A Conceptual Framework for Understanding a Fuzzy System

José M. Alonso, Luis Magdalena

European Centre for Soft Computing, Edificio Científico-Tecnológico,
Gonzalo Gutiérrez Quirós s/n, 33600 Mieres, Asturias, Spain
Email: {jose.alonso,luis.magdalena}@softcomputing.es

Abstract— *The word Interpretability is becoming more and more frequent in the fuzzy literature. It is admitted as the main advantage of fuzzy systems and it should be given a main role in fuzzy modeling. However, although researchers talk a lot about Interpretability, it is even not clear what it really means. Understanding of fuzzy systems is a subjective task which strongly depends on the person (experience, preferences, knowledge, etc.) who makes the assessment. The general context and the specific problem under consideration have a huge influence too. This paper makes a review on works related to Interpretability and presents the proposal of a conceptual framework that can help to understand fuzzy systems. Moreover, it can be considered as a starting point in order to propose a fuzzy index, easily adaptable to the context of each problem as well as to the user quality criteria, for measuring Interpretability.*

Keywords— Fuzzy systems, Interpretability definition and measurement.

1 Introduction

The concept of interpretability appears in many fields (education, medicine, computer science, etc.) under several names like understandability, comprehensibility, intelligibility, transparency, readability, etc. All these terms are usually considered as synonymous what could yield some confusion. However, some authors [1] distinguish between the term "transparency" (readability) referred as an inherent systemic property (related to the view of the model structure as a white-box) and the term "interpretability" (comprehensibility) which has more cognitive aspects because it is always related to human beings, or more specifically to *humanistic systems* (defined by Zadeh as *those systems whose behavior is strongly influenced by human judgment, perception or emotions* [2]). Notice that readability is assumed as a prerequisite for comprehensibility.

Understanding is likely to be one of the most valuable human abilities. Of course, it is related to the human intelligence and the natural language processing capabilities, because human reasoning is mainly supported by language. The most usual way of explaining something to someone is through the use of words, sentences, linguistic expressions, etc. Of course, gestures and symbols are also used as additional communication tools but they only represent other kinds of languages. Unfortunately, knowledge about these kinds of cognitive tasks is still quite reduced. However, let us underline that this work belongs to the field of artificial intelligence and it will focus on analyzing the interpretability of knowledge-based systems, and more specifically of fuzzy rule-based systems (FRBSs). The main goal of this work is to study how comprehensible are such systems from a human point of view, opening a constructive discussion.

The use of linguistic variables [2] to overcome the ineffectiveness of computers in dealing with systems whose behavior is strongly influenced by human judgment, perceptions or emotions was pointed out by Zadeh long time ago: *In order to be able to make significant assertions (...) it may be necessary to abandon the high standards of rigor and precision that we have become conditioned to expect of our mathematical analyses (...) and become more tolerant of approaches which are approximate in nature* [2]. Following the Zadeh's advice if we really want to define a useful index for system modeling, it is necessary to change our mind. Numerical indices should be forgotten and in turn fuzzy indices should be defined, i.e., the focus must be shifted *from computing with numbers to computing with words, from manipulation of measurements to manipulation of perceptions* [3]. In consequence, the right approach to assess interpretability in an effective way consists in proposing a fuzzy index instead of a numerical one. A first attempt was presented in [4] where a hierarchical fuzzy system was used to get an interpretability measure. That proposal opened this way but a lot of work remains to do.

The expressivity of linguistic rules [5] is acknowledged to be quite close to natural language what favors the interpretability because human understanding is made in terms of natural language. That is why it is useful to take into account the experience gained by natural language processing researchers. For instance, the philosopher Paul Grice established the next four conversational maxims [6] which arise from the pragmatics of natural language and they are based on the common sense:

1. *Maxim of Quality*: Do not say what you believe to be false. Do not say anything without adequate evidence.
2. *Maxim of Quantity*: Make your contribution as informative as required for the current purposes of the exchange.
3. *Maxim of Relation*: Be relevant.
4. *Maxim of Manner*: Avoid obscurity of expression. Avoid ambiguity. Be brief. Be orderly.

Keeping the Grice's maxims in mind during the fuzzy modeling process can help to make easier the understanding of FRBSs. The rule base must be coherent avoiding the use of inconsistent rules (*Maxim of Quality*), redundant rules (*Maxim of Quantity*), and ambiguity rules (*Maxim of Manner*). In addition, selecting the most relevant rules (*Maxim of Relation*) will yield more compact and robust systems.

The rest of the paper is structured as follows. Section 2 has a look on definitions of interpretability found in the literature. In addition, it makes a global review on all the aspects that should be taken into account in the interpretability

ISBN: 978-989-95079-6-8

assessment, setting a conceptual framework for characterizing interpretability. Section 3 describes how to combine the main factors included in the proposed framework for measuring interpretability of FRBSs. Finally, section 4 offers some conclusions and points out future works.

2 Understanding a fuzzy system

Authors talk a lot about interpretability but it is not easy to find a formal definition in the literature. Thus, it is necessary to think on the following question: How can interpretability be defined? The first bid to set a formal definition was made by Tarski et al. [7] a very long time ago (in 1953). He formulated a mathematical definition in the context of classical logic, setting the basis for identifying interpretable theories. In short, *assuming T and S are formal theories, T is interpretable in S if and only if there is a way to pass from T to S, assuring that every theorem of T can be translated and proved into S*.

Regarding the fuzzy literature, a similar definition is included as part of the formal framework proposed in [8]. It distinguishes between a formal language *L* (fuzzy logic) used for describing the model under consideration, and a user-oriented language *L'* (usually the natural language) used for explaining the model to the user. If the system is interpretable, the translation from *L* to *L'* should be made by the user with a small effort. In an informal way, people say that a model is interpretable if they are able to describe it easily.

A more formal definition was given by Bodenhofer and Bauer [9]: *Interpretability means possibility to estimate the system's behavior by reading and understanding the rule base only*. Since the rule base understanding strongly depends on the readability of the involved linguistic expressions, the authors focused on analyzing the interpretability at the level of fuzzy partitioning (linguistic variables) from an intuitive and mathematically exact point of view: *The obvious orderings and inclusions of linguistic terms must not be violated by the corresponding fuzzy sets*. As a result, fuzzy partitioning readability was assumed to be a prerequisite to build interpretable FRBSs.

The comprehensibility of a FRBS depends on all its components, i.e., it depends on the knowledge base (KB) transparency but also on the inference mechanism understanding. Previous works [10, 11] have thoroughly analyzed the main factors that influence the KB readability. Also, a complete study on the interpretability constraints most frequently used in fuzzy modeling has been recently published [1].

Fig. 1 describes the main factors to be considered regarding interpretability of FRBSs. It is inspired on the taxonomy of interpretability of fuzzy systems introduced by [12], which is extended adding our own notation and concepts, and also including some of the most significant constraints extracted from [1]. There are two main points of view to be considered when assessing interpretability of FRBSs (*Global description* and *Local explanation*). The global view presents the system as a whole explaining its global behavior and trend. However, the local view focuses on each individual situation, explaining specific behaviors for specific events. For instance, if we had a fuzzy controller for driving a car, the global view would give an idea on the kind of operations it can do (go straight forward, turn on the right/left, speed up, brake, etc.) and even on the driver style (aggressive, sluggish, etc.). On the contrary, the local view would explain each specific manoeuvre.

Additional information is detailed in the following subsections. Pay attention to the fact that both viewpoints could lead to contradictory goals. The first one (*Global description*) prefers rules as compact as possible, while the second one (*Local explanation*) favors the use of complete rules (*The more general rules, the larger the number of rules that can be fired at the same time*).

2.1 Global description (system structure)

In order to assess the simplicity of a FRBS the following assumption is made: *The more compact the KB, the simpler its understanding, i.e., the higher the interpretability*. This reasoning follows the principle of incompatibility formulated by Zadeh [13] in 1973: *As the complexity of a system increases, our ability to make precise and yet significant statements about its behavior diminishes until a threshold is reached beyond which precision and significance become almost mutually exclusive characteristics. The closer one looks at a real-world problem, the fuzzier becomes its solution*.

The global description of a linguistic FRBS can be analyzed looking at different abstraction levels as illustrated on left part of Fig. 1. First, the lowest level corresponds to the level of individual fuzzy sets. It includes those constraints demanded to build interpretable fuzzy sets, regarding mathematical properties of the membership functions. At the second level, there are several constraints with respect to the combination of several fuzzy sets to form a fuzzy partition. The use of linguistic variables favors the readability, but it is not enough to ensure interpretability. Hence, some linguistic constraints must be superimposed to the fuzzy partition definition to be interpretable. Fortunately, Ruspini defined (in 1969) a special kind of partition called Strong Fuzzy Partition (SFP) [14] that satisfies most demanded semantic constraints (distinguishability, coverage, normality, convexity, etc). In practice, satisfying all constraints is almost impossible and useless because they represent a very restrictive set of conditions that usually yield systems with very small accuracy. Notice that looking for a good accuracy-interpretability trade-off is the most complex task of fuzzy modeling. Especially relevant are some recent successful biomedical applications [15].

Once a set of linguistic terms with their associated semantics has been defined, they can be used to express linguistic propositions. Then, several propositions are combined to form fuzzy rules describing the system behavior. However, in addition to the analysis of each individual rule it is needed to study the combination of several rules, achieving the highest abstraction level. Notice that defining a global semantics previous to the rule definition makes easier the rule understanding. Only if all the rules use the same linguistic terms (defined by the same fuzzy sets) it will be possible to make a rule comparison at the linguistic level. In order to get fully meaningful partitions the right linguistic terms should be selected according to the problem context. Nevertheless, matching linguistic terms and fuzzy sets is not a straightforward task, for instance finding good linguistic terms for fuzzy partitions automatically generated from data is sometimes not feasible.

To sum up, the satisfaction of all constraints enumerated on the left part of Fig. 1 guarantees the interpretability of the FRBS from the structural point of view.

ISBN: 978-989-95079-6-8

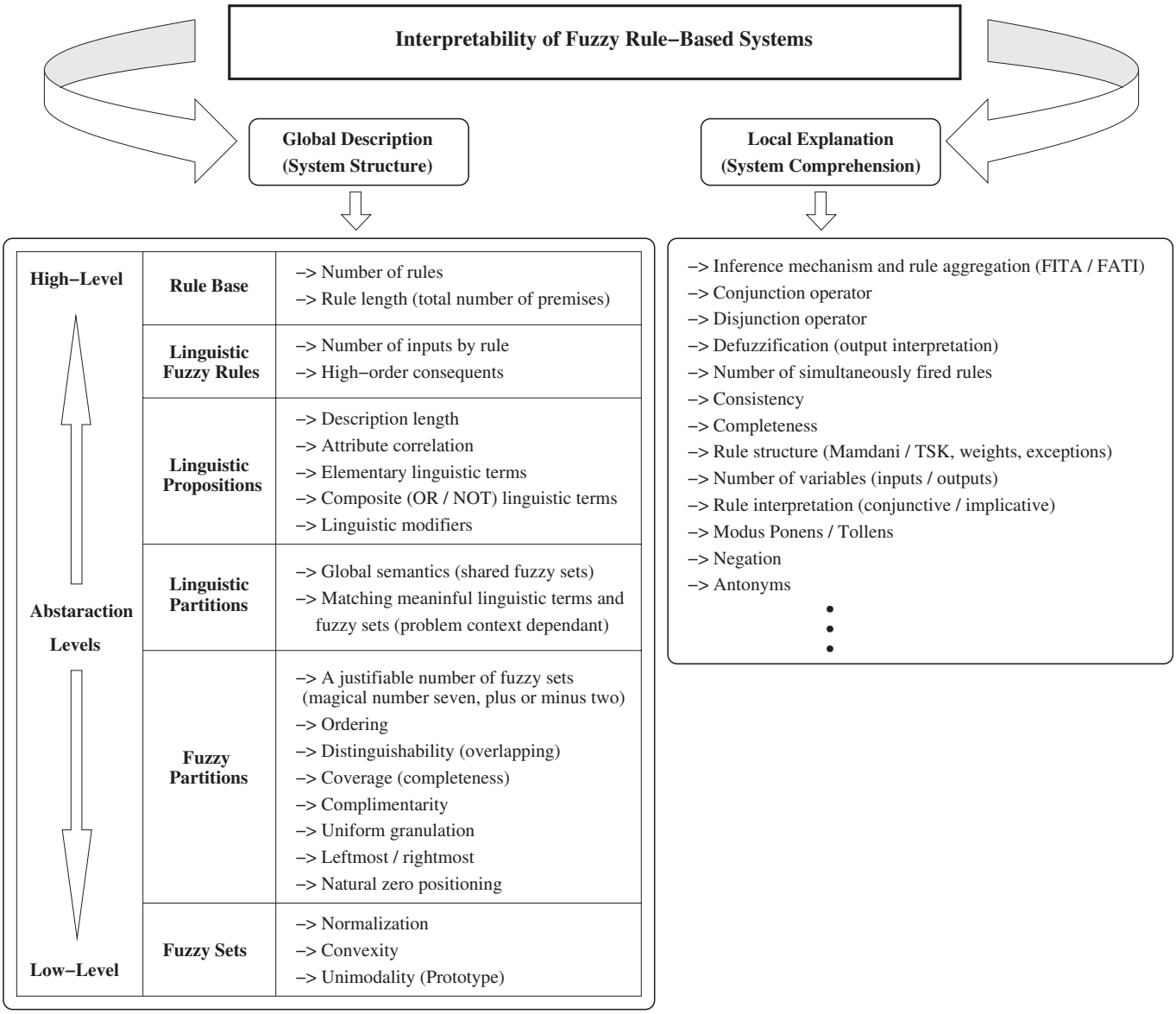

Figure 1: A conceptual framework for characterizing interpretability of FRBSs.

2.2 Local explanation (system comprehension)

Understanding the system behavior from its linguistic description is a very hard task that involves the inference level going beyond the former analysis of the system structure.

In addition, there is a need to give some comments about the inference mechanism implementation distinguishing between FITA (First Infer Then Aggregate) and FATI (First Aggregate Then Infer). It includes the fuzzy operator definitions for conjunction, disjunction, aggregation, and defuzzification. Furthermore, taking into account that as the result of a fuzzy inference several rules can be fired at the same time for a given input vector, the interpretability strongly depends on the number of rules that can be simultaneously fired. The smaller that value, the higher the interpretability. In fact, a model made up of thousand rules (where at maximum ten rules are fired together) may be seen as more interpretable than a model including only one hundred rules (where most of them are simultaneously fired). Notice that the whole rule base should be consistent (not including redundancies, contradictions, etc.) and it should cover most possible situations.

Although Mamdani rules are widely admitted as the more interpretable kind of rules, there are many other rule formats. The second most used rules are the well-known Takagi-Sugeno rules, but there are also rules with exceptions, rules with weights, and so on. The different rule formats can be compared and it is possible to discuss which one is better regarding interpretability from a structural point of view but it is a controversial issue. For instance, for many people the most interpretable rule format is the one they usually work with disregarding its complexity. This proves that many psychological aspects make influence when assessing interpretability. It is a clear example of the "Hammer principle" formulated by Zadeh [16]: *When the only tool you have is a hammer, everything begins to look like a nail*.

Finally, modus Ponens/Tollens must be carefully taken into account. Notice that the fact the all rules are fired at the same time make not easy to establish logical chains of reasoning. It is also necessary to remark that the use of negation and antonyms are quite usual in natural language but their representation using fuzzy logic is still a matter of research.

ISBN: 978-989-95079-6-8

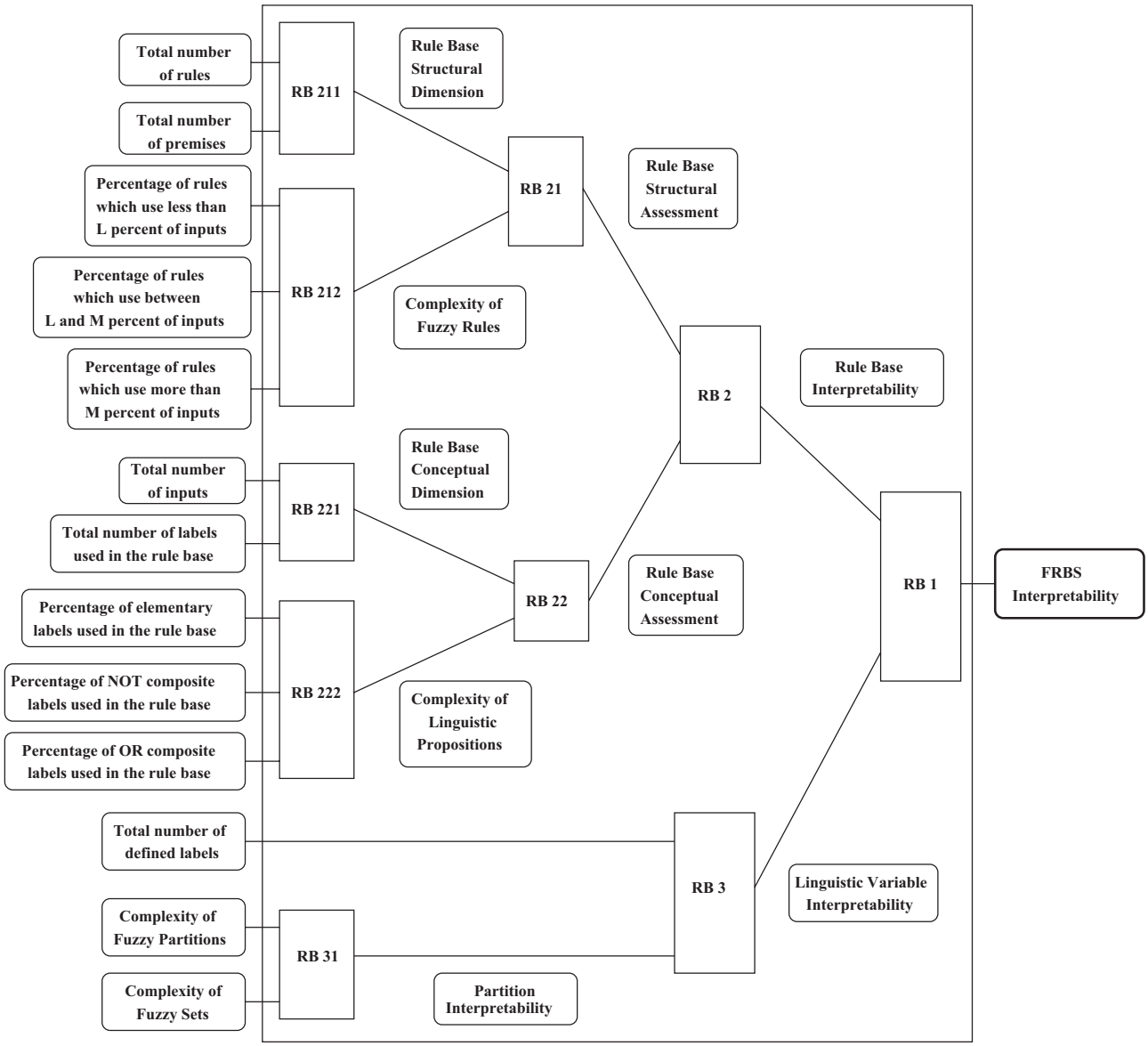

Figure 2: A conceptual framework for assessing interpretability (*Global description*) of FRBSs.

3 Measuring interpretability

Once identified all factors that should be kept in mind regarding interpretability, the definition of a universal interpretability index able to combine all of them becomes a great challenge.

The aim of this section is to introduce a conceptual framework for assessing interpretability. It is represented in the form of a hierarchical diagram in Fig. 2. Of course, this diagram only takes into account interpretability from the structural point of view (*Global description*). The local explanation will be addressed in future works.

To start with, the whole set of factors represented in left part of Fig. 1 has been summarized by a small subset that could be extended in the future. The global diagram can be seen as a flow chart with thirteen inputs (measurable factors to take into account regarding interpretability) and one output (Interpretability measure). The diagram keeps the abstraction levels shown in Fig. 1 (low-level at the bottom of the figure and high-level at the top) and selected inputs are grouped according to the information they convey.

Interpretability of a FRBS is estimated as a combination of two estimators at both low and high level. On the one hand, partition interpretability regards the complexity of each fuzzy set but also the complexity of the whole partition. An estimation of the low-level interpretability (regarding the description of all linguistic variables) is computed adding the number of labels (linguistic terms). Notice that the simple diagram depicted in Fig. 2 does not include anything with respect to the interpretability of linguistic partitions. It is assumed the use of SFPs and global semantics, not entering to the way how linguistic terms are named. On the other hand, high-level interpretability (called *rule base interpretability* in the figure above) involves analysis at the three highest sublevels (linguistic propositions, linguistic fuzzy rules, and rule base).

Drawing an analogy between a set of fuzzy rules and a set of sentences in natural language, the interpretability is assessed regarding both Syntax and Semantics. On the one hand, Syntax can be defined as the *arrangement of words in sentences, clauses, and phrases, and the study of the formation of sentences and the relationship of their component parts*.

ISBN: 978-989-95079-6-8

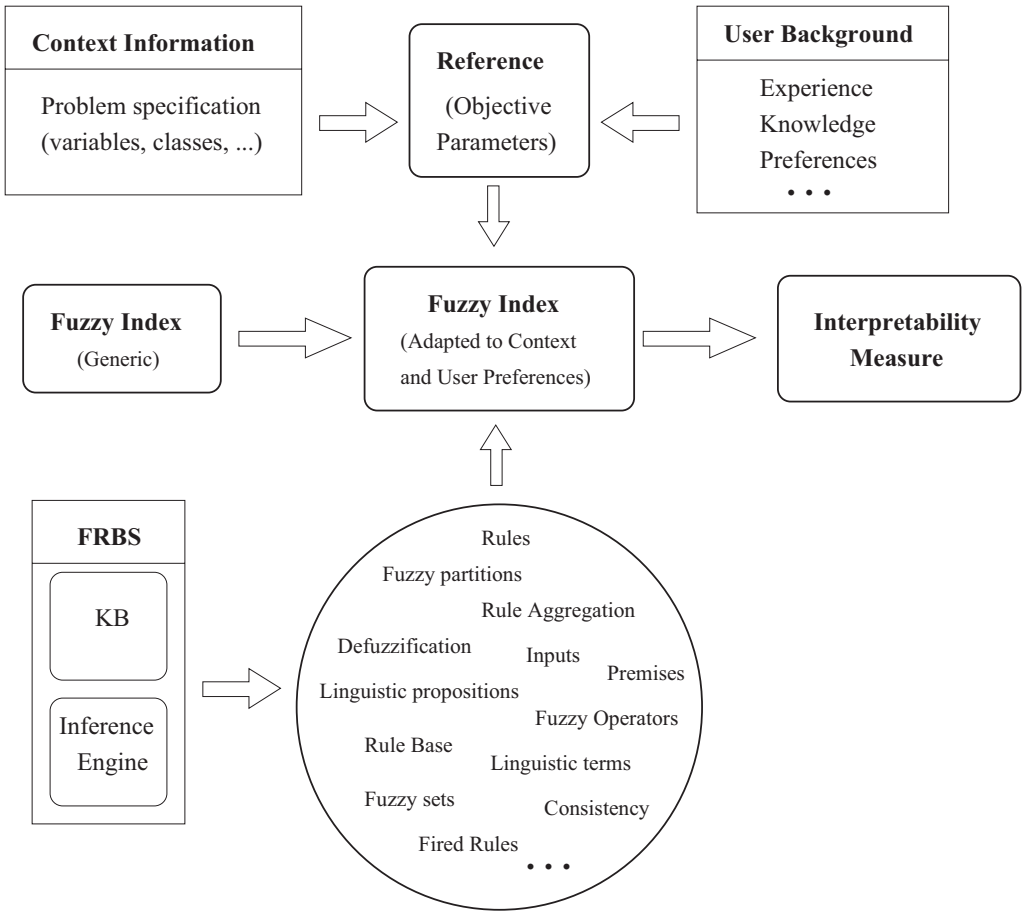

Figure 3: A fuzzy index (adaptable to context problem and user preferences) for measuring Interpretability.

On the other hand, Semantics makes reference to the *Study of meaning*[1]. In our context Syntax is related to dimension and complexity of the rule base (what we have named as *Rule Base Structural Assessment*), computed in a very simplistic way, only considering the total number of rules, premises and premise by rule. It covers both linguistic fuzzy rules and rule base abstraction levels. Semantics in turn takes into account the complexity at the level of linguistic propositions used in the rules (what we have named as *Rule Base Conceptual Assessment*).

The process of measuring something consists in comparing it with a reference (standard unit of measurement) such us a meter for measuring length. However, finding out the suitable reference is not always feasible and the task is especially difficult when measuring non-physical properties. It is widely admitted that interpretability assessment is clearly context dependant. There is not a universal reference; on the contrary the reference will change depending on the problem and depending on the person who makes the assessment. Therefore, the general proposed framework has to be adapted to the specific features of each problem under consideration as well as to the user's background and preferences. For instance, *Total number of rules* may be defined as a linguistic variable made up of five linguistic terms (*Very low*, *Low*, *Medium*, *High*, *Very high*). Nevertheless, the meaning of each linguistic term needs to be defined carefully. *What value should be taken as a prototype for small number of rules (three, ten, one hundred, etc)?* If we were analyzing a FRBS for classification among three kinds of wines, the minimum number of rules should be three, but we need to ask to the people who are going to interact with the system in order to really know how to characterize the linguistic variables. In fact, the perception of interpretability will change depending on the kind of user. The point of view of a system designer who is used to work with fuzzy systems is likely to be very different from the point of view of the domain expert who perfectly knows the problem and how should be the system behavior, but it will be even much more different from the final user who could have only a superficial knowledge of the problem, and who probably has not heard anything about fuzzy logic.

Fig. 3 describes how to build an easily adaptable fuzzy index for assessing interpretability. A generic fuzzy index like the one presented in Fig. 2 has to be tuned and adapted for each problem regarding both the problem definition and the user quality criteria. The system is flexible enough for making an easy adaptation. It consists in defining a reference with all collected information about the problem and the evaluator user (system designer, domain expert, and/or final user). Such reference yields the ranges (universes of discourse) along with the modal points of the fuzzy partitions used to define the input variables (*Total number of rules*, *Total number of premises*, *Percentage of rules which use less than L percent of inputs*, etc.) of the generic fuzzy index. The linked rule bases as well as the intermediate input-output variables could be tuned too.

[1]Both definitions were got from the Encyclopedia Britannica (http://www.britannica.com)

ISBN: 978-989-95079-6-8

4 Final remarks

Previous works has made a great effort to establish the basis for building interpretable fuzzy systems. There are many different works regarding interpretability on the fuzzy literature. Recently, some works have made a global review of the literature putting together contributions of different authors. Following that way, this work has formalized a conceptual framework for characterizing and assessing interpretability of fuzzy systems.

The use of multi-objective approaches is becoming a more and more important topic in fuzzy modeling [17] because of interpretability and accuracy are conflictive goals. In this specific field the interpretability of the model is usually only considered from the point of view of the fuzzy designer. First, it is necessary to make a qualitative and quantitative comparison of all obtained solutions. Then, the best solutions can be selected from a Pareto front regarding the accuracy-interpretability trade-off. It is possible to set a qualitative ranking of solutions based on a comparison per couples, without measuring the interpretability of each individual solution, setting some kind of pre-order is enough. Although there are several accuracy indices, interpretability is measured taking into account only basic parameters what is a strong limitation. Thus, the use of interpretability indices guiding the modeling process could help to achieve better solutions.

Setting qualitative rankings is quite common in the context of semantic web search where retrieved documents have to be ranked before presenting them as answer to a query. For instance, BUDI [18] is a meta-searcher based on fuzzy logic which uses a fuzzy similarity function for comparing documents. It regards the size of the documents, the number of series of words in the same position in both documents, but also the complexity and rarity of words and linguistic propositions. This approach could be extended to the interpretability assessment problem, considering that instead of documents what are going to be compared are the linguistic descriptions of FRBSs.

In the future, experimental analysis must be carried out in order to adapt the theoretical developments to the real worlds. In order to get a universal index adaptable to the user preferences, it is necessary to study how different kinds of user (fuzzy designer, domain expert, and final user) interact with fuzzy systems in a different way and they have different interpretability requirements.

Finally, it is necessary to advance on the paradigm of computing with words and perceptions (CWW/P) [3] which marks an evolution of fuzzy logic, an extension of current theories of fuzzy sets. It is strongly related with meaning, captured by the use of linguistic expressions, words and connectives. As a result, works regarding interpretability assessment can take profit from current research on CWW/P but also they will help to develop new ideas in relation with this new paradigm.

Acknowledgment

This work has been partially supported by the Foundation for the Advancement of Soft Computing (Asturias) and Spanish government (CICYT) under grant: TIN2008-06890-C02-01.

References

[1] C. Mencar and A. M. Fanelli. Interpretability constraints for fuzzy information granulation. *Information Sciences*, 178(24):4585–4618, 2008.

[2] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. *Parts I, II, and III. Information Sciences*, 8, 8, 9:199–249, 301–357, 43–80, 1975.

[3] L. A. Zadeh. From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions. *IEEE Transactions on Circuits and Systems - I: Fundamental theory and applications*, 45(1):105–119, 1999.

[4] J. M. Alonso, L. Magdalena, and S. Guillaume. HILK: A new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism. *International Journal of Intelligent Systems*, 23(7):761–794, 2008.

[5] E. H. Mamdani. Application of fuzzy logic to approximate reasoning using linguistic systems. *IEEE Transactions on Computers*, 26(12):1182–1191, 1977.

[6] H. P. Grice. Logic and conversation. *Cole, P. and Morgan, J. (eds.) Syntax and semantics, New York: Academic Press*, 3:43–58, 1975.

[7] A. Tarski, A. Mostowski, and R. Robinson. *Undecidable Theories*. North-Holland, 1953.

[8] C. Mencar, G. Castellano, and A. M. Fanelli. Some fundamental interpretability issues in fuzzy modeling. In *Joint EUSFLAT-LFA*, pages 100–105, Barcelona, Spain, September, 7-9, 2005.

[9] U. Bodenhofer and P. Bauer. A formal model of interpretability of linguistic variables. *In [19]*, pages 524–545, 2003.

[10] J. Espinosa and J. Vandewalle. Constructing fuzzy models with linguistic integrity from numerical data-afreli algorithm. *IEEE Transactions on Fuzzy Systems*, 8(5):591–600, 2000.

[11] S. Guillaume and B. Charnomordic. Generating an interpretable family of fuzzy partitions from data. *IEEE Transactions on Fuzzy Systems*, 12 (3):324–335, 2004.

[12] S.-M. Zhou and 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(23):3091–3131, 2008.

[13] L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(1):28–44, 1973.

[14] E. H. Ruspini. A new approach to clustering. *Information and Control*, 15(1):22–32, 1969.

[15] P. Fazendeiro, J. V. de Oliveira, and W. Pedrycz. A multiobjective design of a patient and anaesthetist-friendly neuromuscular blockade controller. *IEEE Transactions on Biomedical Engineering*, 54(9):1667–1678, 2007.

[16] L. A. Zadeh. *Applied Soft Computing - Foreword*, 1(1):1–2, 2001.

[17] H. Ishibuchi and Y. Nojima. Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning. *International Journal of Approximate Reasoning*, 44(1):4–31, 2007.

[18] J. Serrano-Guerrero et al. BUDI: Architecture for fuzzy search in documental repositories. *Mathware and Soft Computing*, 16(1):71–85, 2009.

[19] J. Casillas et al. *Interpretability issues in fuzzy modeling*, volume 128. Studies in Fuzziness and Soft Computing, Springer-Verlag, Heidelberg, 2003.

ISBN: 978-989-95079-6-8