Information Sciences 181 (2011) 4340–4360

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures [☆]

M.J. Gacto [a,*], R. Alcalá [b], F. Herrera [b]

[a] *Department of Computer Science, University of Jaen, 23071 Jaen, Spain*
[b] *Department of Computer Science and Artificial Intelligence, University of Granada, 18071 Granada, Spain*

ARTICLE INFO

Article history:
Available online 4 March 2011

Keywords:
Complexity
Semantic interpretability
Linguistic fuzzy rule-based systems

ABSTRACT

Linguistic fuzzy modelling, developed by linguistic fuzzy rule-based systems, allows us to deal with the modelling of systems by building a linguistic model which could become interpretable by human beings. Linguistic fuzzy modelling comes with two contradictory requirements: interpretability and accuracy. In recent years the interest of researchers in obtaining more interpretable linguistic fuzzy models has grown.

Whereas the measures of accuracy are straightforward and well-known, interpretability measures are difficult to define since interpretability depends on several factors; mainly the model structure, the number of rules, the number of features, the number of linguistic terms, the shape of the fuzzy sets, etc. Moreover, due to the subjectivity of the concept the choice of appropriate interpretability measures is still an open problem.

In this paper, we present an overview of the proposed interpretability measures and techniques for obtaining more interpretable linguistic fuzzy rule-based systems. To this end, we will propose a taxonomy based on a double axis: "Complexity versus semantic interpretability" considering the two main kinds of measures; and "rule base versus fuzzy partitions" considering the different components of the knowledge base to which both kinds of measures can be applied. The main aim is to provide a well established framework in order to facilitate a better understanding of the topic and well founded future works.

1. Introduction

Fuzzy modelling (FM), system modelling by Fuzzy rule-based systems (FRBSs), may be considered as an approach used to model a system, making use of a descriptive language based on fuzzy logic with fuzzy predicates. When dealing with the interpretability-accuracy trade-off of the obtained models the FM field may be divided into two different areas depending on which is the main requirement pursued:

1. Linguistic fuzzy modelling (LFM): The main objective is to obtain fuzzy models with good interpretability and it is mainly developed by means of linguistic (or classic Mamdani) FRBSs [49,50]. Linguistic FRBSs are based on linguistic rules, in which the antecedent and the consequent make use of linguistic variables comprised of linguistic terms and the associated fuzzy sets defining their meanings.
2. Precise fuzzy modelling (PFM): The main objective is to obtain fuzzy models with good accuracy, and it is mainly developed by means of Takagi–Sugeno FRBSs [72] or by means of approximate FRBSs, which differ from the linguistic ones in the use of fuzzy variables, i.e., fuzzy sets without an associated meaning.

[☆] Supported by the Spanish Ministry of Education and Science under Grant No. TIN2008-06681-C06-01.

[*] Corresponding author.
E-mail addresses: mjgacto@ugr.es (M.J. Gacto), alcala@decsai.ugr.es (R. Alcalá), herrera@decsai.ugr.es (F. Herrera).

0020-0255/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2011.02.021

In this work, we are going to focus on LFM since this approach allows us to deal with the modelling of systems by building a linguistic model which is naturally closer to interpretability than PFM for human beings. Focusing on LFM, we have to pay attention to two contradictory requirements of the model:

- *Accuracy*: This is the capability to faithfully represent the real system. It should be better as there is a higher similarity between the responses of the real system and the fuzzy model. There are well-defined measures that are widely accepted in order to assess how good the accuracy is. Some well known examples for both, classification and regression, are the percentage of correctly classified patterns from a set of example data (complementarily percentage of misclassified patterns), or the mean square error (MSE) for regression problems as a way to measure how fair the model is with respect to a set of example data, among others.
- *Interpretability*: This is the capacity to express the behavior of the real system in an understandable way. It is a subjective property that depends on the person who makes the assessment. This is related to several factors, mainly the model structure, the number of input variables, the number of fuzzy rules, the number of linguistic terms, the shape of the fuzzy sets, etc. There is still no standard measure to assess how good interpretability is.

Researchers have usually focused on the improvement of the accuracy of the models obtained without paying special attention to the interpretability. Nowadays, the interest of the researchers in interpretability has grown, which has prompted the appearance of a great quantity of work with the purpose of obtaining more interpretable linguistic models. However, two important problems remain to be solved:

- Accuracy and interpretability represent contradictory objectives. The ideal thing would be to satisfy both criteria to a high degree, but since they are contradictory properties it is generally not possible. Because of this, researchers usually focus on obtaining the best trade-off between interpretability and accuracy [12,13], depending on the user's requirements.
- Due to its subjective nature and the large amount of factors involved, the choice of an appropriate interpretability measure is still an open problem. Most researchers would agree on interpretability involving aspects such as: The number of rules being as small as possible; rule premises that are easily understood in terms of structure and contain only a few input variables, linguistic terms that are intuitively comprehensible; etc. Nevertheless, whereas the definition of accuracy in a certain application is straightforward, the definition of interpretability is rather problematic.

In this paper, we present an overview of the proposed interpretability measures and techniques for obtaining more interpretable linguistic FRBSs. To this end, we will propose a taxonomy based on a double axis: "complexity versus semantic interpretability" considering the two main kinds of measures; and "rule base versus fuzzy partitions" considering the different components of the knowledge base (KB) to which both kinds of measures can be applied. This leads to four different quadrants being analyzed: The complexity at the rule base (RB) level; the complexity at the fuzzy partition level; the semantics at the RB level; the semantics at the fuzzy partition level. The main aim is to provide a well established framework in order to facilitate a better understanding of the topic and well founded future works.

We consider that a revision of the existing methodologies and measures, taking into account the interpretability of linguistic FRBSs as a part of the interpretability-accuracy trade-off problem together with a taxonomy, would be very interesting in order to show how the different authors tackled this difficult problem. Although, there are two review papers by Zhou and Gan [75] and by Mencar et al. [54] studying the interpretability of fuzzy systems in general, there is no exhaustive review of interpretability issues that specifically focuses on the area of linguistic FRBSs, which represents an extensive framework deserving a deeper analysis from the mentioned double point of view, complexity versus semantic interpretability.

This paper is arranged as follows. The next section presents the taxonomy proposed for studying interpretability in the linguistic FRBSs area. Section 3 presents the works considering complexity at the RB level. This is one of the quadrants of the taxonomy representing the most extended way to work, i.e., the classic measures to obtain simpler models. In Section 4, we present the measures that the authors proposed to improve the interpretability by taking into account the complexity at the fuzzy partition level. In Section 5, we analyze those works devoted to maintaining semantic interpretability at the RB level. Section 6 includes those works trying to ensure the semantic interpretability at the fuzzy partition level which usually imposes constraints on the MFs by considering measures such as distinguishability, coverage, etc. Section 7 summarizes the current state-of-the-art to assess the interpretability of linguistic FRBSs stressing some important hints on the different quadrants. Finally, in Section 8 we draw some conclusions.

2. Semantic interpretability versus complexity: A taxonomy on linguistic FRBSs

In this section, we propose a specific taxonomy that can help us better understand how the interpretability aspect has been taken into account in the particular framework of linguistic FRBSs. Different works [10,54,75] have proposed interesting taxonomies as a way to study interpretability aspects within the more general area of fuzzy systems:

- Alonso et al. in [10] present a taxonomy including the main factors to be considered in order to assess the interpretability of FRBSs (for both LFM and PFM). Those factors are description and explanation: "On the one hand, the system is viewed as a whole describing its global behavior and trend. On the other hand, each individual situation is analyzed explaining specific behaviors for specific events".

- A taxonomy on interpretability constraints for information granules has been suggested by Mencar et al. [54], considering: "Constraints for fuzzy set, constraints for universe of discourse, constraints for fuzzy information granules, constraints for fuzzy rules, constraints for fuzzy models and constraints for learning algorithms".
- Zhou and Gan [75] propose a taxonomy in terms of low-level interpretability and high-level interpretability. *Low-level interpretability* of fuzzy models is achieved on the fuzzy set level by optimizing membership functions (MFs) in terms of the semantic criteria of MFs (semantics-based interpretability) and *high-level interpretability* is obtained on the fuzzy rule level by conducting overall complexity reduction in terms of certain criteria, such as a moderate number of variables and rules (complexity-based interpretability) and consistency of rules (semantics-based interpretability).

Following our approximation, the two main kinds of approach to take into account the interpretability of linguistic FRBSs are:

1. Complexity-based Interpretability: These approaches are devoted to decreasing the complexity of the obtained model (usually measured as number of rules, variables, labels per rule, etc.).
2. Semantics-based Interpretability: These approaches are devoted to preserving the semantics associated with the MFs. We can find approaches trying to ensure semantic integrity by imposing constraints on the MFs or approaches considering measures such as distinguishability, coverage, etc.

Since both kinds of measures, complexity-based interpretability and semantic-based interpretability, should be considered in both KB components, linguistic fuzzy partition and RB, the taxonomy is based on a double axis:

- Complexity versus semantic interpretability.
- Rule base versus fuzzy partition.

Finally, the proposed taxonomy comes from combining both axes. This leads to the appearance of the following quadrants devoted to analyzing the interpretability of linguistic FRBSs (see Table 1):

- Q_1: Complexity at the RB level, analyzed in Section 3.
- Q_2: Complexity at the fuzzy partition level, analyzed in Section 4.
- Q_3: Semantics at the RB level, analyzed in Section 5.
- Q_4: Semantics at the fuzzy partition level, analyzed in Section 6.

The *Complexity* problem has been solved in the literature in different ways. Some works use techniques such as merging [30,31,47,65] (to reduce the number of MFs) or rule selection [4,27,28,34,36,37,39,42] (to reduce the number or rules) or methods for rule learning [37,42] (directly obtaining simple models). We will also consider these techniques when they explicitly mention complexity reduction as one of their inherent properties.

The following sections analyze the different quadrants trying to emphasize the descriptions of those interpretability criteria/measures proposed within each of them. At this point we would like to clarify that our aim is to provide as complete as possible descriptions of the approaches in general and of the measures used in particular, including original formulations and terminologies (instead of setting a common unified terminology) in order to maintain these descriptions as close as possible to the original ones, thus easing checking the original contributions if any detail remains unclear. In this way, some descriptions could be longer than others due to the need for a good description and not to their own importance. In fact, even though we will draw some conclusions and/or recommendations at the end of the paper, since this is a very subjective and controversial topic it is not our aim to determine which proposals are good or bad but to establish a well described framework in order to help readers have a good global vision in order to easily come to their own conclusions or better approach their own future work.

Table 1
A taxonomy to analyze the interpretability of linguistic FRBSs.

	Rule base level	Fuzzy partition level
Complexity-based interpretability	Q_1 number of rules number of conditions	Q_2 number of membership functions number of features
Semantic-based interpretability	Q_3 consistency of rules rules fired at the same time transparency of rule structure (rule weights, etc.) cointension	Q_4 completeness or coverage normalization distinguishability complementarity relative measures

In addition, some works will be related to only one quadrant, while others may be related to several quadrants simultaneously. There are two possible reasons to include a work within several quadrants. One reason may be that the work seeks to improve interpretability by approaching it from several angles, for example including measures like the number of rules (Q_1) and distinguishability of the MFs (Q_4). Another reason is that, by improving a measure of one quadrant one can also produce an improvement in the measures of a different quadrant, for example reducing the number of MFs (Q_2) as a way to reduce the number of rules (Q_1).

In order to keep descriptions complete, those works related to more than one quadrant will be completely explained once, in the quadrant in which they are more relevant, but clearly indicating which parts are related to other quadrants (they will also be referenced at the end of the other quadrants in order to maintain their completeness). For each quadrant the descriptions of the papers have been included chronologically.

3. Q_1: Complexity at the rule base level

In this section, we analyze those criteria that try to reduce or to control the complexity of the RB. The more commonly used measures are the following:

- *Number of rules*: According to the principle of Occam's razor (the best model is the simplest one fitting the system behavior well), the set of fuzzy rules must be as small as possible under conditions in which the model performance is preserved to a satisfactory level.
- *Number of conditions*: The number of conditions in the antecedent of a rule must not exceed the limit of 7 ± 2 distinct conditions, which is the number of conceptual entities a human being can handle [56]. Furthermore, the number of conditions should be as small as possible in order to ease the readability of the rules. Of course, the model performance must also be maintained to a satisfactory level.

Some considerations we should take into account with respect to the aforementioned measures and their descriptions are:

- Differences in simplicity are only remarkable when they are big enough, for example a system with 30 and another one with 32 (or one with 5 and other one with 3) rules, are in practice, at the same level.
- According to the definition of Zhou et al. [75], a system must be as simple as possible without seriously affecting its accuracy/usefulness. Nevertheless, even though we agree that having a too simple and very bad system cannot be usefully applied to a real problem, these kinds of systems could allow us to have a general idea of the system's behavior. Generalizations are not usually the most accurate, but a generalization can be useful to show trends.

In the following, we provide a brief review of the approaches that directly take into account the number of rules, conditions, etc., or that include techniques to control or to reduce the complexity at the RB level, for example rule selection.

Ishibuchi et al. [39] propose a genetic algorithm for rule selection in classification problems, considering the following two objectives: To maximize the number of correctly classified training patterns and to minimize the number of selected rules. This improves the complexity of the model, thanks to the reduction in the number of rules and the use of "*don't care*" conditions in the antecedent part of the rule. A two-objective genetic algorithm for finding non-dominated solutions in classification problems has also been proposed in [35], with the same two objectives. Then, in [36] both single-objective and two-objective genetic algorithms are studied, to perform the rule selection on an initial set of classification rules involving "*don't care*" conditions and considering the aforementioned objectives: Classification accuracy and number of rules. In addition, Ishibuchi et al. [37] present a multi-objective evolutionary algorithm (MOEA) for classification problems with three objectives: Maximizing the number of correctly classified patterns, minimizing the number of rules and minimizing the number of antecedent conditions. Moreover, they consider two approaches, one for rule selection and a second for rule learning. In [41], they examine the effect of fuzzy partitioning and condition selection in order to find a good trade-off between the number of fuzzy rules and classification performance, by using the said genetic algorithm devoted to maximizing the classification error and to minimizing the number of fuzzy rules. A new approach for regression problems is presented in [40]. They discuss the design of linguistic models with high interpretability considering a fuzzy genetics-based machine learning algorithm and using a Pittsburgh approach. They explain how the formulated linguistic modelling problem can be handled by single-objective and multi-objective genetic algorithms, with three objectives to minimize: Total squared error of the rule set, number of fuzzy rules in the rule set and total rule length of the fuzzy rules in the rule set (or total number of antecedent conditions). In the case of the single-objective approach, they used a weighted sum of the three objectives as a fitness function. Furthermore, they consider that using "*don't care*" as an additional antecedent fuzzy set is necessary in order to linguistically describe high-dimensional nonlinear functions.

Ishibuchi et al. [42] apply an improved MOEA, the multi-objective genetic local search [34] (MOGLS) for classification problems, considering the same approach as in [37] with three objectives: Maximizing the number of correctly classified training patterns, minimizing the number of fuzzy rules, and minimizing the total rule length of fuzzy rules. The approach consists of two phases: First, the method generates candidate rules by using rule evaluation measures and second, the method applies a multi-objective based rule selection. They propose the use of two well-known data mining criteria (confidence and support), in order to find a tractable number of candidate fuzzy if-then rules. More specifically, the confidence indicates the degree of the validity of a rule and the support indicates the degree of coverage of a rule.

An automatic method combining different heuristics for designing fuzzy systems from data in classification problems is proposed by Mikut et al. [55]. They integrate in the algorithm the following components to improve interpretability:

- Generation of rules by decision tree induction and by using a pruning method in order to obtain simple rule conditions and to lead to derived linguistic terms.
- Decreasing the number of generated rules by using an interactive rule selection algorithm, that uses a measure of the relevance, defined as:

$$Q = \underbrace{\left(1 - \frac{E}{E_0}\right)}_{Q_{ac}} Q_{cl}^{\beta} \quad \text{where} \quad Q_{cl} = \prod_{r=1}^{r_{max}} max_j(\hat{p}(B_j|P_r)),$$

  where Q is the compromise between classification accuracy (Q_{ac}) and clearness of the rules (Q_{cl}), β is used to control the compromise ($\beta \geqslant 0$), E is the minimum quadratic error in terms of membership values of the output classes, E_0 is the minimum quadratic error of the trivial model (a rule with an always true premise), r_{max} is the number of rules, and $\hat{p}(B_j|P_r)$ is the probability of "y is in class B_j" for premise P_r.
- Feature selection is used to reduce the number of features (quadrant Q_2) by determining the most important features. To do this, they present a feature relevance measure that reflects the preference and relevance of a feature:

$$M_l = M_{l,ap}^{\alpha} \frac{H(x_l; y)}{H(y)} \quad \text{where} \quad H(y) = -\sum_{j=1}^{m_y} p(B_j)\, lg\,(p(B_j)),$$

  where $H(y)$ is the entropy of the output y, $H(x_l;y)$ is a measure of the average information provided by feature x_l about the class of y, $M_{l,ap}$ is a relevance weight provided by the user, α is the strength of the feature preference (if α is near to zero the influence of a priori preferences diminishes), m_y is the number of classes, $p(B_j)$ is the probability of the event "y is in class B_j" and lg is the logarithm in base 2.

A genetic algorithm to perform genetic tuning combined with a fuzzy rule set reduction process that obtains a compact RB with a reduced number of rules has been proposed by Casillas et al. in [14]. Moreover, they combined linguistic hedges with classic three definition parameter tuning and with domain learning to improve the performance of the system while maintaining its complexity.

Narukawa et al. in [57] propose an adaptation of the well-known NSGA-II [21] in order to reduce complexity by decreasing the number of rules using three different mechanisms: removing overlapping rules, merging similar rules and recombining both very different and similar parents. This algorithm includes the following three objectives: Maximizing the number of correctly classified training patterns, minimizing the number of fuzzy rules and minimizing the total number of antecedent conditions.

An MOEA for classification problems considering a hybridization of the Michigan and the Pittsburgh approaches is proposed by Ishibuchi and Nojima in [38]. They analyze the interpretability-accuracy trade-off of fuzzy systems considering three formulations for multi-objective optimization problems (MOPs) and three formulations for single objective optimization problems (SOPs):

- MOP-1: Maximize f_1 and minimize f_2.
- MOP-2: Maximize f_1 and minimize f_3.
- MOP-3: Maximize f_1, minimize f_2, and minimize f_3.
- SOP-1: Maximize $w_1 \cdot f_1 - w_2 \cdot f_2$.
- SOP-2: Maximize $w_1 \cdot f_1 - w_3 \cdot f_3$.
- SOP-3: Maximize $w_1 \cdot f_1 - w_2 \cdot f_2 - w_3 \cdot f_3$,

where w_1, w_2 and w_3 are specified non-negative weights and f_i represents each objective considered: f_1 is the number of correctly classified training patterns, f_2 is the number of fuzzy rules and f_3 is the total number of antecedent conditions of the fuzzy rules, excluding "*don't care*" conditions. The experimental results in [38] demonstrate "the potential advantages of multi-objective formulation over single-objective ones".

Liu et al. in [47] present a Mamdani neuro-fuzzy system for balancing interpretability and accuracy. The improvement in interpretability takes place thanks to the reduction in the number of MFs, number of rules and number of attributes. They propose reducing the number of rules by using a method for merging the fuzzy sets and by considering a Hebbian ordering. "The Hebbian ordering is used to represent the importance of the rules, where a higher Hebbian ordering indicates a larger coverage of the training points provided by a given rule. The rules with higher importance are more likely to be preserved". The reduction of the number of rules can provoke the appearance of inconsistent rules. However, the authors solve the problem of inconsistent rules by maintaining only the most important rules, with a mechanism for controlling the consistency of the RB (quadrant Q_3).

Alcalá et al. in [1] propose an effective model of tuning for FRBSs combined with a rule selection, considering the linguistic 2-tuples representation scheme introduced in [33], in order to improve the performance and to decrease the complexity of the classic tuning approaches in complex search spaces. The linguistic 2-tuples allows the lateral displacement of the labels (in fact, the MFs) by considering only one parameter (slight displacements to the left/right of the original MFs). Since the three parameters usually considered per label are reduced to only one symbolic translation parameter, this proposal decreases the learning problem complexity, helping to decrease the model error and facilitating a significant decrease in the model complexity.

Cococcioni et al. in [17] propose an efficient modified version of the well-known PAES [46] and appropriate genetic operators have been proposed to learn a set of Mamdani FRBSs with different trade-offs between accuracy and complexity. They propose a Pareto-based multi-objective evolutionary approach in regression problems, considering the following two objectives: Minimizing the MSE and minimizing the global number of conditions.

An MOEA to obtain a set of solutions with different degrees of accuracy and interpretability has been proposed by Pulkkinen et al. in [66]. They use the number of misclassifications, the number of rules and the total rule length as objectives to be minimized. The authors apply the C4.5 algorithm [68] to initialize the evolutionary algorithm. First they create a decision tree through C4.5 and transform this decision tree into a fuzzy classifier. Thus, relevant variables are selected and an initial partition of the input space is performed. Then, they perturb randomly some parameters of this fuzzy classifier to generate the individuals of the initial population of the MOEA. The C4.5 algorithm implicitly includes a mechanism for reducing the number of features (quadrant Q_2).

An enhanced MOEA for regression problems has also been proposed in [4] by Alcalá et al., and deeply discussed in [27] by Gacto et al., which aims to minimize the number of rules together with a classic tuning of the MF parameters (three parameters) by focusing on the most accurate part of the Pareto front in order to find the best trade-off between complexity and accuracy (since both objectives present different levels of difficulty). They used the MSE and the number of rules as objectives to be minimized. A real application of a refined version of the proposed algorithm can be found in [26] as a way of improving the performance of linguistic models obtained from experts by forcing rule reduction in a complex real problem for the control of Heating, Ventilating and Air Conditioning Systems. Alcalá et al. in [3] propose an MOEA for learning RBs and parameters of the MFs of the associated linguistic labels concurrently (they use the linguistic 2-tuples representation [33] by only considering one parameter per MF). These works [3,4,27] generate a set of FRBSs with different near optimal trade-offs between accuracy and complexity for regression problems.

Moreover, in this quadrant we can find works that use as a measure of complexity, the number of rules [2,18–20,28,43,51], the number of conditions [6,58,64,67] or both [8–10,30,31,63,65], combined with measures proposed for other quadrants in which they will be described because of their relevance.

4. Q_2: Complexity at the level of fuzzy partitions

Several measures have been classically considered in the literature to control the complexity at the level of fuzzy partitions. Among them, the most used measures that are found in this quadrant are:

1. Number of Features or Variables: To reduce the dimensionality in high dimensional problems. The reduction of the number of features can improve the readability of the KB.
2. Number of MFs: To control the complexity at the level of fuzzy partitions, it is necessary to have a *moderate number of MFs*. The number of MFs should not exceed the limit of 7 ± 2 distinct MFs, which is the number of conceptual entities a human being can handle [56]. As soon as the number of MFs increases the precision of the system may increase too, but its relevance will decrease.

The number of variables as well as the granularity (number of linguistic terms or MFs) of the fuzzy partitions determine the specificity or generality of the model that can be obtained, and they influence proportionally the number of rules of the obtained models [5]. Therefore, Q_2 is highly related to Q_1 and many of the works in this quadrant can also be found in the previous one. Examples of it are the techniques for rule reduction based on decreasing the number of MFs (merging rules) using similarity measures among MFs.

We have to clarify that, firstly, we will include here only those works performing feature selection as a way to reduce the number of features. The selection of conditions has been taken into account as a different technique since it can cause a variable not to be used in one or several rules but does not disappear from the KB. However, feature selection eliminates the variable from the KB completely. Secondly, we will also include here those works focused on decreasing or controlling the number of MFs of the fuzzy partitions.

In the following works [2,18,19], Cordón et al. and Alcalá et al. present models for embedded evolutionary learning of linguistic fuzzy partitions in regression problems.

The initial proposal in [18] learns the granularity (number of labels) of the fuzzy partitions and the MFs' parameters (their three parameters jointly in the case of triangular membership functions). At the same time the authors in [19] propose an evolutionary algorithm to learn the granularity, scaling factors and the domains (i.e., the variable domain or working range to perform the fuzzy partitioning) for each variable. Alcalá et al. in [2] also propose a method for learning KBs by means of an a priori evolutionary learning of the linguistic fuzzy partition (granularity and translation parameters) that uses the linguistic 2-tuples representation [33]. This methodology allows the reduction of the search space, obtaining more optimal models with high levels of accuracy and simplicity. All these works control the complexity of the RB and linguistic fuzzy partition. In order to do this, they penalize the fitness function with the number of rules obtained, learning models with lower granularities and therefore with a smaller number of rules (quadrant Q_1). The fitness function is defined for minimization:

$$F = w_1 \cdot MSE + w_2 \cdot NR$$

where NR is the number of rules of the obtained KB, w_1 = 1 and w_2 is computed from the KB generated from a linguistic fuzzy partition considering the maximum number of labels ($max - lab$, usually 9) and with the MF parameters, $w2 = \alpha \cdot (MSE_{max-lab} / NR_{max-lab})$ with α being a weighting percentage given by the system expert that determines the trade-off between accuracy and complexity. Values higher than 1.0 search for linguistic models with few rules, and values lower than 1.0 search for linguistic models with high accuracy. Additionally, Cordón et al. in [20] propose an MOEA for performing feature selection together with linguistic fuzzy partition learning, in order to learn the number of labels for each variable and to adjust the shape of each MF in non-uniform fuzzy partitions, using a non-linear scaling function. They consider the following two objectives: To minimize the classification error percentage and to minimize the complexity (number of features and number of conditions). All these methods consider an evolutionary process that learns the number of MFs.

Tikk et al. in [73] present an algorithm for feature selection in order to reduce the complexity in classification problems. They search for features which maximize the average distance between the classes. They propose the use of the sequential backward selection [22] as a search method in order to rank the features. This method removes one feature at each stage of the search process, in this way reducing complexity.

A feature selection algorithm which should be useful in high dimensional classification problems is proposed by Vanhoucke et al. in [74]. This algorithm ranks the input features according to their mutual information, and discards all features deemed irrelevant by a threshold criterion.

An index for classification problems to measure the interpretability (I) of linguistic fuzzy models has been proposed by Nauck in [58], in terms of complexity ($comp$), the degree of coverage ($\overline{cov}$) of the fuzzy partition, and a partition complexity measure ($\overline{part}$) that penalizes partitions with a high granularity. The first measure is the average number of conditions per class in classification problems (quadrant Q_1). The second measure takes into account the coverage of the fuzzy partitions (quadrant Q_4). Thanks to this last measure ($\overline{part}$) it tries to obtain a fuzzy rule based system with a small number of fuzzy sets. This index is defined as:

$$I = comp \cdot \overline{cov} \cdot \overline{part}.$$

In the following, these measures are formulated. The complexity measure ($comp$) is defined as:

$$comp = m / \sum_{i=1}^{r} n_i$$

where m is the number of classes, r is the number of rules and n_i is the number of variables used in the ith rule (also included in the Q_1 quadrant).

The degree of coverage $\overline{cov}$ is defined as the average normalized coverage on cov_i:

$$cov_i = \frac{\int_{x_i} \bar{h}_i(x)dx}{N_i} \quad \text{where } \bar{h}_i(x) = \begin{cases} h_i(x) = \sum\limits_{k=1}^{p_i} \mu_i^{(k)}(x), & \text{if } 0 \leqslant h_i(x) \leqslant 1 \\ \frac{p_i - h_i(x)}{p_i - 1}, & \text{otherwise} \end{cases}$$

where X_i is the domain of the ith variable and this domain is partitioned by p_i fuzzy sets and with $N_i = \int_{x_i} dx$ for continuous domains or with $N_i = |X|$ for discrete domains. The partition complexity measure $\overline{part}$ is the average normalized partition measure on $part_i$:

$$part_i = 1/p_i - 1$$

where p_i is the number of fuzzy sets in the ith variable.

After a preliminary review of interpretability-oriented fuzzy inference systems obtained from data presented by Guillaume in [29], Guillaume and Charnomordic in [30] present a method for generating interpretable fuzzy rules from data. They include a procedure to simplify an RB (quadrant Q_1) in order to get what they call "incomplete rules". These are rules that include "*don't care*" conditions and are defined by only a few variables. Therefore, these rules are easier to interpret than "complete rules" (those considering all the system variables). In [31], they present a fuzzy inference system derivation method together with a simplification algorithm, which includes a mechanism for removing unnecessary rules (quadrant Q_1) combined with a procedure for the selection of variables and fuzzy sets. Additionally, in both works [30,31], they propose a sophisticated distance function, with external and internal distances, which is used to merge fuzzy sets, thus allowing a moderate number of MFs for classification problems to be obtained. They propose the application of the merging of fuzzy sets that minimizes the variation of the D_m index defined for a given size m partition as:

$$D_m = \frac{1}{N(N-1)} \sum_{q,r=1,2,\ldots N, q\neq r} d(q,r)$$

where m is the number of fuzzy sets in the fuzzy partition, N is the number of training data and the pairwise distance $d(q,r)$ will take into account the memberships of the different q and r training points by combining the respective parts of internal and external distances. Internal and external distances as well as the pairwise distance $d(q,r)$ are defined as follows:

The *internal distance* is a measure of membership similarities for a given fuzzy set f and it is computed, given two data points with $\left(x_q^j, x_r^j\right)$ coordinates for the jth dimension, by means of the difference of the membership degrees:

$$d_{int}^f(q,r) = \left|\mu_q^f - \mu_r^f\right|$$

The *external distance* is a measure combining internal distances and prototype distances. It takes into account the point location within the fuzzy set and the relative position of the fuzzy set in the fuzzy partition. The external distance between two points which belong to the f and g fuzzy sets respectively is defined as:

$$d_{ext}^{f,g}(q,r) = \left|\mu_q^f - \mu_r^g\right| + d_{prot}(f,g) + D_c$$

where D_c is a constant correction factor, which ensures that the external distance is always superior to any internal distance, and $d_{prot}(f,g)$ is the prototype distance (numerical or symbolic distance) between the centers of fuzzy sets f and g.

Taking into account that $d_{f,g}(q,r)$ represents respective memberships of the fuzzy sets f and g, and that it is an internal distance if $f = g$ or an external distance otherwise, $d(q,r)$ is defined as:

$$d(q,r) = \frac{1}{\sum_{f=1}^{m} \mu_q^f} \sum_{f=1}^{m} \left[\mu_q^f \frac{1}{\sum_{g=1}^{m} \mu_r^g} \sum_{g=1}^{m} \left[\mu_r^g d_{f,g}(q,r)\right]\right]$$

Additionally in [31], to assess the validity of a fuzzy partition they propose a new index based on the homogeneity of the fuzzy set densities (quadrant Q_4). The density d_f for a fuzzy set f is equal to the ratio of its weight, or fuzzy cardinality w^f defined as $w^f = \sum_{x\in E} \mu_j^f(x)$, divided by the fuzzy set area, where E is the subset of learning samples covered by f. The density homogeneity σ^{FP}, is defined as the density standard deviation for all the fuzzy sets of the fuzzy partition:

$$\sigma^{FP} = \sqrt{(1/m)\sum_{f=1}^{m} (d_f - \bar{d})^2},$$

where $\bar{d}$ is the mean of the fuzzy set densities. From the homogeneity point of view the best partition is the one for which σ^{FP} reaches a minimum.

Additionally, we should also mention in this quadrant some works whose main contributions are in other quadrants but also take into account the number of features [8,55,65,66] or the number of MFs [23,43,59,60], some of them by simply imposing an upper bound on the number of MFs [59,60].

5. Q_3: Semantics at the rule base level

Assuming that the initial fuzzy partitions are interpretable at the semantic fuzzy partition level, this quadrant is related to measures or properties that are devoted to controlling the semantic interpretability at RB level. Mainly, this quadrant takes into account the following properties:

- *Consistency of the RB*, which means the absence of contradictory rules in RB, in the sense that rules with similar premise parts should have similar consequent parts. The analysis of consistency should be different depending on the kind of problem (classification, regression, control, etc.). Most of the works in this quadrant are focused on this property.
- *Number of rules fired at the same time*, which consists of minimizing the number of rules firing that are activated for a given input. There are only a few works that make use of the number of rules fired, among them the works in [6,15,16,51]. However, in our opinion this measure represents a very promising way to preserve the individual meanings of the linguistic rules comprising the RB.

This is the quadrant in which there are fewer works in the literature. In what follows, we describe those works that look to improve the interpretability in this quadrant.

Jin et al. in [44,45] propose a methodology based on evolutionary strategies for generating flexible, complete, consistent and compact fuzzy rule systems from data using evolutionary algorithms. They propose some indices for consistency and coverage (quadrant Q_4) of the linguistic fuzzy system and they integrate them into an aggregated objective function f. The consistency index ($Cons$) is calculated as follows for two given rules $R(i)$ and $R(k)$:

$$Cons(R(i), R(k)) = exp\left\{ -\frac{\left(\frac{SRP(i,k)}{SRC(i,k)} - 1.0\right)^2}{\left(\frac{1}{SRP(i,k)}\right)^2} \right\}$$

where SRP is the similarity of rule premises and SRC is the similarity of rule consequents, and they are calculated as:

$$SRP(i,k) = \min_{j=1}^{n} S(A_{ij}, A_{kj}), \quad SRC(i,k) = S(B_i, B_k)$$

where S is the fuzzy similarity measure defined in [69]. The S measure is a fuzzy relation that expresses the degree to which fuzzy sets A and B are equal and is defined as follows:

$$S(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|},$$

where intersection ($\cap$) and union ($\cup$) are defined by a proper couple of t-norm and t-conorm and $|\cdot|$ is the cardinality of the resulting fuzzy set. This value ranges from 0 to 1. Finally, the degree of inconsistency of a given RB (f_{Incons}) is calculated as:

$$f_{Incons} = \sum_{i=1}^{N} Incons(i),$$

where $Incons(i)$ is the degree of inconsistency for the ith rule. It is defined as:

$$Incons(i) = \sum_{\substack{1 \leqslant k \leqslant N \\ k \neq i}} [1.0 - Cons(R^1(i), R^1(k))] + \sum_{\substack{1 \leqslant l \leqslant L \\ i=1,2,\ldots,N}} [1.0 - Cons(R^1(i), R^2(l))],$$

where R^1 and R^2 denote the RB generated from data and the RB extracted from prior knowledge (since the authors defined this index considering the possibility of including rules provided by experts) and N and L are the rule numbers of R^1 and R^2, respectively.

The proposed approach is applied to the design of a distance controller for cars. In this problem the objective function f is formulated as:

$$f = f_E + \xi \cdot f_{Incons} + f_{Incompl} \quad \text{where} \quad f_E = \sum_{t=1}^{J} \sqrt{(v(t) - v_d(t)^2)},$$

J is the total number of sampled data, $v_d(t)$ is the target velocity, $v(t)$ is the velocity of the controlled car, ξ is a weighting constant to control the consistency level and $f_{Incompl}$ is a penalization constraint to maintain the completeness of the fuzzy partition. Moreover, the coverage and the distinguishability of the fuzzy partitioning of each input variable is examined using the fuzzy similarity measure (S). To keep the MFs with a proper shape, the fuzzy similarity measure of any two neighboring MFs is required to satisfy the following constraint:

$$FSM^- \leqslant S(A_i, A_{i+1}) \leqslant FSM^+$$

where A_i and A_{i+1} are two neighboring fuzzy sets and both FSM^- and FSM^+, are the desired lower and upper bound of the fuzzy similarity measure respectively.

In [43], Jin proposes a methodology based on genetic algorithms and the gradient learning method. He reduces the number of rules (quadrant Q_1) removing redundant rules by means of the previously defined similarity of the rule premise (SRP), proposed in [45]. As in the case of the previous work, it is also devoted to controlling the consistency of the RB. Moreover, the author presents a regularization learning to reduce the number of fuzzy sets (quadrant Q_2) while distinguishability is improved (quadrant Q_4). “The regularization is to drive the similar fuzzy sets to the same fuzzy set during gradient learning so that the interpretability of the fuzzy system can be greatly improved without seriously deteriorating the system performance”. The cost function used for the regularization is defined as:

$$J = E + \gamma\Omega$$

where E is the conventional error function, Ω is the regularization term for merging the similar fuzzy MFs and γ is the regularization parameter ($0 \leqslant \gamma < 1$). If γ takes high values the system performance is degraded, whereas if γ takes low values the interpretability of the fuzzy system is bad. The author assumes that variable x_i has L_i fuzzy subsets A_i, $i = 1,2,\ldots,L_i$. They can then be divided into m_i (initially equal to L_i) groups using a prescribed similarity threshold δ

$$U_{ik} = \{A_i | S'(A_i, A_{k0}) \geqslant \delta)\}; \quad 1 \leqslant k \leqslant L_i$$

where U_{ik} denotes a group of fuzzy subsets that are considered to be similar, A_{k0} is the reference fuzzy set for the group and S' is a different definition of the similarity measure S [69]. This is based on a distance measure, $d(A_1,A_2)$ based on the existence of gaussian MFs (with a_i and b_i the core and width definition parameters for the MF A_i):

$$S'(A,B) = \frac{1}{1 + d(A_1, A_2)}, \quad \text{where} \quad d(A_1, A_2) = \sqrt{(a_1 - a_2)^2 + (b_1 - b_2)^2}$$

The goal of regularization is to drive the similar fuzzy sets into the same fuzzy set, and the regularization term is defined as:

$$\Omega = \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{m_i}\sum_{A_{ij}\in U_{ik}}(a_{ij} - \bar{a}_{ik})^2 + \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{m_i}\sum_{A_{ij}\in U_{ik}}(b_{ij} - \bar{b}_{ik})^2,$$

$$\text{such that } \bar{a}_{ik} = \frac{1}{I_{ik}}\sum_{A_i\in U_{ik}}^{I_{ik}} a_{ij} \text{ and } \bar{b}_{ik} = \frac{1}{I_{ik}}\sum_{A_i\in U_{ik}}^{I_{ik}} b_{ij}$$

where $\bar{a}_{ik}$ and $\bar{b}_{ik}$ are the parameters of the gaussian function (core and width) shared by all the fuzzy subsets in group U_{ik}, n is the total number of inputs, I_{ik} is the number of fuzzy sets in the group U_{ik} and a_{ik} and b_{ik} are the core and the width of each subset A_i in the same group U_{ik}.

Cheong and Lai in [15,16] present a parallel genetic algorithm for obtaining a fuzzy logic controller with some constraints in the RB. This algorithm tries to minimize the number of rules fired at the same time, which is not only devoted to improving the consistency of the RB, but also to reducing the effort needed to understand the meaning of the rules since they present less interactions among them. This helps to provide rules that can be better understood by human beings. However, this was an indirect consequence of their work, as the authors proposed to reduce the number of rules firing at the same time because this "has a negative impact on computation speed". Therefore, the authors did not really consider this as a way to improve interpretability. On the other hand, the authors use semantic constraints that are applied to the optimization process in order to produce well-formed fuzzy sets. In this parallel genetic algorithm, they use strong fuzzy partitions with triangular MFs and only one parameter is used for the evolutionary adaptation of the MFs. This is the location of the center of the triangle while the remaining parameters are moved automatically since the last point of each MF is kept equal to the center point of the following MF. In a normalized universe of discourse in $[-1,1]$, the first and last triangles were fixed to -1 and 1, respectively, and the other triangles constrained to the range $[X_1^s, X_2^s]$, which is defined as follows for the center of the sth triangle:

$$[X_1^s, X_2^s] = (2/n * (s-1) \pm a)$$

where n is the number of fuzzy sets in the universe of discourse, $1 < s < n$, and a is constant with a value determined experimentally (an adequate portion of the discourse universe). These semantic constraints are used in order to guarantee distinguishability (quadrant Q_4).

Pedrycz in [61] analyzes the interpretability of an RB by using two measures, relevance and consistency:

1. The relevance of a rule "is quantified in terms of the data being covered by the antecedent and conclusions standing in the rule". This measure is defined by the author as:

$$rel(A_i \times B_i) = \sum_{k=1}^{N} A_i(x_k) t B_j(y_k),$$

where A_i is the antecedent of the ith rule, B_j is the consequent of the rule, N is the number of data and t is the t-norm used. This measure presents higher values when the rule has more relevance in the RB.
2. The consistency of the rule "expresses how much the rule interacts with the others in the sense that its conclusion is distorted by the conclusion parts coming from other rules". This measure is defined as:

$$Cons(i,R) = \sum_{j=1, j\neq i} cons(i,j),$$

$$cons(i,j) = \frac{1}{N}\sum_{k=1}^{N} A_i(x_k) \equiv A_j(x_k) \to (B_i(y_k) \equiv B_j(y_k)),$$

where i and j are the indexes of the rules in R and → is the implication operator used.

Alonso et al. in [9] propose a methodology for designing Highly Interpretable Linguistic KBs (HILK) for classification problems, considering both expert knowledge and knowledge extracted from data. This methodology includes procedures to merge and to simplify rules, obtaining shorter rules by removing unused variables. In addition, they present a fuzzy interpretability index to quantify the system complexity inspired by Nauck's index, which combines the following six criteria:

- Total number of rules (quadrant Q_1).
- Total number of premises (quadrant Q_1).
- Number of rules which use one input.
- Number of rules which use two inputs.

- Number of rules which use three or more inputs.
- Total number of labels defined per variable.

These criteria are taken as inputs of a fuzzy system. The fuzzy interpretability index is computed as the result of the inference of a hierarchical fuzzy system made up of four linked KBs generated by HILK. This also takes into account the conflicting rules by using the following solutions:

- Totally inconsistent rules are rules with the same premises and different conclusions. When there are several rules with the same premises and different conclusions, only the one covering a bigger number of samples from the training data file is kept, while the others are removed.
- Specialization rule is a rule in which the space of the premises are included in another rule, and both rules have different conclusions, i.e., a specific rule is a specialization of the most general one. There are two different ways to avoid this contradiction:
  - Keep only one rule corresponding to the largest input domain and set the best suited consequent.
  - Keep the most specific rule and split the most general one, to cover the same input domain except the one covered by the specific rule.
- Partially inconsistent rules are rules with no empty intersection in the premises and with different consequents. The way to avoid the intersection in the premises is splitting these rules and choosing the best suited consequent.

Moreover, the authors in [8] present an extension of HILK (called HILK++), that includes feature selection (quadrant Q_2) based on the C4.5 algorithm [68], but also a new index called *RBC* (rule-base complexity) which takes into account the presence of NOT and OR labels in the fuzzy premises. This rule-based complexity index is defined as:

$$RBC = \sum_{j=1}^{NR} \left[\prod_{a=1}^{NI} \overbrace{\left(2 - \frac{LT_a^j}{NL_a}\right)}^{complexity(P_a)} \right]$$

where *NR* is the number of rules, *NI* is the number of inputs used in the rule, NL_a is the number of elementary terms defined in the fuzzy partition of the input I_a and LT_a^j is the number of elementary terms included in linguistic term A_a^i, which is computed by considering the following values:

- One for elementary terms.
- Number of elementary terms combined with OR. For example, the expression "medium or high" is equal to two.
- NL_a minus one half for NOT composite terms, which penalizes NOT against OR composite terms to participation by all but one elementary terms.
- If the input I_a is not considered in the rule then $complexity(P_a) = 1$.

This work also incorporates the consistency analysis presented in [9].

Alonso et al. in [10] propose a methodology in order to analyze different measures (many of them in Q_1) by using a web poll dedicated to determining how different people assess interpretability by giving priority to different criteria. The authors consider ten variables as tentative interpretability indicators: Number of rules, total rule length, number of inputs, number of labels used in the RB, percentage of rules which use less than ten percent of inputs, percentage of rules which use between ten and thirty percent of inputs, percentage of rules which use more than thirty percent of inputs, percentage of elementary labels used in the RB, percentage of OR composite labels used in the RB and percentage of NOT composite labels used in the RB. Finally, they conclude that the results extracted from the poll show the inherent subjectivity of the measures, obtaining a huge diversity of completely different answers. However, "three interpretability indicators turn up as the most significant ones: Total rule length, number of used labels in the RB, and percentage of NOT composite linguistic terms". In fact, the use of NOT composite linguistic terms or even OR disjunctive operators could also affect the transparency of the rule structure. This work is also included in this quadrant, since it uses the analysis of consistency previously described for [9] in order to remove inconsistent rules.

Alonso et al. in [6] propose an embedded HILK, genetic fuzzy partition learning process, developed by means of an MOEA (based on the well-known NSGAII algorithm) which tackles the joint optimization of three objectives: Maximizing the classification rate, maximizing the readability of the system description by means of minimizing the total rule length (quadrant Q_1) and maximizing the comprehensibility of the system explanation by means of minimizing the average number of rules which are fired at the same time. This work is included in this quadrant since it is the first work considering the number of simultaneous fired rules from the interpretability point of view.

Mencar et al. in [53] propose "an approach for automatically evaluating interpretability of rule-based fuzzy classifiers, exploiting the propositional view of rules as a mean to define co-intension". They evaluate how much the semantics of fuzzy rules is coherent with their logical view. The novelty of these works arises from the fact that they present a new way to measure the fuzzy rules' interpretability from the point of view of understandability (comprehensibility) instead of regarding only complexity as is usual.

Márquez et al. in [51] propose an MOEA for regression problems with a mechanism to improve the interpretability in the sense of complexity for linguistic fuzzy rule based systems with adaptive defuzzification, considering the following three objectives:

- Maximizing the accuracy.
- Minimizing of the number of final rules (*#RF*) (Q_1 quadrant). This measure is based on the idea that those rule weights close to 0 represent a low influence of that rule, and it is defined as: $\#RF = \#R - (\textit{number of rule weights close to } 0,\ \textit{i.e}, \leqslant 0.1)$ where $\#R$ is the number of initial rules.
- An interpretability index $R_p\_MR_{TG}$ based on the aggregation of two metrics: Minimizing the number of rules with weights associated (*#RW*) and the average number of rules triggered by each example (MR_{TC}). Thanks to this last measure, this paper is included in the quadrant Q_3.
  This measure *#RW* is based on the idea that weight values close to 1 are those in which the rule is important and could be used without any weight, and so remove this value, thus reducing the system complexity. This is defined by: $\#RW = \#R - (\textit{number of rule weights close to} 1, \textit{i.e}, \geqslant 0.9)$.
  The index MR_{TC} measures the average number of rules triggered by each example and it is calculated as:

  $$MR_{TG} = \frac{\sum_j^M R_{TG}^j}{M}$$

  where M is the number of examples and $\#R_{TG}^j$ is the number of rules triggered by the example j.
  Finally the index is defined as:

  $$R_p\_MR_{TG} = \frac{\#RW + MR_{TC}}{2}$$

In this quadrant, there are a few measures to quantify the semantic interpretability at the level of RB. It would be interesting to propose new measures for this Q_3 quadrant or to fix the most appropriate from the existing ones. As mentioned, some additional aspects could be to measure "the percentage of OR composite labels used in the RB" and/or "the percentage of NOT composite labels used in the RB" as Alonso et al. in [10] suggest from the studies based on the web pool. Recently, the authors in [8] propose the measurement of the OR and NOT composite terms in a complexity rule-based index.

In this quadrant, we should also mention some works that are more relevant for other quadrants but that also consider the consistency of the RB [43,47,65].

6. Q_4: Semantics at the fuzzy partition level

The design process for obtaining an accurate linguistic FRBSs could lead to complex fuzzy partitions which could make interpretation of the system by an expert difficult. Fig. 1 shows an example of a complex fuzzy partition (with huge overlapping among MFs), which deteriorates the global interpretability from the point of view of semantic interpretability. In this quadrant, we look to maintain semantic interpretability at the fuzzy partition level. From a classical point of view, this problem has been tackled by applying some constraints to the MF's definition in order to preserve or to improve some desirable properties. Some of the most important properties defined by the experts in this framework are:

1. *Completeness or Coverage*: The universe of discourse of a variable should be covered by the MFs, and every data point should belong to at least one of the fuzzy sets and have a linguistic representation, i.e., it is required that membership values should not be zero for all the linguistic variable domains.
2. *Normalization*: MFs are normal if there is at least one data point in the universe of discourse with a membership value equal to one, in respect to the maximum membership degree.
3. *Distinguishability*: An MF should represent a linguistic term with a clear semantic meaning and should be easily distinguishable from the remaining MFs of the corresponding variable.

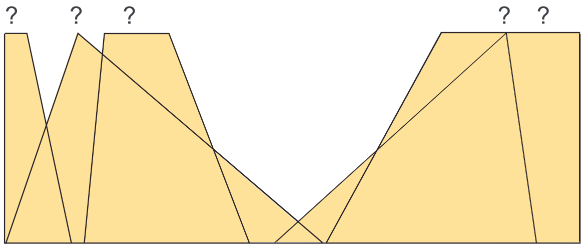

Fig. 1. Fuzzy partition with a poor semantic interpretability.

4. *Complementarity*: For each element of the universe of discourse, the sum of all its membership values should be near to one. This guarantees a uniform distribution of the meanings among the elements.

Taking into account these properties and the semantic constraints classically used to preserve them, we can observe that they represent absolute properties that try to obtain well distributed MFs, ensuring complementary membership degrees between each two adjacent concepts. In fact, *strong fuzzy partitions* satisfy these semantic properties (distinguishability, coverage, normality, complementarity, etc.) to the highest level. These kinds of fuzzy partitions, in which the sum of the membership degrees within the variable domain are equal to 1.0, perfectly meet the required semantic constraints and they are widely assumed to have high semantic interpretability, particularly when the MFs are also uniform. However, it is not always possible to use strong fuzzy partitions or to impose absolute properties. If the system is provided from expert knowledge, the experts can consider another type of fuzzy partitioning more appropriate for the problem. Therefore, it would be interesting to take into account relative measures which try to measure interpretability in respect to the fuzzy partitions intended as the most interpretable ones.

As mentioned above, some of the works that are in this quadrant look to maintain interpretability, introducing *semantic constraints* in the modelling process. By contrast, some others are devoted to proposing semantic interpretability *measures* to quantify or to optimize the semantic interpretability properties. Those works that consider constraints or measures, but simply for setting limits on some values of the MFs, are included in Section 6.1. Those works that propose measures to quantify the fuzzy partition interpretability are studied in Section 6.2.

6.1. Semantic interpretability constraints at the fuzzy partition level

This subsection presents those works that use measures to impose constraints on the MFs or that directly control the limits on the values of the MFs at the fuzzy partition level. In the following, we shortly review these approaches.

The autonomous fuzzy rule extractor with linguistic integrity (AFRELI) algorithm combined with the FuZion algorithm have been proposed by Espinosa et al. in [23]. The FuZion algorithm allows the merging of consecutive MFs, in order to reduce the number of fuzzy sets, and to maintain a justifiable number of MFs (quadrant Q_1).

The proposed algorithm guarantees distinguishability and coverage properties by imposing some constraints. A fundamental parameter of this algorithm is the minimum acceptable distance (M) between the centers of the membership functions. When the value of M is smaller, the number of acceptable MFs per domain will increase, increasing the number of rules and also increasing the complexity of the model. On the other hand, as the value of M increases the number of MFs per domain decreases, reducing the number of rules and increasing the approximation error. This parameter, which must be used to balance the trade-off between interpretability and precision, should be fixed to values between 5–25% of the coverage of the universe of discourse to guarantee semantic integrity.

Peña-Reyes and Sipper in [63] try to obtain linguistic fuzzy models with a good balance between accuracy and interpretability. To achieve this, they consider several constraints by taking into account both semantic and syntactic criteria, in order to obtain interpretable systems. They propose some strategies to satisfy the semantic and syntactic criteria during the definition of the fuzzy model:

1. Considering linguistic labels that cover all the variable domain for satisfying the completeness criterion.
2. The use of normal, orthogonal MFs.
3. Allowing "*don't care*" conditions, which reduce the number of antecedents in the rules (quadrant Q_1).
4. Including a default rule, that reduces the number of rules (quadrant Q_1).

Mencar et al. in [52] have focused on defining a faster alternative to the similarity metrics as a way of measuring the distinguishability of the MFs. The most common measure to quantify distiguishability is similarity S [69], previously explained in Section 5. The problem of the similarity measure S is that its calculation is usually computationally intensive. For this reason the authors propose in [52] the use of a possibility measure (Π) as an alternative whose calculation can be very efficient. This measure between two fuzzy sets A and B is defined as follows:

$$\Pi(A,B) = \sup_{x \in U} \; min\{\mu_A(x), \mu_B(x)\}$$

The possibility measure can also be used to evaluate distinguishability. As for the similarity measure $S(A,B)$, distinguishability can be formally defined as the complement of the possibility between two distinct fuzzy sets, which must not be less than a predetermined threshold δ:

$$\Delta(A,B) = 1 - \Pi(A,B) \geqslant \delta \text{ which implies } \forall A,B \in F : A \neq B \rightarrow \Pi(A,B) \leqslant \vartheta \text{ with } \vartheta = 1 - \delta$$

Pulkkinen and Koivisto in [67] propose "a dynamically constrained multiobjective genetic fuzzy system learning fuzzy partitions, tuning the MFs, and learning the fuzzy rules" for regression problems, considering the following two objectives: MSE and total rule length (sum of the rule lengths) for which reason it can also be found in the Q_1 quadrant. Moreover, the proposed MOEA includes different mechanisms in order to allow a decrease in the number of rules, the number of conditions, the number of MFs, and the number of input variables. Since they are also evolving MF parameters, they propose dynamic constraints in order to guarantee the distinguishability, and the coverage, of fuzzy partitions, using the following conditions:

- Symmetry conditions: The symmetry is guaranteed by definition since they use Gaussian MFs.
- α-condition: To control the intersection point of two MFs. "At any intersection point of two MFs, the membership values are at most α".
- γ-condition: To control the overlapping in the center of each MF. "At the center of each MF, no other MF receives membership values larger than γ".
- β-condition: To ensure that the Universe of Discourse is strongly covered. "At least one MF has its membership value at β".

These constraints must be fixed previously in order to apply the dynamic tuning strategies. The authors recommend as appropriate values the following: $\alpha = 0.8$, $\gamma = 0.25$ and $\beta = 0.05$. They also recommend using the following conditions for the case of highly transparent fuzzy partitions: $\alpha = 0.6$, $\gamma = 0.4$ and $\beta = 0.1$. Moreover, the proposal in [64] is devoted to enable the method proposed in [67] to be used in classification problems as well.

Other works previously explained are also included in this quadrant because they consider constraints or measures for setting limits on some values of the MFs, by using Distinguishability measures [15,16,43] or Coverage measures [44,45].

6.2. Semantic interpretability measures at the fuzzy partition level

This subsection presents those works that propose or use a measure that allows the semantic interpretability of the fuzzy partitions obtained by the different learning techniques used with this aim to be quantified. In what follows they are briefly described.

Valente de Oliveira et al. in [59,60,62] propose some semantic constraints for MFs together with some interpretability metrics or measures, including distinguishability of MFs, moderate number of MFs, natural zero positioning, normality and completeness. Almost all the most used and accepted constraints, or absolute characteristics were proposed in these first works. In this way, the authors try to avoid potential inconsistencies in linguistic fuzzy models. They propose the following expressions for coverage (J_1), and for distinguishability (J_2), and they use them to enforce the interpretability of fuzzy systems during the gaussian MFs' optimization problem using a backpropagation algorithm. These measures are defined for a given external variable as:

$$J_1 = \frac{1}{2}\sum_{k=1}^{N}(x[k] - \bar{x}[k])^2 \quad \text{where} \quad \bar{x}[k] = \frac{\sum_{i=1}^{n}\mu_i(x[k])a_i}{\sum\limits_{i=1}^{n}\mu_i(x[k])},$$

$$J_2 = \frac{1}{2}\sum_{k=1}^{N}\left[\left(\sum_{i=1}^{N}\mu_i^p(x[k])^{i/p} - 1\right)\right]^2,$$

N is the number of training data, n is the total number of elements, MFs, a_i $(i : 1\ldots n)$ are the centers of the generic MFs μ_i, $x[k]$ is the kth numeric sample and p is used to control the strength of the J_2 measure. If $p = 1$ they have a strong influence whereas it can be eliminated if $p \to \infty$.

In [59,60,62], they use a linear combination of the two constraints as a fitness function:

$$\bar{J} = J + K_1 J_1 + K_2 J_2$$

where K_1 and K_2 are positive penalty factors and J is the MSE. Moreover, earlier works [59,60] propose the use of constraints as a moderate number of MFs in the field of artificial neural networks. To do this, they impose an upper bound on the number of MFs (7 ± 2) (quadrant Q_1).

Furuhashi et al. in [25] and Suzuki et al. in [70,71], propose a conciseness measure based on the combination of De Luca and Termini's fuzzy entropy [48] and a measure for the deviation of an MF. They consider that "a fuzzy model is more concise if the MFs are more equidistantly allocated in the universe of discourse, and the shapes of MFs are less fuzzy". De Luca and Termini's fuzzy entropy, $d(A)$, can be used to evaluate the shapes of the MFs. The fuzzy entropy of a fuzzy set A is defined as:

$$d(A) = \int_{x_1}^{x_2}\{-\mu_A(x)\ ln(\mu_A(x)) - (1 - \mu_A(x))\ ln(1 - \mu_A(x))\}dx,$$

where x_1 and x_2 are the left and right points of the support of the fuzzy set A, and $\mu_A(x)$ is the MF of the fuzzy set A. If $\mu_A(x) = \frac{1}{2}$ for all x on the support of A, then the fuzzy entropy of the fuzzy set A is the maximum. Assuming that two membership functions $\mu_A(x)$ and $\mu_B(x)$ are overlapping and for all $x \in [x_1, x_2]$, $\mu_A(x) + \mu_B(x) = 1$, then the fuzzy entropy can be simplified as:

$$d(A) = \int_{x_1}^{x_2}\{-\mu_A(x)\ ln(\mu_A(x))\}dx.$$

On the other hand, the authors define the measure for the deviation of an MF, $r(A)$, as the quantitative measure of the deviation of an MF from the symmetry and it is given as:

$$r(A) = \int_{x_1}^{x_2} \mu_C(x)\ ln\left(\frac{\mu_C(x)}{\mu_A(x)}\right)dx,$$

where $\mu_C(x)$ is the symmetrical MF associated with the fuzzy set A which has the same support as the fuzzy set A. Finally, the conciseness measure dr_{avr} is used to evaluate the shapes and allocations of N_m fuzzy sets $A_i(i = 1,\ldots,N_m)$ on the universe of discourse X of *x-axis*. They define the average conciseness measure as:

$$dr_{avr} = \frac{1}{N_m - 2}\sum_{i=2}^{N_m-1} dr(A_i), \quad \text{where} \quad dr(A) = d(A) + r(A) = -\int_{x_1}^{x_2} \mu_C(x)\ ln(\mu_A(x))dx,$$

$dr(A)$ is used to evaluate the shape and deviation of the membership function and N_m is the number of fuzzy sets in the universe of discourse. They use an MOEA considering the following two objectives: the accuracy of the model and the average conciseness, which has been explained previously.

Fazendeiro et al. in [24] propose an interpretability measure (J), for multi-objective optimization focusing on solving a control of a neuromuscular blockade problem. This index is obtained by means of the aggregation of three indices:

- The first index J_1 is intended to promote the natural localization of the linguistic term Zero.
- Index J_2 penalizes MFs with a poor distinguishability level.
- Index J_3 penalizes the low level of coverage of the universe of discourse.

Additionally, they use normal MFs to satisfy the normalization property. The three indices and the aggregated measures are defined as:

$$\begin{aligned}
&J_1 = K_1\ c_{ZE}^2,\\
&J_2 = K_2 \times \sum_x [(M_p(\zeta_x) - 1)^2 step(M_p(\zeta_x) - 1)],\\
&J_3 = K_3 \times \sum_x [(M_p(\zeta_x) - \epsilon)^2 step(\epsilon - M_p(\zeta_x))],\\
&J = \sum_i J_i,
\end{aligned}$$

where c_{ZE} denotes the center of the membership function of the linguistic term *Zero*, K_1, K_2 and K_3 are constants which allow the tuning of the relative weight of each J_i, the function *step* is the standard unit step function, ζ_x is a fuzzy set representing a real-valued sample x from the universe of discourse, $M_p(\zeta_x)$ is a sigma-count operator and ϵ is the minimum desired level of coverage. The sigma-count operator $M_p(\zeta_x)$ is defined as follows:

$$M_p(\zeta_x) = \sqrt[p]{l_1^p + \cdots + l_n^p},$$

where l_i $(i = 1,\ldots,n)$ is the membership degree of x in the ith linguistic term and p is a positive integer (in the experiments they used $p = 1$). The MOEA proposed optimizes the error and index J separately.

Pulkkinen et al. in [65] present a hybrid genetic fuzzy system [32], to be applied to a bioaerosol detector problem. They initialize the population of the MOEA using a decision tree (which implicitly includes features reduction) and include a simplification mechanism in order to reduce the number of rules (quadrant Q_1) and the number of rule conditions (quadrant Q_1). They use the following three heuristics in the evolutionary process in order to reduce the complexity:

- Remove all the rules with the same antecedent, except one rule selected randomly.
- The inconsistent rules can be rules of different lengths in which all conditions of the shorter rule(s) are present in the longer rule(s). If these inconsistent rules exist in the RB, they only preserve the longer rule.
- If a condition is present in all the rules, they propose the removal of this condition.

This mechanism for controlling the consistency of the RB, is used to prevent the RB from having inconsistent rules (quadrant Q_3). The authors first apply the C4.5 algorithm [68] in order to create a decision tree and this is then used to obtain a fuzzy classifier which helps to initialize the population of the evolutionary process. The C4.5 algorithm implicitly includes a mechanism for reducing the number of features (quadrant Q_2). Moreover, the authors present an MOEA that uses a discrete computation/approximation of the similarity measure (S), proposed in [69] (previously explained in Section 5), for merging fuzzy sets (quadrant Q_2). As stated above, this measure is a fuzzy relation that expresses the degree to which two fuzzy sets, A and B, are equal. If the similarity measure is greater than a given threshold (a suitable value for the threshold is 0.25), then they merge these two fuzzy sets (A and B) to generate a new one C. The merging method creates a common trapezoidal fuzzy set C that replaces the occurrence of the merged trapezoidal fuzzy sets A and B, defined as $\mu_A = (x; a_1, a_2, a_3, a_4)$ and $\mu_B = (x; b_1, b_2, b_3, b_4)$. The fuzzy set C is defined as $\mu_C = (x; c_1, c_2, c_3, c_4)$ where:

$$c_1 = min(a_1, b_1); \quad c_2 = \lambda_2 a_2 + (1-\lambda_2)b_2; \quad c_3 = \lambda_3 a_3 + (1-\lambda_3)b_3; \quad c_4 = max(a_4, b_4),$$

and the parameters $\lambda_2, \lambda_3 \in [0,1]$ determine which of the fuzzy sets A or B has the highest influence on the kernel of C.

Several metrics to guarantee the semantic interpretability of the MFs are presented by the authors in [65]. The proposed semantic interpretability metrics are:

1. Overlap penalty (P_{OL}): It is the length of overlapping for fuzzy sets,

$$P_{OL} = \frac{1}{n_s}\sum_{i=1}^{n_s}\frac{1}{N_{ov}^i}\sum_{j=1}^{N_{ov}^i}\frac{\lambda_{i,j}}{\chi_i} \quad \text{where} \quad N_{ov}^i = \binom{M_i}{2} = \frac{M_i!}{2(M_i-2)!},$$

where n_s is the number of variables selected from the n variables in the initialization, $\lambda_{i,j}$ is the length of the jth overlapping between two MFs for the input variable i, N_{ov}^i is the number of MF pairs in the input variable i, $M_i \geqslant 2$ is the number of active fuzzy sets in the input variable i (if there are only 2 MFs, $N_{ov}^i = 1$) and $\chi_i = ubound_i - lbound_i$ ($ubound_i$ and $lbound_i$ are, respectively, the upper and lower bounds of the ith variable). The overlap penalty is not calculated for a certain variable, in the case of the number of active MFs assigned to it being less than 2.

2. Discontinuity penalty (P_{DC}): It is the proportion of total length of discontinuity for two fuzzy sets.

$$P_{DC} = \frac{1}{n_s}\sum_{i=1}^{n_s}\sum_{j=1}^{G_i}\frac{\psi_{i,j}}{\chi_i}$$

where G_i is the number of discontinuities and $\psi_{i,j}$ is the length of the jth discontinuity in the input variable i.

3. Middle value penalty (P_{MV}): It is used to prevent the relaxed covering of MFs.

$$P_{MV} = \frac{1}{n_s}\sum_{i=1}^{n_s}\delta_i \quad \text{where} \quad \delta_i = \begin{cases} \frac{\delta_i^* - \alpha_L}{1-\alpha_L} & \text{if } \delta_i^* > \alpha_L \\ 0 & \text{if } \delta_i^* \leqslant \alpha_L \end{cases}$$

where α_L is the user specified maximum covering level ($\alpha_L = 0.1$) allowed for an MF in the center definition point of another MF, and δ_i^* is the maximum covering of other MFs at any center definition point of the MFs in variable i. This penalty is not calculated for a certain variable, in the case of the number of active MFs assigned to it being less than 2.

They define a transparency penalty (T) which is used to control distinguishability by including this measure T as an objective in the proposed MOEA together with the misclassification rate (as a metric of accuracy, true positive and false positive rates). T is defined as follows:

$$T = P_{OL} + P_{DC} + P_{MV}.$$

An MOEA to perform context adaptation is presented by Botta et al. in [11]. This algorithm considers the system error (MSE in regression problems) and an interpretability index ($\Phi_Q(P)$) to preserve the fuzzy ordering and good distinguishability by using scaling functions: Core-position modifier, core-width modifier, support-width modifier and generalized positively modifier (i.e., changing the degree of membership of the boundary elements of the fuzzy sets). This interpretability index considers a fuzzy partition, $P = A_1, \ldots, A_i, \ldots, A_N$, consisting of N fuzzy sets where $d_{j,i} = |j - i|$ is the semantic distance between A_j and A_i. For instance, the semantic distance between A_3 and A_1 is 2. The index is defined as:

$$\Phi_Q(P) = \frac{\sum_{1\leqslant i\leqslant N-1}\sum_{i<j\leqslant N}\frac{1}{d_{j,i}}\cdot \mu_Q^{d_{j,i}}(\overbrace{Q \leqslant (A_i, A_j)}^{x})}{\sum_{1\leqslant i\leqslant N-1}\sum_{i<j\leqslant N}\frac{1}{d_{j,i}}}$$

where Q is a fuzzy ordering index, $x = Q \leqslant (A_i, A_j)$ and $\mu_Q^{d_{j,i}}(x)$ are membership degrees of the values of Q defined on the universe $[0,1]$.

Gacto et al. in [28] propose a post-processing MOEA to improve the system accuracy while trying to maintain or even improve the interpretability to an acceptable level. This method includes a rule selection mechanism which is combined with a genetic tuning by using the number of rules (quadrant Q_1) as an objective to be minimized in order to reduce the model complexity. Because of this (different levels of difficulty in the objectives), it also proposes an enhanced algorithm extending the ideas of the MOEAs in [4,27]. Moreover, they proposed an index (namely *GM3M*) that helps preserve the semantic interpretability of linguistic fuzzy systems. This index is devoted to maintaining the original shape of the MFs while a tuning (or any kind of learning or improvement) of their definition parameters is performed, and it becomes the first relative measure of the quality of the linguistic fuzzy partitions, once we know how the most interpretable ones should be. Therefore, it can be used in two different ways. A possible way to measure the interpretability of the MFs is measuring it with regard to uniform strong fuzzy partitions (which usually satisfy absolute semantic constraints or absolute measures to the highest degree). On the other hand, since the concepts and their meaning strongly depend on the problem and person who makes the assessment (the final user), the initial linguistic fuzzy partitions could also be given by an expert.

GM3M is defined as the geometric mean of three metrics, and its values range between 0 (the lowest level of interpretability) and 1 (the highest level of interpretability). The index is defined as:

$$GM3M = \sqrt[3]{\delta \cdot \gamma \cdot \rho}$$

where δ, γ and ρ are three complementary metrics to measure interpretability when a tuning is performed on the MFs, i.e., when the MF definitions are changed, which is usually needed to reach an acceptable accuracy level. The geometric mean is used since in the case of only one of the metrics having very low values (causing low interpretability), small values of *GM3M* are also obtained. Each metric was proposed for working with triangular MFs but they can easily be extended with some small changes in the formulation to gaussian or trapezoidal MFs (see [28] for more details). The said metrics are: MFs displacement (δ), MFs lateral amplitude rate (γ) and MFs area similarity (ρ).

Let us represent the definition parameters of the original and the tuned MF *j* as (a_j,b_j,c_j) and $\left(a'_j, b'_j, c'_j\right)$, which can vary in their corresponding variation intervals $\left[I^l_{a_j}, I^r_{a_j}\right]$, $\left[I^l_{b_j}, I^r_{b_j}\right]$ and $\left[I^l_{c_j}, I^r_{c_j}\right]$, respectively. These intervals determine the maximum variation for each parameter and could be defined in a different way for different problems.

The δ metric can control the displacements in the central point of the MFs. It is based on computing the normalized distance between the central point of the tuned MF and the central point of the original MF, and it is calculated through obtaining the maximum displacement from all the MFs. For each MF_j in the linguistic fuzzy partition, we define $\delta_j = \left|b_j - b'_j\right|/I$, where $I = \left(I^r_{b_j} - I^l_{b_j}\right)/2$ represents the maximum variation for each central parameter. Thus δ^* is defined as $\delta^* = max_j\ \{\delta_j\}$ (the worst case). δ^* takes values between 0 and 1 (values near to 1 show that the MFs present a great displacement). The following transformation is made so that this metric represents proximity (maximization): Maximize$\delta = 1 - \delta^*$.

The γ metric is used to control the MF shapes. It is based on relating the left and right parts of the support of the original and the tuned MFs. Let us define $leftS_j = |a_j - b_j|$ as the amplitude of the left part of the original MF support and $rightS_j = |b_j - c_j|$ as the right part amplitude. Let us define $leftS'_j = \left|a'_j - b'_j\right|$ and $rightS'_j = \left|b'_j - c'_j\right|$ as the corresponding parts in the tuned MFs. γ_j is calculated using the following equation for each *MF*:

$$\gamma_j = \frac{min\{leftS_j/rightS_j, leftS'_j/rightS'_j\}}{max\left\{leftS_j/rightS_j, leftS'_j/rightS'_j\right\}}.$$

Values near to 1 mean that the left and right rate in the original MFs are highly maintained in the tuned MFs. Finally γ is calculated by obtaining the minimum value of γ_j (the worst case): Maximize$\gamma = min_j\{\gamma_j\}$.

The ρ metric is used to control the area of the MF shapes. It is based on relating the areas of the original and the tuned MFs. Let us define A_j as the area of the triangle representing the original MF_j, and A'_j as the new area. ρ_j is calculated using the following equation for each $MF : \rho_j = min\left\{A_j, A'_j\right\}/max\left\{A_j, A'_j\right\}$.

Values near to 1 mean that the original area and the tuned area of the MFs are more similar (less changes). The ρ metric is calculated by obtaining the minimum value of ρ_j (the worst case): Maximize$\rho = min_j\{\rho_j\}$.

They propose a particular MOEA for regression problems with three objectives that are optimized together considering both complexity and semantic interpretability at the same time: Minimization of the MSE, minimization of the number of rules (for which reason it is also included in the Q_1 quadrant) and maximization of the semantic interpretability index. This allows the selection of the most appropriate solution from the final Pareto front depending on the expert preferences. The authors include examples that show how the change in the MFs has been almost imperceptible but involving improvements of 30% in accuracy.

Finally, we want to reference two works here which have their main interests in other quadrants but are also using measures to quantify the fuzzy partition interpretability considering: Distinguishability [31] or Coverage [58] measures.

7. Summarizing the current state-of-the-art to assess the interpretability of linguistic FRBSs

In order to analyze the studied works chronologically, Table (2) shows a summary of the works that consider the interpretability for Linguistic FRBSs grouped by publication date: by years and within each year by alphabetical order.

The large quantity of works published in 2003 was motivated by the following two books [12,13] on the interpretability-accuracy trade-off in the field of FRBSs. However, apart from this, the interest of researchers has increased particularly from 2007, giving rise to the appearance of many works from this year to the present. The use of MOEAs has emerged as a good way to handle interpretability since they allow both complexity and semantic interpretability measures to be optimized together by also taking into account the accuracy of the model.

Taking into account the situation depicted in Table (2), most of the works consider measures included inside the quadrants Q_1 and Q_2 which are considered as the classic interpretability measures. The number of rules in the quadrant Q_1 is one of the more used measures in the literature, for which reason it is possible to consider it as a good measure of complexity at the rule base level. However the total number of conditions seems to be a more complete way since it can consider both the length of the rules and the number of rules, in an simple measure. In the quadrant Q_2 most of the works simply impose restrictions on the maximum number of MFs allowed, even though, depending on the problem (particularly in high dimensional problems) decreasing the number of features should be preferred. Inside quadrant Q_3 there are a few measures but some works propose promising measures such as the consistency of the rules and the more recent number of rules fired

Table 2
Summary of the current state-ofthe-art to assess the interpretability of linguistic FRBSs by years.

Authors	Refs.	Year	Type	Using MOEAs	Q_1		Q_2		Q_3			Q_4	
					NR	NC	NMF	NFEAT.	CONS.	RFIRED.	COIN.	CTR.	MEAS.
Ishibuchi et al.	[35,36,39]	1995, 1997, 1995	CLAS.	X, √, X	√								
Pedrycz and de Oliveira	[62]	1996	REG.										AM
de Oliveira	[59,60]	1999	CTL.				√						AM
Jin et al.	[44,45]	2003, 1999	CTL.						√			√	
Cheong and Lai	[15,16]	2000, 2003	CTL.							√		√	
Espinosa and Vandewalle	[23]	2000	REG.				√					√	
Jin	[43]	2000	CTL.		√		√		√			√	
Cordón et al.	[18,19]	2001	REG.		√		√						
Ishibuchi et al.	[37]	2001	CLAS.	√	√	√							
Suzuki et al.	[25,70,71]	2001, 2003	REG.	√									AM
Cordón et al.	[20]	2003	CLAS.	√		√	√						
Guillaume and Charnomordic	[30]	2003	CLAS.			√	√						
Ishibuchi and Yamamoto	[41]	2003	CLAS.		√								
Ishibuchi and Yamamoto	[40]	2003	REG.		√	√							
Nauck	[58]	2003	CLAS.			√	√						AM
Pedrycz	[61]	2003	REG.						√				
Pena-Reyes et al.	[63]	2003	CLAS.		√	√						√	
Tikk et al.	[73]	2003	CLAS.					√					
Vanhoucke and Silipo	[74]	2003	CLAS.					√					
Guillaume and Charnomordic	[31]	2004	CLAS.		√	√	√						AM
Ishibuchi and Yamamoto	[42]	2004	CLAS.	√	√	√							
Casillas et al.	[14]	2005	CLAS.		√								
Narukawa et al.	[57]	2005	CLAS.	√	√	√							
Mikut et al.	[55]	2005	CLAS.		√			√					
Alcalá et al.	[2]	2007	REG.		√		√						
Alcalá et al.	[1]	2007	REG. & CTL.		√								
Alcalá et al.	[4]	2007	REG.	√	√								
Cococcioni et al.	[17]	2007	REG.	√		√							
Fazendeiro et al.	[24]	2007	CTL.	√									AM
Ishibuchi and Nojima	[38]	2007	CLAS.	√	√	√							
Liu et al.	[47]	2007	CLAS.		√				√				
Mencar et al.	[52]	2007	CLAS.									√	
Alonso et al.	[9]	2008	CLAS.		√	√			√				
Pulkkinen et al.	[65]	2008	CLAS.	√	√	√	√		√				AM
Pulkkinen and Koivisto	[66]	2008	CLAS.	√	√	√	√						
Alcalá et al.	[3]	2009	REG.	√	√								
Alonso et al.	[10]	2009	CLAS.		√	√			√				
Botta et al.	[11]	2009	REG.	√									AM
Gacto et al.	[27]	2009	REG.	√	√								
Pulkkinen et al.	[64,67]	2009,2010	CLAS. & REG.	√		√						√	
Alonso et al.	[6]	2010	CLAS.	√		√				√			
Alonso and Magdalena	[8]	in press	CLAS.		√	√		√	√				
Gacto et al.	[28]	2010	REG.	√	√								RM
Márquez et al.	[51]	2010	REG.	√	√					√			
Mencar et al.	[53]	in press	CLAS.								√	√	

NR = Number of rules, NC = Number of conditions, NMF = Number of membership functions, NFEAT. = Number of features, CONS. = Consistency, RFIRED. = Number of rules fired at the same time, COIN = Cointension, CTR. = Constraints, MEAS. = Measures; CLAS. = Classification, REG. = Regression, CTL. = Control, AM = Absolute measures, RM = Relative measures.

at the same time and co-intension. In quadrant Q_4 there are a lot of works imposing constraints. However, recently new absolute or relative semantic interpretability measures have arisen, which are more suitable to be taken into account for optimization processes. For Q_3 and Q_4 quadrants, there are still no widely accepted measures, which will arise with their use as happened with the number of rules for Q_1.

8. Conclusions

In this contribution, we have presented a review of the interpretability of fuzzy systems focused on the framework of linguistic FRBSs. A complete review of works on the use/proposal of techniques or measures to take into account the interpretability of linguistic FRBSs, as a contribution towards finding a good trade-off between interpretability and accuracy, has been carried out in this paper. To this end, we have proposed a taxonomy with four quadrants (complexity or semantic interpretability at the level of RB or fuzzy partitions) as a way of organizing the different measures or constraints that we find in the literature to control interpretability in this area. We have analyzed the different measures proposed in the different quadrants. Since the interpretability of linguistic FRBSs is still an open problem, this will help researchers in this field determine the most appropriate measure depending on the part of the KB in which they want to maintain/improve interpretability.

After studying the different works on the said topic, we can state that there is no single comprehensive measure to quantify the interpretability of linguistic models. In our opinion, to get a good global measure it would be necessary to consider appropriate measures from all of the quadrants, in order to take into account the different interpretability properties required for these kinds of systems together. The different measures from each quadrant could be optimized as different objectives within a multi-objective framework. This would allow a search for a compromise among the different measures to take place taking accuracy into account. The main problem is that multi-objective optimization algorithms are currently unable to adequately handle much more than three objectives. Therefore, it is also necessary to find a way to combine them into a single index using weights or appropriate aggregation operators in order to give appropriate importance to one or another measure. One possibility is to aggregate complexity-based and sematic-based measures separately, producing two different indexes. This would allow the different trade-offs among accuracy, complexity and semantic interpretability to be arrived at. In this sense, and taking into account the studied works in the different quadrants, we can make the following statements:

- In quadrants Q_1 and Q_2, there are well-known and much used measures to quantify complexity. These measures are widely accepted as the number of rules, number of conditions and number of features. Moreover, these measures are easy to use in practice since, for example, measuring the total number of conditions is a way to also take into account the remaining ones (such as, fewer conditions in a model, smaller number of rules and/or smaller number of features).
- By contrast, there is no agreement regarding the choice of an appropriate measure in the Q_3 and Q_4 quadrants. Nevertheless, we can find two interesting ways of working for each quadrant respectively:
  - On one hand, we consider that the use of the number of rules fired at the same time [6,15,16,51] is a very promising measure or a way of working for Q_3 if it is properly combined or adapted in order to also consider the consistency of the rules.
  - On the other hand, the use of a relative measure, such as the one in [28], defined as a global semantic interpretability index to quantify interpretability with respect to the preferred reference fuzzy partitions provided by an expert (when this is possible), could represent a promising alternative for measuring/comparing the linguistic models in Q_4 since it can take into account the contribution of an expert.

An important aspect that could condition the possibility of using a relative measure for Q_4 is the existence or not of an expert for the problem being solved and the possibility/ability of this expert to define a reference fuzzy partition. It could be argued that for partitions fully generated from data there is no expert or user that could assess the derivation of reference fuzzy partitions. This is true for benchmark problems (which are required for and used to assess the good performance of the proposed techniques) but in a real problem, even if we are going to learn from data, there should be an expert/user/client (the one that needs to understand the final model) that should be able to determine the concepts that he can understand (even if it is by simply using strong fuzzy partitions centered on the concept's modal values). If he requires this kind of linguistic interpretability but is unable to specify the linguistic terms, it would make no sense to apply a linguistic approach since there are some good alternatives that could obtain better accuracy (and this is essential to understand the final linguistic model). In any event, it will not be an easy task since for example the number of features could be too much high. However, it should be at least possible to define some previous simple models to fix a reasonable number of relevant features and/or even to provide some automatically generated transparent fuzzy partitions. This way of working can facilitate this non trivial task for the expert (these are the usual steps when solving a real application for an end user, including some initial trial and error steps).

This alternative approach can even become complementary to the classic alternative of providing some automatically generated transparent fuzzy partitions (including feature selection, granularity specification, transparency measures, etc.), which become the only alternative when it is not possible to obtain the linguistic definitions from an expert or a good alternative when the expert is able to accept or wants to learn new concepts (i.e., he has no strongly pre-established or fixed knowledge and the automatic learning of transparent partitions could even provide new unknown concepts extending the expert's knowledge). In any case, this way of working (automatic definition by absolute measures) also requires some trial and error steps by generating different alternatives that should be analyzed by an expert in order to validate the appropriate fuzzy partitions. Once the final fuzzy partition is accepted/validated by the expert, using an appropriate relative measure can become a complementary approach that could help to assess the semantic interpretability of different linguistic KB learners/optimizers based on the automatically obtained and validated definitions.

On the other hand, an interesting way to analyze how we could combine the measures of the different quadrants (which is still a completely open problem) is to pay attention to what users and/or experts would consider interpretable; for instance, by using a web poll as in [10]. Alonso et al. in [10] propose the use of the knowledge extracted by means of a web poll of researchers familiarized and not familiarized with LFM. This poll is dedicated to determining how different people assess interpretability, giving priority to different criteria. Moreover, in [7] they emphasize the need for combining several indices in order to get a measure incorporating preferences of the user in a flexible and efficient manner. Because of it, the user should be able to personalize the index according to his/her preferences.

Consequently, we must point out that it is necessary to establish the measures for the different quadrants and, in respect to the aggregation of the different measures in a global index, the way in which the measures are combined by selecting the appropriate aggregation operators is not a trivial but an essential task [7].

Finally, another important question is the development of more effective learning or tuning algorithms able to obtain knowledge bases with the best trade-offs. This remains a key issue since considering all the aspects involved in the case of interpretability in an optimum manner is an important task when this is combined with accuracy measures. Designing appropriate algorithms able to handle the increasing number of interpretability objectives/measures, which involve different difficulties, is still an open topic which should be addressed within the specific framework of the interpretability-accuracy trade-off.

References

[1] R. Alcalá, J. Alcalá-Fdez, F. Herrera, A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection, IEEE Transactions on Fuzzy Systems 15 (2007) 616–635.

[2] R. Alcalá, J. Alcalá-Fdez, F. Herrera, J. Otero, Genetic learning of accurate and compact fuzzy rule based systems based on the 2-tuples linguistic representation, International Journal of Approximate Reasoning 44 (2007) 45–64.

[3] R. Alcalá, P. Ducange, F. Herrera, B. Lazzerini, F. Marcelloni, A multi-objective evolutionary approach to concurrently learn rule and data bases of linguistic fuzzy rule-based systems, IEEE Transactions on Fuzzy Systems 17 (2009) 1106–1122.

[4] R. Alcalá, M.J. Gacto, F. Herrera, J. Alcalá-Fdez, 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 (2007) 539–557.

[5] R. Alcalá, Y. Nojima, F. Herrera, H. Ishibuchi, Multiobjective genetic fuzzy rule selection of single granularity-based fuzzy classification rules and its interaction with the lateral tuning of membership functions, Soft Computing, in press. Available from: doi:10.1007/s00500-010-0671-2.

[6] J. Alonso, L. Magdalena, O. Cordón, Embedding HILK in a three-objective evolutionary algorithm with the aim of modeling highly interpretable fuzzy rule-based classifiers, in: 4th International Workshop on Genetic and Evolving Fuzzy Systems (GEFS2010), IEEE, Mieres, Spain, 2010, pp. 15–20.

[7] J.M. Alonso, L. Magdalena, Combining user's preference and quality criteria into a new index for guiding the design of fuzzy systems with a good interpretability-accuracy trade-off, in: IEEE World Congress on Computational Intelligence, 2010, pp. 961–968.

[8] J.M. Alonso, L. Magdalena, HILK++: an interpretability-guided fuzzy modeling methodology for learning readable and comprehensible fuzzy rule-based classifiers, Soft Computing, in press. Available from: doi:10.1007/s00500-010-0628-5.

[9] J.M. Alonso, L. Magdalena, S. Guillaume, HILK: a new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism, International Journal of Intelligent Systems 23 (2008) 761–794.

[10] J.M. Alonso, L. Magdalena, G.G. Rodríguez, Looking for a good fuzzy system interpretability index: an experimental approach, International Journal of Approximate Reasoning 51 (2009) 115–134.

[11] A. Botta, B. Lazzerini, F. Marcelloni, D.C. Stefanescu, Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index, Soft Computing 13 (2009) 437–449.

[12] J. Casillas, O. Cordón, F. Herrera, L. Magdalena, Accuracy improvements in linguistic fuzzy modeling, Studies in Fuzziness and Soft Computing, vol. 129, Springer, 2003.

[13] J. Casillas, O. Cordón, F. Herrera, L. Magdalena, Studies in Fuzziness and Soft Computing, vol. 128, Springer, 2003.

[14] J. Casillas, O. Cordón, M.J. del Jesus, F. Herrera, Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction, IEEE Transactions on Fuzzy Systems 13 (2005) 13–29.

[15] F. Cheong, R. Lai, Constraining the optimization of a fuzzy logic controller using an enhanced genetic algorithm, IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics 30 (2000) 31–46.

[16] F. Cheong, R. Lai, Constrained optimization of genetic fuzzy systems, J. Casillas, O. Cordón, F. Herrera, L. Magdalena (Eds.), Accuracy Improvements in Linguistic Fuzzy Modeling, Springer-Verlag, 2003, pp. 46–71.

[17] M. Cococcioni, P. Ducange, B. Lazzerini, F. Marcelloni, A pareto-based multi-objective evolutionary approach to the identification of Mamdani fuzzy systems, Soft Computing 11 (2007) 1013–1031.

[18] O. Cordón, F. Herrera, L. Magdalena, P. Villar, A genetic learning process for the scaling factors granularity and contexts of the fuzzy rule-based system data base, Information Science 136 (2001) 85–107.

[19] O. Cordón, F. Herrera, P. Villar, Generating the knowledge base of a fuzzy rule-based system by the genetic learning of data base, IEEE Transactions on Fuzzy Systems 9 (2001) 667–674.

[20] O. Cordón, M. del Jesus, F. Herrera, L. Magdalena, P. Villar, A multiobjective genetic learning process for joint feature selection and granularity and context learning in fuzzy rule-based classification systems, in: J. Casillas, O. Cordón, F. Herrera, L. Magdalena (Eds.), Interpretability Issues in Fuzzy Modeling, Springer-Verlag, 2003, pp. 79–99.

[21] K. Deb, A. Pratab, S. Agrawal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation 6 (2002) 182–197.

[22] P.A. Devijver, J. Kittler, Pattern Recognition: A Statistical Approach, Prentice Hall, London, 1982.

[23] J. Espinosa, J. Vandewalle, Constructing fuzzy models with linguistic integrity from numerical data-AFRELI algorithm, IEEE Transactions on Fuzzy Systems 8 (2000) 591–600.

[24] P. Fazendeiro, J.V. de Oliveira, W. Pedrycz, A multiobjective design of a patient and anaesthetist-friendly neuromuscular blockade controller, IEEE Transactions on Biomedical Engineering 54 (2007) 1667–1678.

[25] T. Furuhashi, T. Suzuki, On interpretability of fuzzy models based on conciseness measure, in: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'01), 2001, pp. 284–287.

[26] M. Gacto, R. Alcalá, F. Herrera, A multi-objective evolutionary algorithm for an effective tuning of fuzzy logic controllers in heating, ventilating and air conditioning systems, Applied Intelligence, in press. Available from: doi:10.1007/s10489-010-0264-x.

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

[28] M.J. Gacto, R. Alcalá, F. Herrera, 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 (2010) 515–531.

[29] S. Guillaume, Designing fuzzy inference systems from data: an interpretability-oriented review, IEEE Transactions on Fuzzy Systems 9 (2001) 426–443.

[30] S. Guillaume, B. Charnomordic, A new method for inducing a set of interpretable fuzzy partitions and fuzzy inference systems from data, in: J. Casillas, O. Cordón, F. Herrera, L. Magdalena (Eds.), Interpretability Issues in Fuzzy Modeling, Springer-Verlag, 2003, pp. 148–175.

[31] S. Guillaume, B. Charnomordic, Generating an interpretable family of fuzzy partitions from data, IEEE Transactions on Fuzzy Systems 12 (2004) 324–335.

[32] F. Herrera, Genetic fuzzy systems: taxonomy current research trends and prospects, Evolutionary Intelligence 1 (2008) 27–46.

[33] F. Herrera, L. Martínez, A 2-tuple fuzzy linguistic representation model for computing with words, IEEE Transactions on Fuzzy Systems 8 (2000) 746–752.

[34] H. Ishibuchi, T. Murata, Multi-objective genetic local search algorithm, in: Proceedings of Third IEEE International Conference on Evolutionary Computation, Japan, 1996, pp. 119–124.

[35] H. Ishibuchi, T. Murata, I.B. Türksen, Selecting linguistic classification rules by two-objective genetic algorithms, in: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Vancouver, Canada, 1995, pp. 1410–1415.

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

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

[38] H. Ishibuchi, Y. Nojima, Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning, International Journal of Approximate Reasoning 44 (2007) 4–31.

[39] H. Ishibuchi, K. Nozaki, N. Yamamoto, H. Tanaka, Selecting fuzzy ifthen rules for classification problems using genetic algorithms, IEEE Transactions on Fuzzy Systems 3 (1995) 260–270.

[40] H. Ishibuchi, T. Yamamoto, Interpretability issues in fuzzy genetics-based machine learning for linguistic modelling, in: J. Lawry, J.G. Shanahan, A.L. Ralescu (Eds.), Modelling With Words: Learning Fusion, and Reasoning within a Formal Liguistic Representation Framework, Lecture Notes in Computer Science, vol. 2873, Springer-Verlag, Berlin, Heidelberg, 2003, pp. 209–228.

[41] H. Ishibuchi, T. Yamamoto, Trade-off between the number of fuzzy rules and their classification performance, J. Casillas, O. Cordón, F. Herrera, L. Magdalena (Eds.), Accuracy Improvements in Linguistic Fuzzy Modeling, Springer-Verlag, 2003, pp. 72–99.

[42] H. Ishibuchi, T. Yamamoto, Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining, Fuzzy Sets and Systems 141 (2004) 59–88.

[43] Y. Jin, Fuzzy modeling of high-dimensional systems: Complexity reduction and interpretability improvement, IEEE Transactions on Fuzzy Systems 8 (2000) 212–221.

[44] Y. Jin, Generating distinguishable, complete consistent and compact fuzzy systems using evolutionary algorithms, J. Casillas, O. Cordón, F. Herrera, L. Magdalena (Eds.), Accuracy Improvements in Linguistic Fuzzy Modeling, Springer-Verlag, 2003, pp. 100–118.

[45] Y. Jin, W. von Seelen, B. Sendhoff, On generating FC^3 fuzzy rule systems from data using evolution strategies, IEEE Transactions on Systems Man and Cybernetics 29 (1999) 829–845.

[46] J.D. Knowles, D.W. Corne, Approximating the non dominated front using the pareto archived evolution strategy, Evolutionary Computation 8 (2000) 149–172.

[47] F. Liu, C. Quek, G.S. Ng, A novel generic hebbian ordering-based fuzzy rule base reduction approach to Mamdani neuro-fuzzy system, Neural Computation 19 (2007) 1656–1680.

[48] A. de Luca, S. Termini, A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory, Information and Control 20 (1972) 301–312.

[49] E.H. Mamdani, Application of fuzzy algorithms for control of simple dynamic plant, in: Proceedings of IEEE, vol. 121, 1974, pp. 1585–1588.

[50] E.H. Mamdani, S. Assilian, An experiment in linguistic synthesis with a fuzzy logic controller, International Journal of Man-Machine Studies 7 (1975) 1–13.

[51] A. Marquez, F. Márquez, A. Peregrin, A multi-objective evolutionary algorithm with an interpretability improvement mechanism for linguistic fuzzy systems with adaptive defuzzification, in: IEEE World Congress on Computational Intelligence, 2010, pp. 277–283.

[52] C. Mencar, G. Castellano, A.M. Fanelli, Distinguishability quantification of fuzzy sets, Information Sciences 177 (2007) 130–149.

[53] C. Mencar, C. Castiello, R. Cannone, A.M. Fanelli, Interpretability assessment of fuzzy knowledge bases: a cointension based approach, International Journal of Approximate Reasoning, in press. Available from: doi:10.1016/j.ijar.2010.11.007.

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

[55] R. Mikut, J. Jakel, L. Grall, Interpretability issues in data-based learning of fuzzy systems, Fuzzy Sets and Systems 150 (2005) 179–197.

[56] G.A. Miller, The magical number seven plus or minus two: some limits on our capacity for processing information, The Psychological Review 63 (1956) 81–97.

[57] K. Narukawa, Y. Nojima, H. Ishibuchi, Modification of evolutionary multiobjective optimization algorithms for multiobjective design of fuzzy rule-based classification systems, in: Proceedings of the 2005 IEEE International Conference on Fuzzy Systems, Reno, USA, 2005, pp. 809–814.

[58] D. Nauck, Measuring interpretability in rule-based classification systems, in: Proceedings of the 12th IEEE International Conference on Fuzzy Systems, vol. 1, 2003, pp. 196–201.

[59] J.V. de Oliveira, Semantic constraints for membership function optimization, IEEE Transactions Systems, Man, and Cybernitics – Part A: Systems and Humans 29 (1999) 128–138.

[60] J.V. de Oliveira, Towards neuro-linguistic modeling: constraints for optimization of membership functions, Fuzzy Sets and Systems 106 (1999) 357–380.

[61] W. Pedrycz, Expressing relevance interpretability and accuracy of rule-based systems, in: J. Casillas, O. Cordón, F. Herrera, L. Magdalena (Eds.), Interpretability issues in fuzzy modeling, Springer-Verlag, 2003, pp. 546–567.

[62] W. Pedrycz, J.V. de Oliveira, Optimization of fuzzy models, IEEE Transactions Systems, Man, and Cybernetics Part B 26 (1996) 627–636.

[63] C.A. Peña-Reyes, M. Sipper, Fuzzy CoCo: Balancing accuracy and interpretability of fuzzy models by means of coevolution, in: J. Casillas, O. Cordón, F. Herrera, L. Magdalena (Eds.), Accuracy Improvements in Linguistic Fuzzy Modeling, Springer-Verlag, 2003, pp. 119–146.

[64] P. Pulkkinen, A multiobjective genetic fuzzy system for obtaining compact and accurate fuzzy classifiers with transparent fuzzy partitions, in: Proceedings of the 8th International Conference Machine Learning and Applications, Miami Beach, FL, 2009, pp. 89–94.

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

[66] P. Pulkkinen, H. Koivisto, Fuzzy classifier identification using decision tree and multiobjective evolutionary algorithms, International Journal of Approximate Reasoning 48 (2008) 526–543.

[67] P. Pulkkinen, H. Koivisto, A dynamically constrained multiobjective genetic fuzzy system for regression problems, IEEE Transactions on Fuzzy Systems 18 (2010) 161–177.

[68] J.R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufman, 1993.

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

[70] T. Suzuki, T. Furuhashi, Conciseness of fuzzy models, in: J. Casillas, O. Cordón, F. Herrera, L. Magdalena (Eds.), Interpretability issues in fuzzy modeling, Springer-Verlag, 2003, pp. 568–586.

[71] T. Suzuki, T. Kodama, T. Furuhashi, H. Tsutsui, Fuzzy modeling using genetic algorithms with fuzzy entropy as conciseness measure, Information Sciences 136 (2001) 53–67.

[72] T. Takagi, M. Sugeno, Fuzzy identification of systems and its application to modeling and control, IEEE Transactions on Systems, Man, and Cybernetics 15 (1985) 116–132.

[73] D. Tikk, T. Gedeon, K. Wong, A feature ranking algorithm for fuzzy modelling problems, in: J. Casillas, O. Cordón, F. Herrera, L. Magdalena (Eds.), Interpretability Issues in Fuzzy Modeling, Springer-Verlag, 2003, pp. 176–192.

[74] V. Vanhoucke, R. Silipo, Interpretability in multidimensional classification, in: J. Casillas, O. Cordón, F. Herrera, L. Magdalena (Eds.), Interpretability issues in fuzzy modeling, Springer-Verlag, 2003, pp. 193–217.

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