of rules classified by relevance quarters and interpretability indexes: Accuracy-Interpretability view, P-QR, Rel_{RH}, improved linguistic models. (a) Preserved rules (%) (b) Improved distribution (%)

NefProx

	Low Relevance				***Medium-Low Rel.***				***Medium-High Rel.***				***High Relevance***			
Best *Inter*																
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	5.50	13.09	31.47	16.24	15.42	27.06	49.52	32.78	38.01	50.37	68.37	55.19	61.25	49.58	60.16	48.75
(b)	48.89	62.18	65.75	63.05	11.09	11.34	11.68	12.02	10.57	7.00	6.04	7.40	18.33	8.37	5.42	6.42
Median *Acc-Inter*																
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	14.66	21.56	34.72	21.64	27.50	34.80	50.07	37.36	48.31	55.13	73.30	60.00	44.17	50.83	61.59	51.67
(b)	61.84	67.39	66.67	66.29	12.44	10.26	11.01	10.52	7.74	5.98	5.83	6.77	6.87	5.26	5.38	5.31
Best *Acc*																
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	26.32	34.27	37.76	30.87	40.14	49.09	52.75	46.18	65.92	66.48	74.81	70.74	57.50	61.46	65.04	62.71
(b)	62.43	67.39	66.83	66.24	12.32	10.67	11.02	10.91	7.55	5.70	5.55	6.38	6.60	5.13	5.50	5.36

L-IRL

	Low Relevance				***Medium-Low Rel.***				***Medium-High Rel.***				***High Relevance***			
Best *Inter*																
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	10.65	10.47	25.96	12.37	25.59	27.18	55.82	31.55	40.46	38.84	64.38	23.52	55.15	56.97	64.09	46.36
(b)	50.90	49.35	61.56	57.48	21.00	22.52	20.35	22.35	12.71	12.27	9.37	7.48	15.40	15.87	8.73	12.69
Median *Acc-Inter*																
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	19.07	19.11	31.10	18.10	40.00	39.58	59.81	39.81	54.84	54.03	64.92	35.35	63.94	62.58	66.36	53.64
(b)	57.99	57.79	63.77	62.57	20.22	20.44	19.47	19.94	10.33	10.51	8.69	7.73	11.47	11.27	8.07	9.77
Best *Acc*																
	NR	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*	*NR*	*NMF*	*Inc*	*Sim*
(a)	39.57	34.53	36.81	35.96	59.20	56.57	60.09	57.42	71.91	65.99	64.11	60.75	75.45	75.61	71.36	71.67
(b)	65.38	63.90	64.99	66.03	18.21	18.38	18.51	17.60	8.33	8.62	8.32	7.68	8.08	9.09	8.18	8.69

ratio of selection, or preservation, of original rules. Under 40% of *Low Relevance* rules are selected for both linguistic FRBS models (see table 13 (a)). This % implies around $60-90\%$ of *Low Relevance* rules in the improved models.

On the other hand, the selection rate for *High Relevance* rules is very high ($> 70\%$) in comparison with *Low Relevance*. Here also, there are *High Relevance* rules that are not preserved: $25\%-60\%$.

As in the scatter case, the improved models are made for *Low Relevance* rules in a very high %: from around 60% up to 90% (see tables 13(b) and 15(b)). No differences exist between the interpretability indexes used.

Table 16: Distribution (%) of rules according to relevance quarters by other approaches

Orthogonal Transformation	FRBS Model	***Low Relevance***	***Medium-Low Relevance***	***Medium-High Relevance***	***High Relevance***
		Best *Inter*			
OLS	ScatA	85.51	9.39	2.41	2.69
	ScatB	89.67	6.87	1.72	1.75
	LingA	77.63	11.72	5.19	5.47
	LingB	80.67	10.02	5.43	3.88
P-QR	ScatA	69.84	18.75	8.24	3.18
	ScatB	64.78	25.05	6.74	3.43
	LingA	71.23	12.27	11.18	5.32
	LingB	71.96	14.16	8.33	5.54
		Median *Acc-Inter*			
Orthogonal Transformation	**FRBS Model**	***Low Relevance***	***Medium-Low Relevance***	***Medium-High Relevance***	***High Relevance***
OLS	ScatA	85.80	9.15	2.52	2.53
	ScatB	90.96	6.08	1.52	1.44
	LingA	81.74	9.74	4.03	4.49
	LingB	83.22	9.19	4.56	3.04
P-QR	ScatA	69.48	19.20	8.02	3.30
	ScatB	64.01	25.64	6.98	3.36
	LingA	73.93	12.02	7.72	6.33
	LingB	73.58	14.21	9.02	3.19
		Best *Acc*			
Orthogonal Transformation	**FRBS Model**	***Low Relevance***	***Medium-Low Relevance***	***Medium-High Relevance***	***High Relevance***
OLS	ScatA	86.38	8.91	2.36	2.34
	ScatB	91.55	5.76	1.33	1.36
	LingA	84.93	7.56	3.90	3.61
	LingB	85.98	7.98	3.81	2.22
	WM-R	86.73	6.66	3.42	3.20
	GFS-RS-T	79.42	9.48	4.71	6.39
	S_{SP2}	73.27	14.01	6.30	6.42
P-QR	ScatA	70.22	18.07	8.53	3.18
	ScatB	64.30	25.47	6.94	3.29
	LingA	77.50	10.68	7.23	4.59
	LingB	75.99	12.39	6.62	5.01
	WM-R	78.82	12.06	3.99	5.13
	GFS-RS-T	73.69	13.70	5.45	7.15
	S_{SP2}	65.31	15.75	8.02	10.92

These "surprising" results concerning rules with *High* and *Low Relevance* are supported by the Relevance metric, which is defined locally, meaning for each rule individually. However, accuracy is globally defined for the FRBS model taking into account all the components of FRBS (knowledge base, inference system, etc.), not only the fuzzy rule set.

The results show that rules with high relevance can be removed and/or rules with low relevance preserved. So, the idea of removing the latter does not match with the results obtained, nor with always preserving the most relevant rules. This seems contradictory, but it fits with the nature of both types of rule.

In general, rules with low relevance, according to relevance based on orthogonal transformations, can concern rules with very low levels of activation, except for some cases of the input space, or very similar activation levels for most input/output spaces; whereas that rules with high relevance correspond to very variable levels of activation for most cases of the input/output space. The first case can imply "exception" cases in the input space, or rules covering the input space very homogeneously; while the second can imply rules covering most of the input/output universe on different levels.

In order to reach the Accuracy-Interpretability trade-off, it may be necessary to preserve or remove some rules, those that allow the trade-off to be reached by considering the structure and operators of each FRBS and data. Here, accuracy as a global index plays a very powerful role in comparison with the interpretability or relevance of the rules, which are only based on the fuzzy rule set of the FRBS. In other words, this implies there is interaction between rules that depends on the nature and design of each FRBS, which can modify the perception based on the rule relevance.

In this way, some low relevant rules are preserved due to their considerable impact on the accuracy of the FRBS.

On the other hand, removing some relevant rules can be explained in terms of Interpretability: these rules are very relevant, so here this implies much variability of activation regarding most input/output data spaces. This fact can imply redundancy, or even incoherency, between rules, so interpretability works like a global metric and could lead to them being removed if the impact on accuracy were assumable.

In terms of rule Relevance, this is a local measurement for every rule concerning its variability based meaningfulness for each FRBS and problem. However, it does not imply how relevant or useful each rule, or the whole rule set of an FRBS, is when dealing with accuracy, or another objective, as a global issue of an FRBS. It is only possible to know the local relevance of each rule, or the average relevance of the rules contained in the rule set, but not how they interact for each type of FRBS, operators, etc. So a complementary global measure is interesting.

On the other hand, the FRBS performance is a global measurement of the effective relevance of the whole rule set of a particular FRBS, which can be a complementary index to the local relevance of the rules, so as to gain a global view of the relevance of a fuzzy rule set.

As in the previous section, a comparison in terms of relevance with models obtained by other approaches can be seen in Table 16: showing the distribution (%) of selected rules classified by relevance quarters for these cases. Analyzing the rule relevance based distribution for the Best *Acc* models by OLS, the percentage of rules with *Low Relevance* selected by this relevance based approach is lower than the other reported models which are based on $Inter = NR$. These differences are even higher when Median *Acc*-$Inter$ and Best $Inter$ models are analyzed, and even higher still with P-QR. As for *Low Relevance* selected rules, once more, this relevance based approach has selected a higher number of these meaningful rules for all cases.

In fact, using rule relevance, a good Accuracy-Interpretability trade-off allows more relevant rules to be selected than other approaches obtaining a similar, or even better, FRBS performance.

7. Lessons learned

This exhaustive experimental work and analyses concerning the Accuracy-Interpretability-Relevance trade-off has permitted us to show rule relevance as a serious factor in selecting the most adequate rules for improving the Accuracy-Interpretability trade-off, for both scatter and linguistic FRBSs. It even seems possible to improve both the accuracy and interpretability of the FRBSs simultaneously. This performance has been checked regarding other reported approaches, with positive statistical results.

The results show that FRBSs have been improved in over 98% of cases, increasing relevance and decreasing *error*, *number of rules* and *membership functions*, *incoherence* and *similarity*.

In fact, Accuracy is improved by between 48.46% - 62.18% for scatter models and 43.79% - 65.33% for linguistic ones. Interpretability improves by between 23.31% - 52.81% for scatter and 43.51% - 73.44% for linguistic ones, and Relevance by between 8.41% - 47.36% for scatter FRBSs and 40.50% - 190.24% for linguistic ones. No serious differences were found between the relevance criteria proposed. Regarding the relevance based on orthogonal transformations, in general, OLS achieves the best improvements for both scatter and linguistic models. P-QR gets the best ones for Accuracy and SVD for Interpretability in scatter models.

Furthermore, the idea of removing *Low Relevance* rules and preserving *High Relevance* rules from FRBSs has been shown to be a debatable issue. *Low Relevance* rules have important contributions in both scatter and linguistic FRBSs, but their influence in scatter FRBSs is higher. Around 6%-50% of the *Low Relevance* rules are preserved and about 20%-50% of the *High Relevance* rules must be dropped to reach this same trade-off.

This fact is linked with the global or local nature of the metrics considered and how many times a rule is fired, its level of activation, its similarity regarding other rules and its contribution to the FRBS output. This implies a blend between these metrics must be addressed for a more effective and coherent relevance rule.

8. Concluding remarks

The target of this work is to check the concept relevance of the FRBS rule in order to get a better Accuracy-Interpretability trade-off, using a MOEA based rule selection. In this sense, exhaustive experimental work and analyses have been carried out.

This proposal has shown rule relevance to be a serious factor for the Accuracy-Interpretability trade-off, for both scatter and linguistic FRBSs, selecting the most adequate rules for a better Accuracy-Interpretability trade-off. A comparison regarding other reported approaches has shown the performance of rule relevance for this goal.

Moreover, the role of rules based on their relevance has been seen to be contradictory and debatable: some low relevant rules have a high impact for the FRBS trade-off and some high relevant rules are not adequate for this trade-off.

This fact suggests that the usual definition of rule Relevance, as a local measurement, should be complemented by another global index that takes into account rule interactions and the nature of the FRBS involved. Accuracy by Error can be a complementary measurement for this rule Relevance, showing the global Relevance of the FRBS model, including its fuzzy rule set. In some terms, Interpretability can take on this role to some extent, depending on its definition of Interpretability regarding the nature and components of the FRBS.

The work in progress is based on the characterization of the rules to be preserved and removed, as well as the study of formulations of relevance rules that permit a better coherency regarding rules with *High* and *Low Relevance* for a correct balance between the Accuracy and Interpretability of the FRBS.

References

Alcalá, R., Gacto, M. J., Herrera, F., Alcalá-Fdez, J., 2007. A multi-objective genetic algorithm for tuning and rule selection to obtain accurate and compact linguistic fuzzy rule-based systems. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 15 (5), 539 – 557.

Alcalá, R., Nojima, Y., Herrera, F., Ishibuchi, H., December 2011. Multiobjective genetic fuzzy rule selection of single granularity-based fuzzy classification rules and its interaction with the lateral tuning of membership functions. Soft Computing 15 (12), 2303 – 2318.

Alcalá-Fdez, J., Alonso, J., 2016. A survey of fuzzy systems software: taxonomy, current research trends and prospects. IEEE Transactions on Fuzzy Systems 24 (1), 40 – 56.

Alcalá-Fdez, J., Sánchez, L., García, S., del Jesus, M. J., Ventura, S., Garrell, J. M., Otero, J., Romero, C., Bacardit, J., Rivas, V. M., Fernández, J. C., Herrera, F., Febraury 2009. KEEL: a software tool to assess evolutionary algorithms for data mining problems. Soft Computing 13 (3), 307 – 318.

Alonso, J. M., Castiello, C., Mencar, C., 2015. Interpretability of fuzzy systems: Current research trends and prospects. In: Springer Handbook of Computational Intelligence. Springer, pp. 219–237.

Alonso, J. M., Magdalena, L., 2011. HILK++: an interpretability-guided fuzzy modeling methodology for learning readable and comprehensible fuzzy rule-based classifiers. Soft Computing 15, 1959 – 1980.

Alonso, J. M., Magdalena, L., González-Rodríguez, G., December 2009. Looking for a good fuzzy system interpretability index: An experimental approach. International Journal of Approximate Reasoning 51 (1), 115 – 134.

Antonelli, M., Ducange, P., Lazzerini, B., Marcelloni, F., December 2011. Learning knowledge bases of multi-objective evolutionary fuzzy systems by simultaneously optimizing accuracy, complexity and partition integrity. Soft Computing 15 (12), 2335 – 2354.

Cano Izquierdo, J. M., Dimitriadis, Y. A., Gómez Sánchez, E., López Coronado, J., May 2001. Learning from noisy information in FasArt and Fasback neuro-fuzzy systems. Neural Networks 14 (4-5), 407–425.

Casillas, J., Cordón, O., Herrera, F., Magdalena, L., 2003a. Accuracy improvements to find the balance interpretability-accuracy in fuzzy modeling: An overview. In: Casillas, J., Cordón, O., Herrera, F., Magdalena, L. (Eds.), Accuracy Improvements in Linguistic Fuzzy Modelling. Vol. 129 of Studies in Fuzziness and SoftComputing. Springer-Verlag, Berlin Heildelberg, pp. 3–24.

Casillas, J., Cordón, O., Herrera, F., Magdalena, L., 2003b. Interpretability improvements to find the balance interpretability-accuracy in fuzzy modeling: An overview. In: Casillas, J., Cordón, O., Herrera, F., Magdalena, L. (Eds.), Interpretability Issues in Fuzzy Modelling. Vol. 128 of Studies in Fuzziness and SoftComputing. Springer-Verlag, Berlin Heildelberg, pp. 3–22.

Casillas, J., Martínez, P., Benítez, A. D., 2009. Learning consistent, complete and compact sets of fuzzy rules in conjunctive normal form for regression problems. Soft Computing 13, 451 – 465.

Cordón, O., 2011. A historical review of evolutionary learning methods for mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems. International Journal of Approximate Reasoning 52, 894 – 913.

Cordón, O., Herrera, F., 1997. A three-stage evolutionary process for learning descriptive and approximate fuzzy logic controller knowledge bases from examples. International Journal of Approximate Reasoning 17 (4), 369 – 407.

Cordón, O., Herrera, F., 2001. Hybridizing genetic algorithms with sharing scheme and evolution strategies for designing approximate fuzzy rule-based systems. Fuzzy Sets and Systems 118, 235 – 255.

Destercke, S., Guillaume, S., Charnomordic, B., September 2007. Building an interpretable fuzzy rule base from data using orthogonal least squares - Application to a depollution problem. Fuzzy Sets and Systems 158 (18), 2078 – 2094.

Devroye, L., Györfi, L., Lugosi, G., 2013. A probabilistic theory of pattern recognition. Vol. 31. Springer Science & Business Media.

Eshelman, L. J., 1991. The CHC adaptive search algorithm : How to have safe search when engaging in nontraditional genetic recombination. Foundations of Genetic Algorithms 1, 265–283.

Fazzolari, M., Alcalá, R., Nojima, Y., Ishibuchi, H., Herrera, F., 2013a. A review of the application of multiobjective evolutionary fuzzy systems: Current status and further directions. IEEE Transactions on Fuzzy Systems 21, 45 – 65.

Fazzolari, M., Giglio, B., Alcalá, R., Marcelloni, F., Herrera, F., 2013b. A study on the application of instance selection techniques in genetic fuzzy rule-based classification systems: Accuracy-complexity trade-off. Knowledge-Based Systems 54, 32 – 41.

Fernández, A., López, V., del Jesus, M. J., Herrera, F., 2015. Revisiting evolutionary fuzzy systems: Taxonomy, applications, new trends and challenges. Knowledge-Based Systems 80, 109–121.

Gacto, M. J., Alcalá, R., Herrera, F., 2010. Integration of an index to preserve the semantic interpretability in the multi-objective evolutionary rule selection and tuning of linguistic fuzzy systems. IEEE Transactions on Fuzzy Systems 18 (3), 515 – 531.

Gacto, M. J., Alcalá, R., Herrera, F., 2011. Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures. Information Sciences 181, 4340 – 4360.

Galende, M., Gacto, M. J., Sainz, G., Alcalá, R., 2014. Comparison and design of interpretable linguistic vs. scatter FRBSs: Gm3m generalization and new rule meaning index for global assessment and local pseudo-linguistic representation. Information Sciences 282, 190–213.

Galende, M., Sainz, G. I., Fuente, M. J., March 2012. Complexity reduction and interpretability improvement for fuzzy rule systems based on simple interpretability measures and indices by bi-objective evolutionary rule selection. Soft Computing 16 (3), 451 – 470.

Golub, G. H., Van Loan, C. F., 2012. Matrix computations. Vol. 3. JHU Press.

Guillaume, S., Charnomordic, B., 2003. A new method for inducing a set of interpretable fuzzy partitions and fuzzy inference systems from data. In: Casillas, J., Cordón, O., Herrera, F., Magdalena, L. (Eds.), Interpretability Issues in Fuzzy Modelling. Vol. 128 of Studies in Fuzziness and SoftComputing. Springer-Verlag, Berlin Heildelberg, pp. 148–175.

Herrera, F., 2008. Genetic fuzzy systems: Taxonomy, current research trends and prospects. Evolutionary Intelligence 1, 27 – 46.

Hjørland, B., 2010. The foundation of the concept of relevance. Journal of the American Society for Information Science and Technology 61 (2), 217–237.

Horn, L., Ward, G., 2008. The handbook of pragmatics. Vol. 26. John Wiley & Sons.

Ishibuchi, H., Murata, T., Türksen, I. B., July 1997. Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems. Fuzzy Sets and Systems 89 (2), 135 – 150.

Ishibuchi, H., Nakashima, T., Murata, T., August 2001. Three-objective genetics-based machine learning for linguistic rule extraction. Information Sciences 136 (1-4), 109 – 133.

Ishibuchi, H., Nojima, Y., January 2007. Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning. International Journal of Approximate Reasoning. Special Issue on Genetic Fuzzy Systems and the Interpretability-Accuracy Trade-off 44 (1), 4 – 31.

Ishibuchi, H., Nojima, Y., 2013. Repeated double cross-validation for choosing a single solution in evolutionary multi-objective fuzzy classifier design. Knowledge-Based Systems 54, 22 – 31.

Ishibuchi, H., Nojima, Y., 2015. Multiobjective Genetic-Fuzzy Systems. In: Springer Handbook of Computational Intelligence. Springer, pp. 1479 – 1498.

Ishibuchi, H., Nozaki, K., Yamamoto, N., Tanaka, H., August 1995. Selecting fuzzy if-then rules for classification problems using genetic algorithms. IEEE Transactions on Fuzzy Systems 3 (3), 260 – 270.

Kacprzyk, J., Pedrycz, W., 2015. Springer Handbook of Computational Intelligence. Springer.

Karray, F. O., De Silva, C. d., 2004. Soft Computing and Intelligent Systems Design. Theory, Tools and Applications. Addison Wesley.

Keynes, J. M., 2013. A treatise on probability. Courier Corporation.

Kim, J., Suga, Y., Won, S., 2006. A new approach to fuzzy modeling of nonlinear dynamic systems with noise: relevance vector learning mechanism. Fuzzy Systems, IEEE Transactions on 14 (2), 222–231.

Konar, A., 2005. Computational Intelligence: Principles, techniques and applications. Springer-Verlag, Berlin.

Krone, A., Taeger, H., 2001. Data-based fuzzy rule test for fuzzy modelling. Fuzzy Sets and Systems 123 (3), 343–358.

Kundu, M. K., Chowdhury, M., Bulò, S. R., 2015. A graph-based relevance feedback mechanism in content-based image retrieval. Knowledge-Based Systems 73, 254–264.

Kwan, P. W., Welch, M. C., Foley, J. J., 2015. A knowledge-based decision support system for adaptive fingerprint identification that uses relevance feedback. Knowledge-Based Systems 73, 236–253.

Liu, H., Motoda, H., 2012. Feature selection for knowledge discovery and data mining. Vol. 454. Springer Science & Business Media.

Magdalena, L., 2015. Fuzzy Rule-Based Systems. In: Springer Handbook of Computational Intelligence. Springer, pp. 203 – 218.

Márquez, A. A., Márquez, F. A., Peregrín, A., April 2012. A mechanism to improve the interpretability of linguistic fuzzy systems with adaptive defuzzification based on the use of a multi-objective evolutionary algorithm. International Journal of Computational Intelligence Systems 5 (2), 297 – 321.

Mencar, C., Castiello, C., Cannone, R., Fanelli, A., 2011. Interpretability assessment of fuzzy knowledge bases: A cointension based approach. International Journal of Approximate Reasoning 52, 501 – 518.

Mencar, C., Fanelli, A., December 2008. Interpretability constraints for fuzzy information granulation. Information Sciences 178 (24), 4585 – 4618.

Mikut, R., Jäkel, J., Gröll, L., March 2005. Interpretability issues in data-based learning of fuzzy systems. Fuzzy Sets and Systems 150 (2), 179 – 197.

Nauck, D., Kruse, R., January 1999. Neuro-fuzzy systems for function approximation. Fuzzy Sets and Systems 101 (2), 261–271.

Nguyena, C. H., Hoangc, V. T., Nguyenc, V. L., 2015. A discussion on interpretability of linguistic rule based systems and its application to solve regression problems. Knowledge-Based Systems 88, 107 – 133.

Okabe, M., Yamada, S., 2005. Learning filtering rulesets for ranking refinement in relevance feedback. Knowledge-Based Systems 18 (2), 117–124.

Oliveira, J. d., January 1999. Semantic constraints for membership function optimization. IEEE Transactions on Systems, Man and Cybernetics. Part A: Systems and Humans 29 (1), 128 – 138.

Pancho, D. P., Alonso, J. M., Cordón, O., Quirin, A., Magdalena, L., 2013. FINGRAMS: Visual representations of fuzzy rule-based inference for expert analysis of comprehensibility. IEEE Transactions on Fuzzy Systems 21 (6), 1133 – 1149.

Pedrycz, W., 2003. Expressing relevance interpretability and accuracy of rule-based systems. In: Casillas, J., Cordón, O., Herrera, F., Magdalena, L. (Eds.), Interpretability Issues in Fuzzy Modelling. Vol. 128 of Studies in Fuzziness and SoftComputing. Springer-Verlag, Berlin Heildelberg, pp. 547–567.

Pulkkinen, P., Hytönen, J., Koivisto, H., 2008. Developing a bioaerosol detector using hybrid genetic fuzzy systems. Engineering Applications of Artificial Intelligence 21 (8), 1330–1346.

Rey, M. I., Galende, M., Fuente, M., Sainz, G. I., October 2012. Checking orthogonal transformations and genetic algorithms for selection of fuzzy rules based on interpretability-accuracy concepts. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 20 (Suppl. 2), 159 – 186.

Roubos, H., Setnes, M., August 2001. Compact and transparent fuzzy models and classfiers through iterative complexity reduction. IEEE Transactions on Fuzzy Systems 9 (4), 516–524.

Salgado, P., 2008. Rule generation for hierarchical collaborative fuzzy system. Applied Mathematical Modelling 32 (7), 1159–1178.

Setnes, M., 2003. Simplification and reduction of fuzzy rules. In: Casillas, J., Cordón, O., Herrera, F., Magdalena, L. (Eds.), Interpretability Issues in Fuzzy Modelling. Vol. 128 of Studies in Fuzziness and SoftComputing. Springer-Verlag, Berlin Heildelberg, pp. 278–302.

Setnes, M., Babuška, R., Kaymak, U., van Nauta Lemke, H., June 1998. Similarity measures in fuzzy rule base simplification. IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics 28 (3), 376 – 386.

Wang, L.-X., Mendel, J., November/December 1992. Generating fuzzy rules by learning from examples. IEEE Transactions on Systems, Man and Cybernetics 22 (6), 1414–1427.

Yen, J., Wang, L., February 1999. Simplifying fuzzy rule-based models using orthogonal transformation methods. IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics 29 (1), 13–24.

Yu, L., Liu, H., 2004. Efficient feature selection via analysis of relevance and redundancy. Journal of Machine Learning Research 5, 1205–1224.

Zhou, S.-M., Gan, J. Q., June 2007. Constructing L2-SVM-based fuzzy classifiers in high-dimensional space with automatic model selection and fuzzy rule ranking. IEEE Transactions on Fuzzy Systems 15 (3), 398 – 409.

Zhou, S.-M., Gan, J. Q., 2008. Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modelling. Fuzzy Sets and Systems 159, 3091 – 3131.

Zhou, S.-M., Garibaldi, J. M., John, R. I., Chiclana, F., June 2009. On constructing parsimonious type-2 fuzzy logic systems via influential rule selection. IEEE Transactions on Fuzzy Systems 17 (3), 654 – 667.

Zitzler, E., Laumanns, M., Thiele, L., 2001. SPEA2: Improving the strength pareto evolutionary algorithm for multiobjective optimization. In: Proc. Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems. Barcelona, Spain, pp. 95–100.