An Experimental Study on the Interpretability of Fuzzy Systems

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— *Interpretability is one of the most significant properties of Fuzzy Systems which are widely acknowledged as gray boxes against other Soft Computing techniques such as Neural Networks usually regarded as black boxes. It is essential for applications with high human interaction (decision support systems in medicine, economics, etc). The use of accuracy indices to guide the fuzzy modeling process is broadly extended. In turn, although there have been a few attempts to define Interpretability indices, we are still far away from having a universal index. With the aim of evaluating the most used indices an experimental analysis (in the form of a web poll) was carried out yielding some useful clues to keep in mind regarding Interpretability assessment. Results extracted from the poll show the inherent subjectivity of the measure because we collected a huge diversity of answers. Nevertheless, comparing carefully all the answers, it was possible to find out some interesting user profiles.*

Keywords— Fuzzy modeling, Interpretability assessment.

1 Introduction

Fuzzy modeling (FM), i.e., system modeling with fuzzy rule-based systems (FRBSs), is an important and active research line inside the fuzzy community. Fuzzy Logic (FL) was introduced by Zadeh [1] (in 1965) and its semantic expressivity, using linguistic variables [2] and linguistic rules [3], favors the interpretability of the modeled system (at least from the structural transparency viewpoint) because it is quite close to expert natural language. From 1965 to 1990, fuzzy designers focus on modeling highly interpretable systems, mainly working with expert knowledge and a few simple linguistic rules. Then, researchers realized that to deal with complex systems expert knowledge was not enough. Thus, from 1990 to 2000, the main effort was made regarding the accuracy of the final model, building complicated fuzzy rules with high accuracy (applying machine learning techniques to extract knowledge from data) but disregarding the model interpretability because automatically generated rules are rarely as readable as desired. Nowadays, a new challenge lies in looking for compact and robust systems with a good accuracy-interpretability trade-off.

Regarding the interpretability assessment of FRBSs, their comprehensibility depends on all their components, i.e., it depends on the knowledge base (KB) transparency but also on the inference mechanism understanding. There are also some crucial psychological factors; for instance, for some people the most interpretable models are the ones they are used to work with, disregarding the model complexity. Anyhow, previous works [4, 5] have thoroughly analyzed the main factors (rule base and fuzzy partitioning) that influence the KB readability. Also, a complete study on the interpretability constraints most frequently used in literature has been recently published [6]. However, once identified all such elements, the current challenge lies in how to combine them in order to obtain a good index. Unfortunately, only a few works have dealt with this issue. As explained by [7] it is possible to distinguish two main interpretability levels: Low-level or fuzzy set level, and high-level or fuzzy rule level. There are some works regarding interpretability measurement at low-level [8, 9, 10, 11, 12] which consist of mathematical formulas to evaluate the main partition properties such as distinguishability, similarity, coverage, overlapping, etc. These indices are usually used to preserve the interpretability of fuzzy systems automatically generated from data. They are also used as part of tuning processes devoted to increase the accuracy of the final model while keeping good interpretability. Furthermore, there are some simple indices, mainly applied to multi-objective fuzzy genetics-based machine learning, regarding the rule base interpretability [13]: (1) *Number of rules*; (2) *Total rule length* (addition of the number of premises defined in all the rules); (3) *Average rule length* (total rule length divided by the number of rules).

However, only a few researchers have tackled with the challenge of defining an index covering both low and high levels. The first one was the *Nauck's index* [14], a numerical index designed (in 2003) to evaluate fuzzy rule-based classification systems, and computed as the product of three terms: $I_{Nauck} = Comp \times Part \times Cov$. *Comp* represents the complexity of a classifier measured as the number of classes divided by the total number of premises. *Part* stands for the average normalized partition index overall input variables. It is computed as the inverse of the number of labels minus one (two is the minimum number of linguistic terms in a partition) for each input variable. Finally, *Cov* is the average normalized coverage degree of the fuzzy partition. It is equal to one for strong fuzzy partitions (SFPs).

A second global index was defined in 2006 [15] and improved in 2008 [16]. It consists of a *fuzzy index* which was initially inspired on the Nauck's index. Six variables (total number of rules, total number of premises, number of rules which use one input, number of rules which use two inputs, number of rules which use three or more inputs, and total number of labels defined by input) are considered as inputs of a fuzzy system and they are grouped, according to the information they convey. In consequence, the interpretability index is computed as the result of inference of a hierarchical fuzzy system made up of four linked KBs. It is specially designed for the context of classification problems solved by means of a specific kind of FRBSs generated following the HILK (Highly Interpretable Linguistic Knowledge) [16] methodology, and assuming SFPs for all system variables.

The rest of the paper is structured as follows. Section 2

ISBN: 978-989-95079-6-8

Table 1: Comparison of interpretability indices (measures).

	Method	Number of rules	Total rule length	Average rule length	Nauck's index	Fuzzy index
KB1	CL-FDT-DS-FDT	20	43	2.15	0.0174	0.452
KB2	CL-FDT-DS-FDT-S	5	9	1.8	0.1667	0.92
KB3	CL-FDT-DS-WM	53	643	12.132	0.0011	0.144
KB4	CL-FDT-DS-WM-S	8	16	2	0.1484	0.839
KB5	CL-WM-DS-FDT	21	49	2.333	0.0153	0.444
KB6	CL-WM-DS-FDT-S	5	10	2	0.2625	0.919
KB7	CL-WM-DS-WM	46	545	11.848	0.0013	0.192
KB8	CL-WM-DS-WM-S	3	6	2	0.3056	0.924
KB9	FDT-S	8	19	2.375	0.0763	0.814
KB10	WM-S	6	18	3	0.0873	0.742
KB11	FDT-P	32	94	2.937	0.0079	0.392
KB12	FDT-P-S	6	15	2.5	0.1714	0.837

Table 2: Comparison of interpretability indices (ranking).

Index	+ Interpretability -
Number of rules	KB8, KB2/KB6, KB10/KB12, KB4/KB9, KB1, KB5, KB11, KB7, KB3
Total rule length	KB8, KB2, KB6, KB12, KB4, KB10, KB9, KB1, KB5, KB11, KB7, KB3
Average rule length	KB2, KB4/KB6/KB8, KB1, KB5, KB9, KB12, KB11, KB10, KB7, KB3
Nauck's index	KB8, KB6, KB12, KB2, KB4, KB10, KB9, KB1, KB5, KB11, KB7, KB3
Fuzzy index	KB8, KB2, KB6, KB4, KB12, KB9, KB10, KB1, KB5, KB11, KB7, KB3

makes a comparison of several interpretability indices in an experimental analysis. Results extracted from a web poll show clearly the intrinsic subjectivity of the measure. As it is explained in section 3, although we got a huge diversity of answers, at first glance completely different, after looking carefully it was possible to find out some interesting user profiles. Finally, section 4 offers some conclusions and points out future works.

2 Experimental analysis

With the aim of making a fair (qualitative and quantitative) comparison of the five indices remarked in the previous section (Number of rules, Total rule length, Average rule length, Nauck's index, and fuzzy index) this experimental study deals with the well known benchmark classification problem called WINE which data set is freely available at the UCI[1] (University of California, Irvine) machine-learning repository. It contains 178 instances coming from results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines.

Following the HILK modeling methodology [16] twelve KBs, of several sizes, have been generated for the WINE recognition problem. Looking for maximizing the interpretability of final KBs, a global semantics (based on the use of SFPs) is defined previously to rule definition. As a result, for each KB all the rules use the same linguistic terms (defined by the same fuzzy sets), and rule comparison can be done at the linguistic level. The original data set was randomly divided into two subsets taking 50% of data for training and the remaining part for test. HILK copes with different rule induction techniques in order to get enough diversity. Second column in Table 1 contains the abbreviations of the combined methods. CL means clustering previous to rule induction, WM represents the well-known Wang and Mendel's algorithm [17], FDT stands for the popular Fuzzy Decision Tree algorithm [18], DS is data selection in training set previous to rule induction, P means pruning of the tree, and S stands for simplification procedure. All selected algorithms are implemented in KBCT [19], a free software tools for designing fuzzy systems. WM and FDT implementations differ from the original ones in the fuzzy partition design step. Interpretable fuzzy partitions are defined previous to rule induction (in our study five labels per variable were initially defined). Details about the induction algorithms are explained in the cited literature.

The five interpretability indices have been compared from both quantitative (Table 1) and qualitative (Table 2) viewpoints. Table 1 includes results of computing the five selected interpretability indices for the twelve generated KBs. The comparison of the obtained values lets us set rankings from the interpretability point of view. Notice that KBs with equivalent interpretability are set at the same level separated by symbol "/" (see Table 2). From a quantitative point of view the ranges of values are completely different among all the indices. Each individual value only makes sense in comparison with the other values obtained by the same index. From a qualitative point of view we would rather choose those indices yielding a ranking without ambiguities (Total rule length, Nauck's index, and Fuzzy index), i.e., those indices able to produce a full order distinguishing among all pairs of KBs. As expected, we have achieved five different rankings because each interpretability index follows different criteria. Nevertheless, looking carefully it is easy to appreciate that all rankings are somehow similar, the same KBs usually appear at the beginning (KB2, KB4, KB6, KB8) or at the end (KB3, KB7).

However, a last question still remains to be answered: How to know which index is the best one? Since the measure of interpretability is clearly subjective the only way to answer this question is asking people. For that reason, a web poll was addressed to FL experts (50%) as well as people who are not familiar with FL (50%). The study is made regarding the

[1] http://www.ics.uci.edu/~mlearn/MLSummary.html

ISBN: 978-989-95079-6-8

twelve KBs generated for the WINE problem. The goal is to compare the most popular interpretability criteria, including people used (and not used) to work with fuzzy systems that can be (or not) fond of wines. Since interpretability extremely depends on the kind of user, let us add a short comment. In the context of FM, there are three kinds of users:

- The *final user* of the modeled system. In most cases, he/she will interact with the system providing data and/or receiving system suggestions and advices for making decisions. The user will only trust on the system if the system output is coherent according to his/her background. Notice that the use of a comprehensible model can help the final user to understand the system output.
- The *system designer* who has to be an expert on fuzzy logic in order to produce a good model useful for the final user of the application. A transparent (gray-box) model structure is really appreciated for the future model maintenance and update.
- The *domain expert* who will explain the system behavior to the system designer during the model design stage. In addition, he/she will be in charge of validating system running. Since domain experts usually do not know anything about fuzzy logic a clearly readable model description is required to make easier the validation stage.

In our study, FL experts are mainly considered as system designers but due to the nature of the problem they also can act as domain experts and even as final users. In turn, non-FL users are only viewed as domain experts or final users. Twenty six answers were collected. They show a huge diversity what clearly illustrates how different users have very different criteria to measure interpretability. Three main questions were asked as part of the poll:

1. *How much interpretable are the twelve KBs?* Each user was asked to give an interpretability measure for each KB. Such measure was represented by an interval (min-max), i.e., the range in which it should be included, between zero and one hundred. However, only a few users were willing to answer to this question with numerical values. In fact, we realize that people find much more natural to make approximate reasoning based on the use of linguistic terms like *Highly interpretable*, *Moderately interpretable*, etc. In addition collected values show a huge variance. In consequence, it does not make sense drawing statistical conclusions from stored data. According to these results it can be argued that people get into difficulties when they have to give numerical indices as computers usually do.

2. *What is the KB interpretability ranking?* Users were asked to rank the KBs according to their preferences from the interpretability point of view: One for the most interpretable KB, and twelve for the least interpretable one. Since all users were willing to answer this question, an interesting conclusion can be drawn: *People feel much more confident setting rankings than giving numerical values*. In order to set a ranking it is necessary to compare all the KBs (by couples) but it does not imply setting individual measures. First column of Table 3 includes the user identifier, setting in brackets if the user is used to work with fuzzy systems (F) or not (NF). Second column of the table shows rankings given by users. As it can be seen at first glance there is a huge variance. Only two couples of users (1-26 and 4-11) gave exactly the same order. Nevertheless, looking carefully answers are not so different. The global order is more or less the same for all users but when two KBs are quite close regarding interpretability the final ranking choice depends in many subtle details, and as a result, there is a clearly subjective choice at the end.

   The comparison between rankings provided by the users (Table 3) and rankings derived from the computed interpretability indices (Table 2) lets us evaluate the goodness of such indices. However, only user 3 and fuzzy index yield the same ranking. In order to make a deeper analysis, we have computed Euclidean distance from each of the five interpretability indices, x (first line of the table), to all the twenty six users, y, according to equation 1 where x_i means the ranking position of KB_i regarding index x and y_i is the ranking position of KB_i regarding user y.

   $$d_{x,y} = \sqrt{\sum_{i=1}^{12} |x_i - y_i|} \tag{1}$$

   Computed distances give an idea on how different (comparing positions of each KB in selected rankings) indices and user's answers are. Most user's answers are closer to the rankings obtained using *Number of rules*. There are also many answers closer to *Total rule length* and *Fuzzy index*. In fact, it is possible to identify several groups of users (look at Table 4) whose answers fit better with some indices but there is a lot of overlapping among groups. Moreover, approximately the same number of F and NF users belongs to all the groups. This is due to the fact that in general none of the computed indices fit properly in user's answers.

3. *What are the most relevant aspects to consider when assessing interpretability?* Each user was asked to give short comments explaining what he/she considers good strategies and/or key criteria to measure interpretability. Some of the most useful comments are listed below:

   - *A common heuristic reasoning is the following. First look at the total number of rules. Second, if there is ambiguity between some of the knowledge bases it is needed to check the total number of premises. Then, if there is still ambiguity it is necessary to analyze the complexity of the linguistic terms*. This suggests making the ranking in different abstraction levels, adding new criteria only when they are needed.
   - *I prefer short rules considering at most 5 features than fewer rules with a long size*. This remarks that the number of inputs by rule is a main criterion. However, different people have different views about what must be considered as a small

ISBN: 978-989-95079-6-8

Table 3: Ranking of KBs extracted from the poll results.

User	+ Interpretability -
user1 (F)	KB8, KB2, KB6, KB12, KB10, KB4, KB9, KB1, KB5, KB11, KB7, KB3
user2 (F)	KB6, KB2, KB4, KB1, KB8, KB5/KB9, KB10, KB12, KB11, KB7, KB3
user3 (F)	KB8, KB2, KB6, KB4, KB12, KB9, KB10, KB1, KB5, KB11, KB7, KB3
user4 (NF)	KB2, KB6, KB8, KB12, KB9, KB10, KB4, KB1, KB5, KB11, KB7, KB3
user5 (F)	KB2, KB6, KB8, KB9, KB10/KB12, KB4, KB1, KB5, KB11, KB7, KB3
user6 (F)	KB8/KB12, KB2/KB6, KB4/KB9/KB10, KB1/KB5, KB11, KB3, KB7
user7 (F)	KB8, KB6, KB2, KB12, KB10, KB9, KB4, KB1, KB5, KB3, KB11, KB7
user8 (F)	KB8, KB2, KB6, KB12, KB9, KB4, KB10, KB1, KB5, KB11, KB7, KB3
user9 (NF)	KB2, KB9, KB12, KB8, KB6, KB10, KB5, KB4, KB1, KB11, KB3, KB7
user10 (NF)	KB6, KB2, KB9, KB12, KB4, KB8, KB5, KB1, KB11, KB10, KB7, KB3
user11 (NF)	KB2, KB6, KB8, KB12, KB9, KB10, KB4, KB1, KB5, KB11, KB7, KB3
user12 (NF)	KB8, KB12, KB2, KB6, KB9, KB4, KB10, KB5, KB1, KB11, KB7, KB3
user13 (F)	KB2, KB6, KB8, KB9, KB12, KB4/KB10, KB1/KB5/KB11, KB3/KB7
user14 (NF)	KB8, KB2/KB6, KB12, KB9, KB4, KB10, KB1, KB5, KB11, KB7, KB3
user15 (NF)	KB8, KB2, KB12, KB6, KB10, KB1, KB5, KB9, KB4, KB11, KB7, KB3
user16 (NF)	KB8, KB6, KB2, KB12, KB10, KB4, KB9, KB5, KB1, KB11, KB7, KB3
user17 (NF)	KB8, KB6, KB2, KB12, KB4, KB9, KB5, KB1, KB11, KB10, KB3, KB7
user18 (NF)	KB2/KB12, KB4/KB6/KB8/KB10, KB5/KB9, KB11, KB1, KB7, KB3
user19 (NF)	KB2, KB4, KB6, KB11, KB5, KB1, KB9, KB12, KB10, KB8, KB7, KB3
user20 (F)	KB2, KB8, KB9, KB12, KB4, KB6, KB10, KB11, KB1, KB5, KB7, KB3
user21 (NF)	KB2, KB6, KB8, KB12, KB9, KB4, KB10, KB11, KB5, KB1, KB7, KB3
user22 (F)	KB8, KB6, KB2, KB9, KB12, KB10, KB4, KB5, KB11, KB1, KB3, KB7
user23 (NF)	KB8, KB2, KB6, KB12, KB10, KB9, KB4, KB1, KB5, KB11, KB7, KB3
user24 (F)	KB2, KB8, KB6, KB12, KB4, KB9, KB1, KB10, KB5, KB11, KB7, KB3
user25 (F)	KB8, KB2/KB4/KB6/KB9/KB10/KB12, KB1/KB5/KB11, KB3/KB7
user26 (F)	KB8, KB2, KB6, KB12, KB10, KB4, KB9, KB1, KB5, KB11, KB7, KB3

Table 4: Groups of users regarding computed interpretability indices.

Index	Users	F	NF
Number of rules	4, 5, 7, 9, 11, 13, 15, 16, 22, 23, 25	5	6
Total rule length	1, 8, 12, 14, 18, 21, 25, 26	4	4
Average rule length	2, 19	1	1
Nauck's index	6, 12, 25	2	1
Fuzzy index	3, 8, 10, 14, 17, 20, 21, 24, 25	5	4

number of inputs by rule. This problem arises from the intrinsic ambiguity of natural language: What means small? The same word has different meanings in different contexts, but even in the same context it has different meanings for different people. The use of fuzzy logic formalizes a precise meaning for each word coping with this kind of ambiguity.

- *With respect to words (linguistic variables and terms), the better choice of words within the context of the problem, the more accurate interpretation.* Understanding strongly depends on the context of the problem. For instance, it is easy to see how different the meaning of *High* is when talking about people, buildings, or mountains.

3 User profiles

With the aim of finding out some user profiles from Table 3 we have applied a hierarchical clustering analysis [20]. Two dendrograms were built using Ward's method [21] and squared Euclidean distance (see Fig. 1): The first one (on the top part of the figure) only regarding fuzzy expert users where two groups (SF1 and SF2) are clearly identified. The second one (on the bottom part of the figure) including only non-fuzzy users where only one group (SNF1) can be defined. The rest of users are progressively added by the clustering algorithm but they stay at long distance.

It is possible to extract a prototype user profile from each group. According to our experience designing and assessing interpretable fuzzy systems, and keeping in mind the conclusions derived from the web poll, ten variables were selected as tentative interpretability indicators: (1) Number of rules; (2) Total rule length; (3) Percentage of rules which use less than ten percent of inputs; (4) Percentage of rules which use between ten and thirty percent of inputs; (5) Percentage of rules which use more than thirty percent of inputs; (6) Number of inputs; (7) Number of labels used in the rule base; (8) Percentage of elementary labels used in the rule base; (9) Percentage of OR composite labels used in the rule base; (10) Percentage of NOT composite labels used in the rule base.

The task consists of discovering those indicators that can be considered as key to distinguish among groups. From the rankings provided by users it can be induced the order relation between KBs. Assuming that each ranking is based on a comparison per couples of all KBs, Table 3 is translated into a data set with the following format. Each column give the difference between the ten selected indicators (listed above) for each couple of KBs (A and B). The last column includes one

ISBN: 978-989-95079-6-8

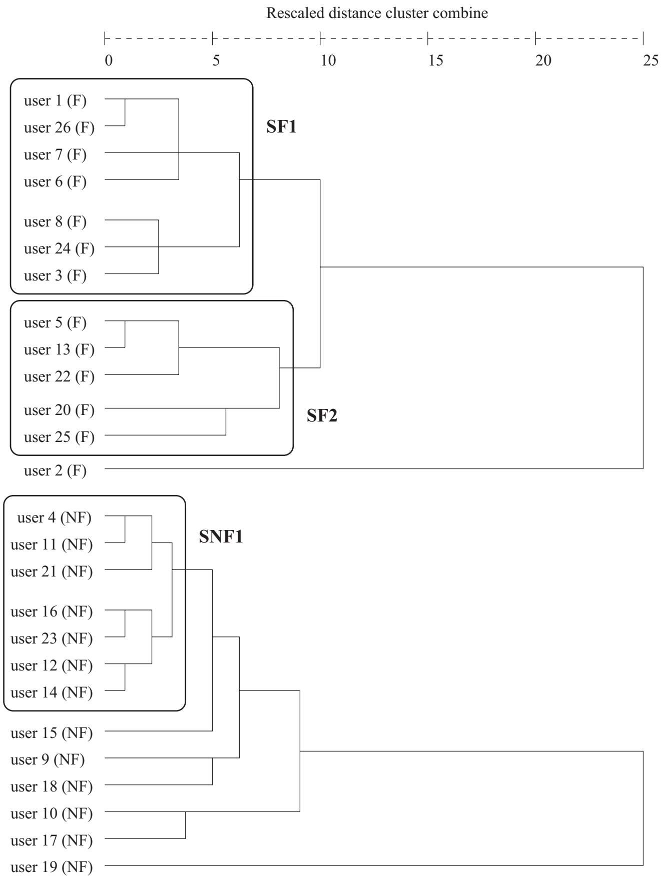

Figure 1: Groups of users by hierarchical clustering.

of three labels (1:*No*, 2:*Yes*, 3:*I don't know*) answering to the question: is A more interpretable than B? The whole data set was divided into three subsets including only the comparisons related to the users belonging to each group.

Using HILK methodology [16] (FDT-S method) and the previously generated data subsets we have built three fuzzy classifiers, one for each group of users. For each couple of KBs (A and B) the fuzzy classifiers give as output (1) *A is more interpretable than B*, (2) *B is more interpretable than A*, or (3) *I don't know*, what means that A and B are similar from an interpretability point of view. The comparison among all KBs yields a ranking for each group. Figure 2 presents obtained rankings as well as the interpretability indicators (inputs of the fuzzy classifier) that define the six prototype user profiles we were looking for. Three indicators are not selected by any of the prototypes: (3) Percentage of rules which use less than ten percent of inputs; (5) Percentage of rules which use more than thirty percent of inputs; and (9) Percentage of OR composite labels used in the rule base. It seems that all users agree that a small percentage of inputs by rule is good for interpretability and a large percentage is bad, while composite propositions including OR are seen as easily readable. Notice that each prototype is defined using at least three indicators, i.e., the use of basic indices is not enough.

In addition, for each group we have computed the distance (mean and variance) between the prototype and all users included in the group. SF1 represents a quite compact group where the prototype is defined using only three indicators, while SF2 and SNF1 are less compact and because of that they need more indicators. Although clustering of F and NF users should intuitively yield more homogeneous groups, in practice there is still a lot of diversity inside each group. In the case of F users, prototypes achieve medium mean distance with small variance. SF2 takes into account six indicators because it is made up of only five users quite heterogeneous. It results specially interesting the fact that *Total rule length* which seemed to be a very important basic indicator according to Table 4 only is taken into account by group SF2. It is not nearly the most relevant indicator in comparison with the subsets of indicators emerging from the clustering analysis. Finally, it could be argued that fuzzy users (SF1 and SF2) give more homogeneous answers but regarding more complex criteria than non-fuzzy users (SNF1). Finally, as a result of the heterogeneity of NF users only seven of them give more or less similar rankings. In consequence, four indicators were relevant for this group but yielding large mean and variance.

4 Conclusions

Assessing interpretability is a very challenging and complex task due to the inherent subjectivity of the measure. In order to evaluate existing indices we have set up a first experimental study, for simplicity limited to twelve rule bases assuring most interpretability constraints described as essential in the literature. As a result, assuming knowledge bases under study are interpretable the study focus on quantifying interpretability and comparing obtained results with assessment provided by people in a web poll. None of the evaluated indices gave good results in comparison with rankings provided by human beings. A lot of work remains still to be done so that finding a universal index. However, results derived from our experimental study offer some interesting clues.

First, it is necessary to define a new index flexible enough to be easily adaptable to the problem context and user preferences. Such index must take into account many subtle details combined in the context of computing with words and perceptions based on fuzzy logic technology. In addition, obtaining a numerical value is not needed in most applications where the important thing is to set an appropriate ranking.

Second, a hierarchical fuzzy framework has been proved as a powerful tool to imitate the usual way of people reasoning. It mainly consists of taking a few interpretability indicators as guide to discriminate between two knowledge bases, adding more criteria only when it is necessary because the compared knowledge bases are not distinguishable at first glance.

Finally, we have focused on interpretability from a structural point of view but there are many cognitive aspects that should be analyzed. Therefore more experimental studies are needed. Obviously, as a first step our study has been limited to a very specific kind of fuzzy rule-base systems. Of course, in the future it would be interesting to make a comparison of different rule base structures.

Acknowledgment

Authors would like to acknowledge people who have filled up the web poll. Their suggestions were very interesting and they have contributed so much to this work.

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

ISBN: 978-989-95079-6-8

SF1

(1) Number of rules
(4) Percentage of rules which use between 10 and 30 percent of inputs
(7) Number of used labels

SF2

(2) Total rule length
(4) Percentage of rules which use between 10 and 30 percent of inputs
(6) Number of inputs
(7) Number of used labels
(8) Percentage of elementary labels
(10) Percentage of NOT labels

SNF1

(1) Number of rules
(4) Percentage of rules which use between 10 and 30 percent of inputs
(7) Number of used labels
(10) Percentage of NOT labels

		Distance	
	Ranking	Mean	Variance
SF1	KB8, KB2, KB6, KB12, KB4, KB9/KB10, KB1, KB5, KB11, KB7, KB3	2.616	0.597
SF2	KB2/KB8, KB6, KB9/KB12, KB4, KB10, KB1, KB5, KB11, KB3, KB7	3.842	0.551
SNF1	KB8, KB2, KB6, KB12, KB9, KB10, KB4, KB1, KB5, KB11, KB7, KB3	2.789	3.086

Figure 2: Prototype user profiles.

References

[1] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

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

[4] 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.

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

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

[7] 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.

[8] A. Botta, B. Lazzerini, F. Marcelloni, and D. C. Stefanescu. Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index. *Soft Computing*, 13(5):437–449, 2009.

[9] P. Fazendeiro and J. V. de Oliveira. A working hypothesis on the semantics/accuracy synergy. In *Joint EUSFLAT-LFA*, pages 266–271, Barcelona, Spain, September, 7-9, 2005.

[10] Y. Jin, W. von Seelen, and B. Sendhoff. On generating FC^3 fuzzy rule systems from data using evolution strategies. *IEEE Transactions on Systems, Man, and Cybernetics*, 29(6):829–845, 1999.

[11] C. Mencar, G. Castellano, and A. M. Fanelli. Distinguishability quantification of fuzzy sets. *Information Sciences*, 177(1):130–149, 2007.

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

[13] 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.

[14] D. D. Nauck. Measuring interpretability in rule-based classification systems. In *FUZZ-IEEE 2003*, volume 1, pages 196–201, St. Louis, Missouri, USA, May, 25-28, 2003.

[15] J. M. Alonso, S. Guillaume, and L. Magdalena. A hierarchical fuzzy system for assessing interpretability of linguistic knowledge bases in classification problems. In *IPMU06. Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 348–355, Paris, France, July, 2-7, 2006.

[16] 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.

[17] L.-X. Wang and J. M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man and Cybernetics*, 22(6):1414–1427, 1992.

[18] H. Ichihashi, T. Shirai, K. Nagasaka, and T. Miyoshi. Neuro-fuzzy ID3: A method of inducing fuzzy decision trees with linear programming for maximizing entropy and an algebraic method for incremental learning. *Fuzzy Sets and Systems*, 81(1):157–167, 1996.

[19] J. M. Alonso, S. Guillaume, and L. Magdalena. KBCT: A knowledge management tool for fuzzy inference systems. *Free software under GPL license, available in http://www.mat.upm.es/projects/advocate/kbct.htm*, 2003.

[20] S. C. Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967.

[21] J. H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.

ISBN: 978-989-95079-6-8