Analyzing interpretability of fuzzy rule-based systems by means of fuzzy inference-grams

J. M. Alonso, O. Cordón, A. Quirin, and L. Magdalena
European Centre for Soft Computing
Mieres, 33600, Spain
Email: {jose.alonso,oscar.cordon,arnaud.quirin,luis.magdalena}@softcomputing.es

***Abstract*—Since the proposal of Zadeh and Mamdani's seminal ideas, interpretability is acknowledged as one of the most appreciated and valuable characteristics of fuzzy system identification methodologies. It represents the ability of fuzzy systems to formalize the behavior of a real system in a human understandable way. Interpretability analysis involves two main points of view: readability of the knowledge base description (regarding complexity of fuzzy partitions and rules) and comprehensibility of the fuzzy system (regarding implicit and explicit semantics embedded in fuzzy partitions and rules, but also the fuzzy reasoning method). Readability has been thoroughly treated by many authors who have proposed several criteria and metrics. Unfortunately, comprehensibility has almost never been considered because it involves some cognitive aspects related to the human reasoning which are very hard to formalize and to deal with. This paper proposes the creation of fuzzy systems' inference maps, so-called fuzzy inference-grams (fingrams) by analogy with scientograms used for visualizing the structure of science. Fingrams show graphically the interaction between rules at the inference level in terms of co-fired rules, i.e., rules fired at the same time by a given input vector. The analysis of fingrams offers many possibilities: measuring the comprehensibility of fuzzy systems, detecting redundancies and/or inconsistencies among fuzzy rules, discovering the most significant rules, etc. Some of these capabilities are explored in this initial work.**

I. Introduction

Interpretability of a fuzzy system involves the skill of the specific end-user who interprets its linguistic description with the aim of conceiving the significance of the system behavior. In consequence, characterizing and assessing interpretability is a very subjective task which strongly depends on the background (experience, preferences, knowledge, etc.) of the person who makes the assessment [1].

It is worthy to highlight that interpretability is a distinguishing capability of fuzzy systems which is really appreciated in most applications. It becomes an essential requirement for those applications that involve an extensive interaction with human beings. For instance, decision support systems in medicine [2] must be easily understandable, for both physicians and patients, with the aim of being reliable, i.e., widely accepted and successfully applicable.

Unfortunately, fuzzy systems are not interpretable *per se*. Of course, the use of linguistic variables and rules [3], [4] favors interpretability because of their high semantic expressivity close to natural language. Nevertheless, there are many different issues which must be taken into account in order to design interpretable fuzzy systems. Firstly, several interpretability constraints [5] have to be imposed along the whole design process with the aim of producing a fuzzy system with the required interpretability level, i.e., a system capable of being understood, explicated or accounted for by a human being. As a result, interpretability is usually achieved at the cost of penalizing accuracy. For this reason, most fuzzy systems are built only paying attention to accuracy and so jeopardizing interpretability. Even in those cases, authors usually claim their fuzzy systems are much more interpretable than those based on black-box techniques, like neural networks, because they are based on fuzzy logic. Such kind of claims is quite questionable and should be rejected because they are deceptive. Notice that, obtaining interpretable fuzzy systems is a matter of design which must be carefully considered to avoid producing fuzzy systems so hardly interpretable that they become useless black-boxes from the interpretability point of view.

The assessment of interpretability has to face two main issues [1]: Readability of the system description; and comprehensibility of the system explanation. Of course, the analysis has to take into account all elements included in a fuzzy system, from the lowest (fuzzy partitions) to the highest (fuzzy rules) abstraction levels [6].

Most previous works [7], [8] only analyze the readability of the designed fuzzy system. Moreover, the analysis of readability usually is reduced to a basic analysis of complexity, i.e., it consists of counting the number of elements included in the knowledge base (number of rules, premises, linguistic terms, etc). Other contributions also analyze structural properties of fuzzy partitions [9] such as distinguishability, coverage, and so on. Recently, a few authors have shown the importance of extending the analysis of readability to evaluate the implicit and explicit semantics embedded in a fuzzy system [10], [11]. Of course, keeping a small number of linguistic terms is also appreciated due to the limits of human processing capabilities [12]. Moreover, the selection of the right linguistic terms is essential to yield interpretable systems.

Although there has been a huge effort for defining, characterizing, and assessing interpretability in the last decade, there is still a lot of work to be done. Namely, the analysis related to comprehensibility of the system explanation is almost negligible. Understanding the system behavior from its linguistic description is a very hard task that involves the inference level going beyond the analysis of the system structure readability.

This work presents a new methodology for analyzing the fuzzy inference layer that may be really useful during the design process. It is mainly based on the adaptation of given techniques for visualizing scientific information to the visual analysis of the inference process of fuzzy systems. It also introduces a new index for assessing interpretability of fuzzy rule-based systems facing the challenge of assessing the comprehensibility of the system explanation.

The rest of the contribution is organized as follows. Section II presents some preliminaries including basic aspects related to interpretability assessment, a brief overview on existent methodologies for visual representation and analysis of fuzzy systems, and a short introduction to the most extended techniques for designing visual science maps. Section III introduces the new proposed methodology. As a first step, the general approach is particularized for the analysis of fuzzy rule-based classifiers (FRBCs). Section IV summarizes the experiments carried out along with the achieved results. Finally, some conclusions and future works are sketched in Section V.

II. Preliminaries

A. Assessing Interpretability of Fuzzy Rule-based Systems

While regarding accuracy assessment it is easy to find universal indices commonly accepted, this is not the case when dealing with interpretability assessment. The evaluation of accuracy consists of measuring the difference between the outputs of the model and the real system. For instance, the mean square error and the number of misclassified cases are widely used for regression and classification problems, respectively. On the contrary, assessing interpretability remains an open hot topic. Moreover, finding out a universal index for interpretabiliy seems to be an impossible mission since it is strongly affected by subjectivity. In fact, it is necessary to look for two kind of complementary indices, objective and subjective ones. On the one hand, objective metrics are needed to make feasible fair comparisons among different fuzzy systems. On the other hand, subjective measures are demanded when looking for personalized fuzzy systems where it is required having an index flexible enough to be easily adaptable to end-user's expectations.

A complete taxonomy on the existing interpretability measures has been recently proposed in [13]. Authors identify four groups of indices as result of combining two different criteria, the nature of the interpretability index (complexity vs. semantic) and the elements of the fuzzy knowledge base that it considers (partitions vs. rule base). Namely, the four groups are: (1) Complexity at rule base level; (2) complexity at partition level; (3) semantic interpretability at rule base level; and (4) semantic interpretability at partition level.

Most well-known existing interpretability indices correspond to groups (1) and (2). They only focus on readability (complexity) of fuzzy systems. In consequence, they are objective indices since they are limited to count the number of elements (features, membership functions, rules, premises, etc.) included in the knowledge base.

Indices included in group (4) usually measure the degree of fulfillment of semantic constraints that should be overimposed during the design process. It is widely admitted that working with the so-called Strong Fuzzy Partitions (SFPs) [14] satisfies all semantic constraints required to have interpretable fuzzy partitions from the structural point of view (coverage, normalization, distinguishability, etc).

Finally, group (3) is the one that contains fewer works in the literature. It comprises some indices mainly devoted to evaluate the rule base consistency. In addition, there are only some works [15], [16], [17] dealing with the number of rules simultaneously fired for a given input. The novel index proposed in this paper belongs and thus will extend this reduced group.

B. Visual Analysis of Fuzzy Rule-based Systems

There are not many papers tackling with visual analysis of the inference process of fuzzy systems. Most of them are limited to visual descriptions. Probably, this is due to the well-known linguistic expressivity of such systems what gives prominence to linguistic representations. However, when dealing with complex problems, even when the design is made carefully to maximize interpretability, the number of rules can become huge because of the curse of dimensionality characteristic of fuzzy rule-based systems. In those cases, looking for a plausible linguistic explanation of the inferred output, derived from the linguistic description of the fuzzy knowledge base, is not straightforward. When many rules are fired at the same time for a given input, explaining the inferred output as an aggregation of all the involved rules is not easy.

Some authors [18] have searched for understandable ways of interpreting the system output in terms of describing the inferred output possibility distribution by a set of previously defined linguistic terms along with some linguistic modifiers and connectives. As an alternative, other authors have bet for searching visual explanations of the system output. Ishibuchi et al. [19], [20] established a set of design constraints with the aim of producing groups of rules with only two antecedent conditions that can be plotted in a two-dimensional (2D) space. They look for a visual representation able to explain the output of fuzzy rule-based classifiers to human users. Nevertheless, considering only two antecedents per rule is a strong limitation that may penalize the accuracy of the system.

A complete analysis of visualization requirements for fuzzy systems is provided in [21]. It gives an overview on existing methodologies to yield 2D and 3D graphical representations of fuzzy systems. It comprises visualization of fuzzy data, fuzzy partitions, and fuzzy rules. Different alternatives are available depending on the requirements of the end-user. Moreover, requirements may change according to visualization tasks to perform: Interactive exploration; automatic computer-supported exploration; receiving feedback from users; and capturing users' profiles and adaptation.

The most relevant works to obtain visual representations of multi-dimensional fuzzy rules are those developed by Berthold et al. [22], [23]. They make a mapping from a high-dimensional feature space onto a two-dimensional space which maintains the pair-wise distances between rules. The established mapping also displays an approximation of the rule spread and overlapping. As a result, it is possible to visualize and explore multi-dimensional fuzzy rule bases in a 2D graphical representation. Authors claim such representation yields a user friendly and interpretable exploratory analysis. However, the complexity of the analysis grows exponentially with the number of features and rules to be displayed. In consequence, in complex problems with many rules the interpretation of the resultant graph is not straightforward.

C. Scientograms: Design and Applications

Even though "*constructing a great map of the sciences has been a persistent dream since the Middle Age*" [24], it has been during the last few years when it has become a strong need mainly due to the success of Internet which acts as catalyst. In consequence, isolated research groups are almost non-existent and the number of scientific publications in very multidisciplinary fields has been increased very quickly. Fortunately, the recent advances on computer visualization [25], [26], [27], [28], [29] make feasible high quality and fast visual representations of very large scientific domains.

The term *scientogram* is coined in the specialized literature to make reference to visual science maps, i.e., visual representations of scientific domains. Vargas-Quesada and Moya-Anegón [24] proposed a methodology to create scientograms with the aim of illustrating interactions between authors and papers through citations and co-citations. The basic idea turns up from the notion of paper co-referencing which represents the frequency with which two documents are cited by others. It is possible to talk about author co-citation, journal co-citation, co-citation of classes and categories, etc. Obviously, depending on the kind of co-citation selected the information that can be extracted from the generated maps is different.

The standardized co-citation measure was defined by Salton and Bergmark [30] and is computed by the next equation:

$$MCN(ij) = \frac{Cc(ij)}{\sqrt{c(i) \cdot c(j)}} \tag{1}$$

where Cc means co-citation, c stands for citation, while i and j represent two different entities (authors, documents, journals, categories, institutions, countries, etc).

The combination of entity co-citation and social networks analysis through the use of the Pathfinder algorithm [31] has proved to be able of getting high quality, schematic visualizations of the resultant networks in various fields such as psychology (to represent the cognitive structure of a subject [32]), software development (for debugging of multi-agent systems [33]), or scientometrics (for the analysis of large scientific domains [34]).

The Pathfinder algorithm is in charge of pruning the network defined from the original co-citation matrix (that can be seen as the adjacency matrix of a graph) keeping only the most relevant links. This is essential to make feasible a good visual interpretation of the final Pathfinder networks (PFNETs).

There are many different methods for the automatic visualization of PFNETs. The spring embedder family of methods is the most widely used in the area of Information Science. Spring embedders assign coordinates to the nodes in such a way that the final graph will be pleasing to the eye, and that the most important elements are located in the center of the representation. Kamada-Kawai's algorithm [35] is one of the most extended methods to perform this task. Starting from a circular position of the nodes, it generates networks with aesthetic criteria such as the maximum use of available space, the minimum number of crossed links, the forced separation of nodes, the build of balanced maps, etc. Notice that, the combination of entities co-citation, PFNETs, and Kamada-Kawai makes the entities that most sources share with the rest, tend to situate themselves toward the center.

Concerning the analysis of scientograms, according to [24] there are three main measures of centrality that yield useful information with the aim of identifying the most significant nodes of a PFNET: *Degree of Centrality* (regarding the number of direct links gathering in a node), *Centrality of Closeness* (measuring the distance among nodes), and *Centrality of Intermediation or Betweeness* (looking at nodes that act as link between other nodes contained in the shortest path).

III. Proposal

As said before, this paper proposes a new methodology for visual representation and exploratory analysis of the fuzzy inference process of fuzzy rule-based classifiers (FRBCs). It is based on the generation of FRBC inference maps, so-called fuzzy inference-grams (fingrams) by analogy with the scientograms used for visualizing scientific information. The following subsections explain how to generate fingrams, how to analyze them with the aim of making a exploratory analysis of the fuzzy inference process, and how to derive interpretability measures from the fingrams.

A. Fingram generation

Fingrams show graphically the interaction between fuzzy rules at the inference level in terms of co-fired rules, i.e., rules fired at the same time by a given input. First, from a data set we build a square matrix M that contains all interactions among N rules regarding the proportion of co-fired rules.

$$M = \begin{pmatrix} 1 & a_{12} & \dots & a_{1N} \\ a_{21} & 1 & \dots & a_{2N} \\ \dots & \dots & \dots & \dots \\ a_{N1} & a_{N2} & \dots & 1 \end{pmatrix} \tag{2}$$

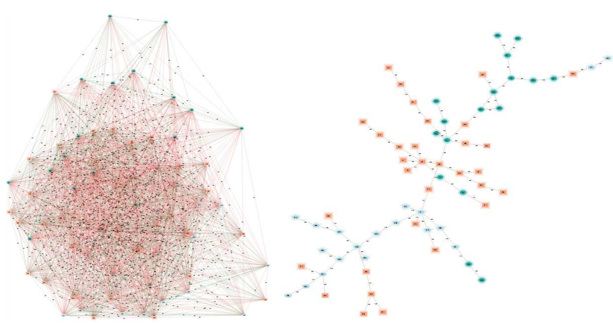
$$a_{ij} = \frac{SFR_{ij}}{\sqrt{FR_i \cdot FR_j}} \tag{3}$$

where a_{ij} is a measure of co-firing inspired on the co-citation measure expressed by equation 1. SFR_{ij} means the number of samples for which rules R_i and R_j are simultaneously fired, while FR_i and FR_j count respectively the total number of samples for which the same rules R_i and R_j are fired, without taking care if they are fired together or not. Notice that, a_{ij} is thus normalized and matrix M is symmetrical.

The number of times a rule is fired is computed in an inferential way for all available data samples. Hence, it is extremely dependant on the goodness (quantity and quality) of the available experimental data. In addition, collecting significant data (covering most possible situations) implies making many experiments which is normally costly in terms of time and ultimately money. Therefore, looking for a more homogeneous view of the inference process based on considering also unknown test samples we have used the Synthetic Minority Over-sampling Technique (SMOTE) [36], which is usually applied to the construction of classifiers from imbalanced datasets, for generating some synthetic test data. As a result, in our context, SMOTE duplicates the number of samples, introducing synthetic data only in those areas of the input space where real experimental data already exist. Notice that, we discarded the introduction of randomly generated synthetic data in the whole input space to avoid the generation of unfeasible data samples that would yield to non-realistic conclusions in our analysis.

Once matrix M is generated, we can use the Kamada-Kawai's algorithm to display the complete fingram. It gives a global view on the interpretability of the FRBC according to the number of nodes (rules), and links (co-firing relations and degrees). Nevertheless, as it happens for the case of scientograms, this initial fingram is normally quite dense and difficult to analyze even for medium size problems. Thus, it is worthy to run the previously mentioned scaling algorithm (Pathfinder[1]) with the aim of pruning the network before printing and exploring the generated fingram. Notice that, the pruned fingram preserves all relevant information at global level thanks to the properties of PFNETs.

Fig. 1 shows an illustrative example of two fingrams made up of 68 rules before and after running Pathfinder. As it can be easily appreciated, it is impossible to see anything on the left side of the figure. Whatever analysis only makes sense after pruning the initial fingram.



(a) Initial fingram (b) Pruned fingram

Fig. 1. Example of fingrams before and afer running Pathfinder.

In our context, the nodes in the fingrams represent fuzzy rules of FRBCs. Rules yielding the same class are depicted by the same symbol (pentagon, rectangle, ellipse, etc). In addition, the number of surrounding lines is related to the complexity of the selected rules (one line means two premises, two lines mean three premises, and so on). Furthermore, edges (links) among nodes represent rule co-firing information. Notice that, thanks to the combination of rule co-firing, PFNETs, and Kamada-Kawai's algorithm, information related to the inference process of FRBCs is displayed in pretty nice scalable fingrams. As a side effect, the most relevant rules, i.e., those rules more often fired tend to be located toward the center of the pruned fingrams, while non-significant rules go to the periphery. Hence, the structure of fingrams is quite informative. On the other hand, redundancies (links among rules of the same class) are painted in green while inconsistencies (links among rules of different classes) are remarked with red color. Moreover, the thickness of links is proportional to their weights. We propose two techniques of pruning a fingram. The first one consists of applying the Pathfinder algorithm directly on the complete fingram. Thus, information related to both redundancies and inconsistencies is taken into account to prune the less salient links. The second one is composed of two steps. Firstly, we remove the redundant links of the complete fingram. Secondly, Pathfinder is applied to the non-redundant graph. In such a way, the pruning of the graph is made only regarding inconsistencies.

Finally, it is important to highlight that our proposal is not affected by the well-known curse of dimensionality problem of fuzzy systems that implies the number of fuzzy rules grow exponentially with the number of inputs. First, nodes represent directly rules instead of premises. Second, PFNETs have been successfully applied to the analysis of large scientific domains representing thousands of co-cited entities [24]. In consequence, fingrams are able to display the interactions among thousands of rules in the form of highly interpretable trees. Even when the number of rules is huge the pruned fingram can be still comfortably viewed by any expert.

B. Fingram exploratory analysis

The analysis of fingrams offers many possibilities. For instance, one can directly analyze its global structure by the exploration of the number and the location of the apparent groups of rules, analyze the respective location of the rules coding for different classes, etc. We would like to remark the following exploratory tasks: Discovering the most significant rules in the knowledge base; and detecting redundancies and/or inconsistencies among fuzzy rules.

On the one hand, the usual Centrality measures that are commonly used in the analysis of scientograms can also be successfully applied to find out the most significant rules within a FRBC. As a first approach, we advocate for the use of the so-called *Degree of Centrality*. This means that we will point out as the most significant rules those corresponding to the nodes that concentrate the larger number of links in a fingram. Remind that they tend to be located toward the center of pruned fingrams.

On the other hand, the interaction among fuzzy rules at inference level is very difficult to appreciate by only reading

[1]We have selected MST-PathFinder [37] a recently published variant of Pathfinder algorithm able to generate large visual science maps in cubic time.

the linguistic description of a FRBC. As a first step, in this contribution we will concentrate on the analysis at rule base level. It depends on the rule description but also on the inference mechanism. Obviously, such analysis is different depending on the kind of problem faced. In classification problems, redundancies and inconsistencies must be handled as conflicts to be solved. Solving redundancies implies removing redundant rules what contributes to get better interpretability. In addition, the interested reader is referred to [38] where there is a detailed explanation of some possible consistency problems along with a methodology to detect and correct such inconsistencies. Of course, from the interpretability point of view it would be desirable to have only one rule that directly yields the right inferred output. Anyway, this may produce a huge number of rules what is also undesirable.

Even when a rule base is fully consistent at linguistic level, there may arise some possible redundancies and/or inconsistencies at inference level because of the rule aggregation procedure made as part of the inference process. Such potential conflicts are difficult to detect mainly because they are partially hidden since they are typically produced by new unknown data that were not taken into account during the learning stage. For instance in classification problems, it may happen that several rules are fired at the same time for a new given input vector and as result several outputs are activated with degrees higher than zero. When two different classes are activated with very similar degrees the situation can be labeled as an ambiguity case. Such situation is not desirable, no matter if the system is (or not) able to yield the right output class, because a slight modification in the input data may yield a wrong output. We can conclude that a FRBC producing many ambiguity cases is a non-reliable system and should be corrected.

Looking at pruned fingrams we can discover redundancies (when the co-fired rules yield the same output class) and inconsistencies (when the co-fired rules yield different output classes). The larger the link weights (co-firing degree computed by equation 3) are, the larger the interaction between rules is, and the larger the degree of redundancy or inconsistency results.

It should be noted that, because of the specific way the pruning and the drawing is done, the most salient links and nodes are likely to be drawn in the center, and those less relevant in the periphery. Thus, those rules that correspond to nodes located in the periphery of a fingram, especially those connected with a low value (the weight of the associated link is small) to the remaining graph, are good candidates to be pruned. This could have an interesting collateral advantage since removing such rules is likely to increase interpretability while keeping almost the same accuracy. We will check it in the experimental section. A basic simplification procedure may consist of finding and removing those non-relevant rules normally located at the periphery of the graph. Moreover, regarding to fingrams we can set a ranking of rules according to their relevance and then running a linguistic simplification procedure like the one proposed in [38].

C. *Interpretability assessment based on fingrams*

We assume that fuzzy partitions are interpretable and the matching among linguistic terms and fuzzy sets is supervised and approved by an expert. Notice that interpretable fuzzy partitions must represent prototypes that are meaningful for the end-user. Then, given a rule format along with an inference mechanism, the system comprehensibility can be evaluated looking only at rule level. Our assumption is the following: The larger the number of simultaneously fired rules for a given input vector, the smaller the comprehensibility of the FRBC.

Fingrams give us information related to the proportion of co-fired rules that should be considered when assessing interpretability. Equation 4 formalizes the Knowledge Base Comprehensibility Index ($KBCI$), our novel proposal of interpretability index:

$$KBCI = \begin{cases} 0, \; if \; \sum_{i=1}^{N} \sum_{j=1}^{N} [(P_i + P_j) \cdot a_{ij}] \geq MAX \\ 1 - \sqrt{\frac{\sum_{i=1}^{N} \sum_{j=1}^{N} [(P_i + P_j) \cdot a_{ij}]}{MAX}}, \; otherwise \end{cases} \quad (4)$$

where N is the total number of rules. P_i and P_j count the number of premises (antecedent conditions) in rules i and j, while a_{ij} is the measure of co-firing for the same rules i and j; it is computed by equation 3. And MAX is a maximum value established to get a normalized measure in the interval [0,1]. It should be fixed by the designer of the FRBC, looking at the maximum number of rules that may be acceptable (by an end-user) for each specific problem according to its inherent complexity (number of inputs, output classes, available training data, etc). According to our experimentations, we suggest setting MAX greater or equal than one thousand times the multiplication of the number of classes (C) by the number of inputs (I) by the number of training samples (T). Hence, $MAX \geq C \cdot I \cdot T \cdot 10^3$. We have set $MAX = 10^7$.

IV. Experimental Analysis

This experimental study deals with the well-known WINE benchmark classification problem whose dataset is freely available at the KEEL[2] (Knowledge Extraction based on Evolutionary Learning) machine-learning repository. It contains 178 instances coming from the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars (3 classes of wines). In addition, the analysis determined the quantities of 13 constituents (*Alcohol*, *Malic acid*, *Ash*, *Alcalinity of ash*, *Magnesium*, *Total phenols*, *Flavanoids*, *Nonflavanoids phenols*, *Proanthocyanins*, *Color intensity*, *Hue*, *OD280/OD315 of diluted wines*, and *Proline*) taken as inputs.

As already said, this analysis considers only FRBCs. In this contribution they have been generated following the HILK (Highly Interpretable Linguistic Knowledge) fuzzy modeling methodology [38], [39]. We have chosen HILK because it is especially thought for making easier the design process of

[2] A free software tool available online at [http://sci2s.ugr.es/keel/].

interpretable FRBCs. To do so, it imposes several constraints (SFPs, global semantics, Mamdani rules [4], etc.) during the design phase. The rule base is made up of rules of form:

$$\textbf{If } \underbrace{X_a \text{ is } A_a^i}_{Premise\ P_a} \textbf{ AND} \ldots \textbf{AND } \underbrace{X_z \text{ is } A_z^j}_{Premise\ P_z} \textbf{ Then } Y \text{ is } C^n$$

where C^n is the selected output class; X_a is the name of the input variable a; and A_a^i represents the label i of such variable. Namely, A_a^i can be one of the elementary terms in the SFP or a composite term defined as a convex hull of adjacent elementary terms corresponding to OR and NOT combinations [40]. These kinds of rules are usually known as DNF rules. Notice that, the absence of an input in a rule means that it is not considered in the evaluation of such rule. This special kind of premises is usually referred as *Don't care* premises [41]. Because several output classes can be activated since several fuzzy rules can be fired at the same time by the same input vector, the winner rule fuzzy reasoning mechanism is considered.

HILK methodology is implemented as part of the free software tool GUAJE[3]. Moreover, the new methodology for visualizing and exploring fuzzy rule bases proposed in this paper is also implemented in that tool. The drawing of the graphs themselves is done using another freeware tool named Graphviz[4] [42].

The rest of this section is devoted to show the utility of the new methodology proposed in this paper through some illustrative examples. Please notice that, probably there are better rule bases for the WINE problem in the fuzzy literature. We do not care about that because our goal is to explain the new methodology instead of looking for the best solution for this specific problem.

As a starting point, Fig. 1 shows fingrams of a FRBC automatically generated with GUAJE for the WINE problem. Uniform strong fuzzy partitions with three triangular fuzzy sets are initially generated for each input. The rule base is made up of 68 rules with a total number of premises (in the following we will refer it by Total Rule Length or TRL) equals to 422. This means an Average Rule Length (ARL) of 6.206 even though the total number of inputs is 13. This FRBC exhibits a good accuracy on the whole original dataset (ACC=0.978) but the Average number of Fired Rules (AFR) is high (AFR=22.489). In consequence, it may be deemed as not very interpretable. Computed on the complete fingram (Fig. 1(a)), the $KBCI$ index is 0.61266, a low value that confirms the feeling one may perceive after visualizing the plotted fingram. From the complete graph, we generated (running the Pathfinder algorithm) a pruned fingram that is detailed in Fig. 1(b). The number of edges decreased consequently, showing a much more interpretable representation of the same FRBC. Notice that this representation can be even simplified furthermore, by eliminating some peripherical rules, as said previously.

[3]A free software tool for generating understandable and accurate fuzzy rule-based classifiers in a Java environment [http://www.softcomputing.es/guaje].

[4]A free software tool available online at [http://www.graphviz.org/]

Then, we have run the linguistic simplification procedure proposed in the HILK methodology as a refining step with the aim of getting better interpretability while preserving accuracy. The simplification affects to both partitions and rules. As a result, a new simplified FRBC is obtained. The number of inputs passes from 13 to 8, but the most impressive reduction is related to the number of rules which drops dramatically from 68 to only 8. Accuracy remains the same (ACC=0.978) while all interpretability indicators are clearly improved (TRL=26, ARL=3.25, AFR=3.792). As expected, the interpretability index is also increased ($KBCI$=0.95433).

Fig. 2 shows the new fingrams (pictures (a), (c), and (d)) of the simplified FRBC along with the detailed list of simplified rules (picture (b)). These fingrams are very illustrative and easy to interpret. Let us explain all the information they provide. Picture (a) plots the complete non-pruned fingram. It shows all interaction among rules and even though the number of rules is tractable this fingram is still quite dense. As explained in the previous section, the next step consists of pruning it by means of Pathfinder, with the aim of highlighting the most significant nodes and links. We consider two cases. First, picture (c) prints the whole pruned fingram. Thus, both redundancies and inconsistencies were considered. Second, picture (d) depicts the fingram resultant of applying Pathfinder only considering inconsistency cases. That is all redundancies were removed from the complete fingram before pruning. In such a way, only the information related to inconsistencies is used to prune the graph.

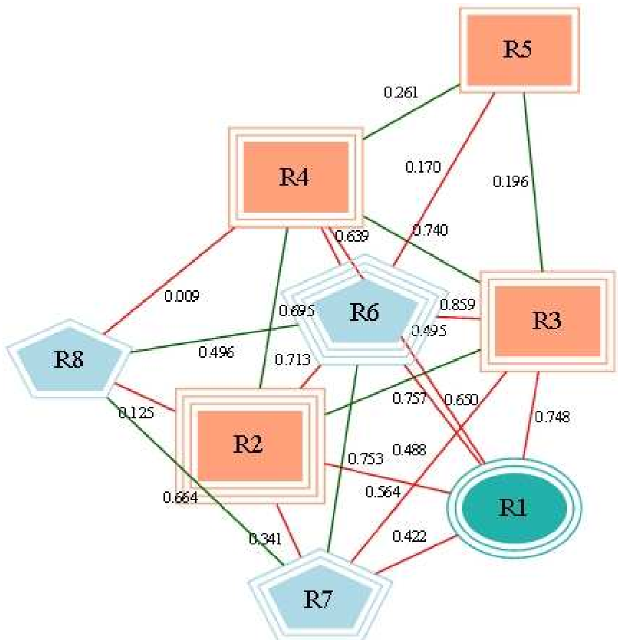
To sum up, picture (a) gives a global overview of all the rule interactions, picture (c) puts the spotlight on the main redundancies and inconsistencies, and picture (d) concentrates only on remarking the most risky inconsistencies. Notice that, picture (d) lets us discover links that were hidden in picture (c) because of the presence of redundant links with high weights. For instance, the link between R5 and R6 appears in picture (d) but not in picture (c). This is because of the higher weight of the links between R4 and R5.

From fingrams in Fig. 2, R5 and R8 are identified as good candidates to be removed with the aim of simplifying even more the FRBC. On the contrary, R3 and R6 are pointed out as key rules. Changes on accuracy and interpretability are almost negligible when removing peripherical rules like R5 (ACC=0.972, TRL=24, ARL=3.429, AFR=3.747, $KBCI$=0.9545) or R8 (ACC=0.972, TRL=24, ARL=3.429, AFR=3.534, $KBCI$=0.95455). However, removing a central rule like R3 (ACC=0.893, TRL=23, ARL=3.286, AFR=3.067, $KBCI$=0.96379) or R6 (ACC=0.949, TRL=20, ARL=2.857, AFR=2.848, $KBCI$=0.97306) has a much more remarkable effect regarding both interpretability and accuracy.

V. Conclusions and Future Works

This paper has introduced a new powerful methodology for exploratory analysis of fuzzy rule-based systems. A first version is already implemented for FRBCs. It can be freely downloaded as part of the free software tool GUAJE at:

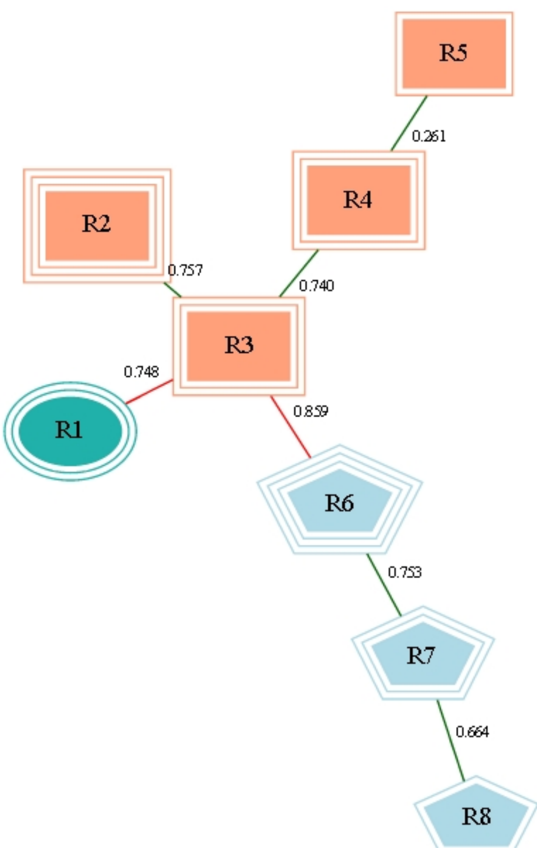
http://www.softcomputing.es/guaje



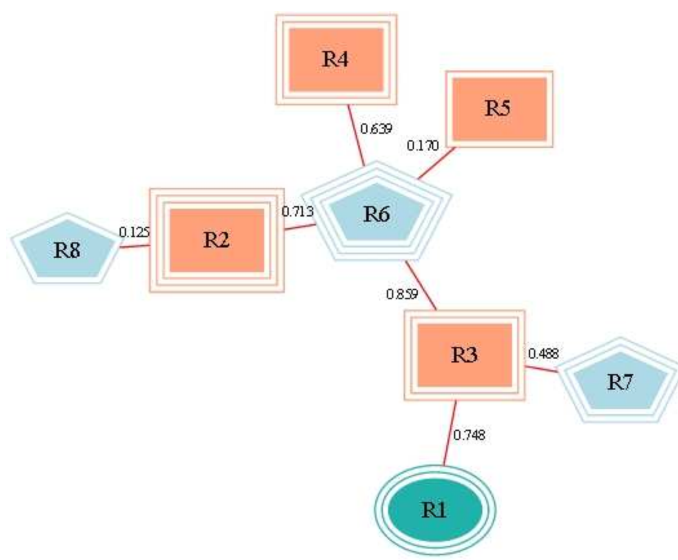
(a) Complete fingram (before running MST-PathFinder)

R1:	**If** Flavanoids is low **AND** Hue is low **AND** OD280/OD315 of diluted wines is low **Then** Class is Wine3
R2:	**If** Malic acid is low **AND** Flavanoids is low **AND** Hue is high **AND** OD280/OD315 of diluted wines is NOT(high) **Then** Class is Wine2
R3:	**If** Flavanoids is average **AND** Proline is low **AND** OD280/OD315 of diluted wines is NOT(low) **Then** Class is Wine2
R4:	**If** Alcohol is low **AND** Flavanoids is average **AND** Proline is average **Then** Class is Wine2
R5:	**If** Alcohol is low **AND** Flavanoids is high **Then** Class is Wine2
R6:	**If** Alcohol is average **AND** Total phenols is low **AND** Flavanoids is average **AND** Color intensity is high **AND** Proline is average **AND** OD280/OD315 of diluted wines is average **Then** Class is Wine1
R7:	**If** Alcohol is high **AND** Flavanoids is average **AND** Proline is average **Then** Class is Wine1
R8:	**If** Flavanoids is average **AND** Proline is high **Then** Class is Wine1

(b) Rule base linguistic description



(c) Pruned fingram (Redundancies and Inconsistencies)



(d) Pruned fingram (Only Inconsistencies)

Fig. 2. Fingrams for a reduced FRBC.

In addition, we have proposed a novel interpretability index that takes into account the comprehensibility of fuzzy systems looking at the correspondence between the linguistic description and the inference process. According to the taxonomy given by [13], the new index tackles with semantic at rule base level where there are almost no proposals in the fuzzy literature. In the future we will extensively validate the methodology and look for other co-firing metrics able to yield additional information about consistency, generality and/or specificity of rules.

ACKNOWLEDGMENT

This work has been partly supported by the European Centre for Soft Computing (ECSC) located at Mieres-Asturias (Spain) and the Spanish Ministerio de Ciencia e Innovación under projects TIN2008-06890-C02-01 and TIN2009-07727, including EDRF funding.

REFERENCES

[1] J. M. Alonso, L. Magdalena, and G. González-Rodríguez, "Looking for a good fuzzy system interpretability index: An experimental approach," *Int. Journal of Approximate Reasoning*, vol. 51, pp. 115–134, 2009.

[2] D. Nauck and R. Kruse, "Obtaining interpretable fuzzy classification rules from medical data," *Artificial Intelligence in Medicine*, vol. 16, pp. 149–169, 1999.

[3] L. A. Zadeh, "Outline of a new approach to the analysis of complex systems and decision processes," *IEEE Trans. on SMC*, vol. 3, pp. 28–44, 1973.

[4] E. H. Mamdani, "Application of fuzzy logic to approximate reasoning using linguistic systems," *IEEE Trans. on Computers*, vol. 26, no. 12, pp. 1182–1191, 1977.

[5] C. Mencar and A. M. Fanelli, "Interpretability constraints for fuzzy information granulation," *Inf. Sciences*, vol. 178, pp. 4585–4618, 2008.

[6] 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*, vol. 159, pp. 3091–3131, 2008.

[7] S. Guillaume, "Designing fuzzy inference systems from data: An interpretability-oriented review," *IEEE Trans. on Fuzzy Systems*, vol. 9, no. 3, pp. 426–443, 2001.

[8] J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, *Interpretability issues in fuzzy modeling*. Studies in Fuzziness and Soft Computing, Springer-Verlag, Heidelberg, 2003, vol. 128.

[9] J. V. de Oliveira, "Semantic constraints for membership function optimization," *IEEE Trans. on SMC. Part A, Systems and Humans*, vol. 29, no. 1, pp. 128–138, 1999.

[10] M. J. Gacto, R. Alcalá, and 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 Trans. on Fuzzy Systems*, vol. 18, no. 3, pp. 515–531, 2010.

[11] C. Mencar, C. Castiello, R. Cannone, and A. Fanelli, "Interpretability assessment of fuzzy knowledge bases: a cointension based approach," *Int. Journal of Approximate Reasoning*, In press, http://dx.doi.org/10.1016/j.ijar.2010.11.007.

[12] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *The Psychological Review*, vol. 63, no. 2, pp. 81–97, 1956.

[13] M. J. Gacto, R. Alcalá, and F. Herrera, "Interpretability of linguistic fuzzy rule-based systems: An overview on interpretability measures," *Inf. Sciences*, In press, 2011.

[14] E. H. Ruspini, "A new approach to clustering," *Information and Control*, vol. 15, no. 1, pp. 22–32, 1969.

[15] J. M. Alonso, L. Magdalena, and O. Cordón, "Embedding HILK in a three-objective evolutionary algorithm with the aim of modeling highly interpretable fuzzy rule-based classifiers," in *Fourth Int. Workshop on Genetic and Evolutionary Fuzzy Systems (GEFS)*, 2010, pp. 15–20.

[16] A. Marquez, F. Marquez, and A. Peregrin, "A multi-objective evolutionary algorithm with an interpretability improvement mechanism for linguistic fuzzy systems with adaptive defuzzification," in *FUZZIEEE*, 2010, pp. 277–283.

[17] F. Cheong and R. Lai, "Constraining the optimization of a fuzzy logic controller using an enhanced genetic algorithm," *IEEE Trans. on SMC. Part B, Cybernetics*, vol. 30, no. 1, pp. 31–46, 2000.

[18] C. Moraga, "Metasemantics of Mamdani systems," *Research report, FSC:2010-13, European Centre for Soft Computing*, pp. 1–14, 2010.

[19] H. Ishibuchi, Y. Kaisho, and Y. Nojima, "Designing fuzzy rule-based classifiers that can visually explain their classification results to human users," in *3rd Int. Workshop on Genetic and Evolutionary Fuzzy Systems (GEFS)*, 2008, pp. 5–10.

[20] ——, "A visual explanation system for explaining fuzzy reasoning results by fuzzy rule-based classifiers," in *NAFIPS*, 2008, pp. 1–6.

[21] B. Pham and R. Brown, "Analysis of visualization requirements for fuzzy systems," in *Int. conference on computer graphics and interactive techniques (GRAPHITE)*, 2003, pp. 181–187.

[22] M. R. Berthold and L. O. Hall, "Visualizing fuzzy points in parallel coordinates," *IEEE Trans. on Fuzzy Systems*, vol. 11, no. 3, pp. 369–374, 2003.

[23] T. R. Gabriel, K. Thiel, and M. R. Berthold, "Rule visualization based on multi-dimensional scaling," in *FUZZIEEE*, 2006, pp. 66–71.

[24] B. Vargas-Quesada and F. Moya-Anegón, *Visualizing the structure of science*. Springer-Verlag, 2007.

[25] C. Chen, *Information visualization and virtual environments*. Springer, 1999.

[26] K. Borner, C. Chen, and K. Boyack, "Visualizing knowledge domains," *Annual Review of Information Science and Technology*, vol. 37, pp. 179–255, 2003.

[27] L. Leydesdorff, "Clusters and maps of science journals based on bi-connected graphs in the journal citation reports," *Journal of Documentation*, vol. 60, pp. 371–427, 2004.

[28] K. Boyack, R. Klavans, and K. Borner, "Mapping the backbone of science," *Scientometrics*, vol. 64, pp. 351–374, 2005.

[29] I. Samoylenko, T.-C. Chao, W.-C. Liu, and C.-M. Chen, "Visualizing the scientific world and its evolution," *Journal of the American Society for Information Science and Technology*, vol. 57, pp. 1461–1469, 2006.

[30] G. Salton and D. Bergmark, "A citation study of computer science literature," *IEEE Trans. on Professional Communication*, vol. 22, pp. 146–158, 1979.

[31] D. Dearholt and R. Schvaneveldt, "Properties of pathfinder networks," in *Pathfinder associative networks: Studies in knowledge organization*, R. Schvaneveldt, Ed. Ablex Publishing Corporation, 1990, pp. 1–30.

[32] R. W. Schvaneveldt, F. T. Durso, and D. W. Dearholt, "Network structures in proximity data," *The psychology of learning and motivation: Advances in research and theory*, no. 24, pp. 249–284, 1989.

[33] E. Serrano, A. Quirin, J. Botia, and O. Cordon, "Debugging complex software systems by means of pathfinder networks," *Inf. Sciences*, vol. 180, no. 5, pp. 561–583, 2010.

[34] C. Chen, *Information visualization: Beyond the horizon.* Springer, 2004.

[35] T. Kamada and S. Kawai, "An algorithm for drawing general undirected graphs," *Information Processing Letters*, vol. 31, no. 1, pp. 7–15, 1989.

[36] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 341–378, 2002.

[37] A. Quirin, O. Cordón, J. Santamaría, B. Vargas-Quesada, and F. Moya-Anegón, "A new variant of the pathfinder algorithm to generate large visual science maps in cubic time," *Information Processing and Management*, vol. 44, pp. 1611–1623, 2008.

[38] J. M. Alonso, L. Magdalena, and S. Guillaume, "HILK: A new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism," *Int. Journal of Intelligent Systems*, vol. 23, no. 7, pp. 761–794, 2008.

[39] J. M. Alonso and L. Magdalena, "HILK++: an interpretability-guided fuzzy modeling methodology for learning readable and comprehensible fuzzy rule-based classifiers," *Soft Computing*, In press, DOI 10.1007/s00500-010-0628-5 2010.

[40] S. Guillaume and L. Magdalena, "An or and not implementation that improves linguistic rule interpretability," in *11th Int. Fuzzy Systems Association World Congress*, Beijing, China, 2005, pp. 88–92.

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

[42] E. R. Gansner and S. C. North, "An open graph visualization system and its applications to software engineering," *Software - Practice and Experience*, vol. 30, no. 11, pp. 1203–1233, 1999.