International Journal of Approximate Reasoning 52 (2011) 501–518

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

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

Interpretability assessment of fuzzy knowledge bases: A cointension based approach

C. Mencar *, C. Castiello, R. Cannone, A.M. Fanelli

Department of Informatics, University of Bari, Bari 70125, Italy

ARTICLE INFO

Article history:
Received 26 April 2010
Revised 18 November 2010
Accepted 19 November 2010
Available online 30 November 2010

Keywords:
Computing with words
Cointension
Interpretability
Fuzzy rule-based classifier
Boolean logic
Logical view

ABSTRACT

Computing with words (CWW) relies on linguistic representation of knowledge that is processed by operating at the semantical level defined through fuzzy sets. Linguistic representation of knowledge is a major issue when fuzzy rule based models are acquired from data by some form of empirical learning. Indeed, these models are often requested to exhibit interpretability, which is normally evaluated in terms of structural features, such as rule complexity, properties on fuzzy sets, partitions and so on. In this paper we propose a different approach for evaluating interpretability that is based on the notion of cointension. The interpretability of a fuzzy rule-based model is measured in terms of cointension degree between the explicit semantics, defined by the formal parameter settings of the model, and the implicit semantics conveyed to the reader by the linguistic representation of knowledge. Implicit semantics calls for a representation of user's knowledge which is difficult to externalise. Nevertheless, we identify a set of properties – which we call "logical view" – that is expected to hold in the implicit semantics and is used in our approach to evaluate the cointension between explicit and implicit semantics. In practice, a new fuzzy rule base is obtained by minimising the fuzzy rule base through logical properties. Semantic comparison is made by evaluating the performances of the two rule bases, which are supposed to be similar when the two semantics are almost equivalent. If this is the case, we deduce that the logical view is applicable to the model, which can be tagged as interpretable from the cointension viewpoint. These ideas are then used to define a strategy for assessing interpretability of fuzzy rule-based classifiers (FRBCs). The strategy has been evaluated on a set of pre-existent FRBCs, acquired by different learning processes from a well-known benchmark dataset. Our analysis highlighted that some of them are not cointensive with user's knowledge, hence their linguistic representation is not appropriate, even though they can be tagged as interpretable from a structural point of view.

1. Introduction

Computing with words (CWW) is a recent paradigm for processing information through words instead of numbers [1–3]. Its inception is justified by the need of providing for a tool to represent and manipulate knowledge in linguistic form rather than numerical. The ability of manipulating words and inferring knowledge that can be expressed in linguistic forms gives added value to intelligent systems, because of their wider applicability in several real-world contexts in which accuracy is not the main (or the only) concern. Hence, CWW may be beneficial in human-centric fields such as medicine, psychology, economics and linguistics, where a main concern is the need of communicating knowledge to users.

* Corresponding author.
E-mail addresses: mencar@di.uniba.it (C. Mencar), castiello@di.uniba.it (C. Castiello), raffaelecannone@di.uniba.it (R. Cannone), fanelli@di.uniba.it (A.M. Fanelli).

0888-613X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2010.11.007

Historically, the approach used for designing expert systems is mainly based on symbolic representation of knowledge derived by a syntactic inference process. Indeed an appropriate choice of symbols leads to a linguistic – usually structured – representation that can be read by users. On the other hand, CWW operates at the semantic level, i.e. inference is carried out by taking into account the semantic definition of words (and their connectives). These two approaches are equivalent in the case of Boolean semantics, i.e. when words refer to sets of objects and connectors correspond to set operations. Unfortunately, natural language is very far from being expressible as Boolean expressions: fuzziness is an intrinsic feature of most linguistic terms that do not belong to mathematics. As a consequence, the utility of expert systems is strictly limited in contexts where knowledge can be expressed by extremely sharp concepts. CWW comes into play when there is the need of representing and manipulating knowledge where fuzziness is an important feature. Fuzziness, indeed, can be captured by using fuzzy sets, and fuzzy knowledge can be manipulated by fuzzy set theory (FST), which is the formal underpinning of CWW.

The greatest enhancement of CWW over FST is mainly methodological. In CWW emphasis is put on the representation of knowledge, while FST is used as a tool for knowledge manipulation and inference. Hence, interpretability of knowledge is a necessary requirement in CWW. CWW does not directly address interpretability, being focused on the structures and rules for representing and manipulating knowledge. However, granting interpretability is a main issue when designing CWW-based systems, since the lack of interpretability damages all the benefits of CWW. Without interpretability of knowledge, purely numerical methods can be an effective alternative.

Interpretability is an ill-posed property, as there are several main issues to be addressed concerning its definition, its achievement and its assessment. Whilst several works in literature try to capture the property of interpretability with a collection of constraints (both crisp and fuzzy), here we adopt a more general approach by considering the notion of "cointension", firstly defined by Zadeh in [4]. Roughly speaking, cointension can be viewed as a relation between concepts such that two concepts are cointensive if they refer to almost the same object. Thus, in our view, a knowledge base is interpretable if its semantics is cointensive with the knowledge acquired by the user.

The point of departure of our approach relies on a close involvement of user understanding in the definition of interpretability. This could allow for a more effective assessment of interpretability with respect to the structural approach based on interpretability constraints, whose fulfilment cannot assure that the knowledge base is actually cointensive with user's knowledge. Indeed, interpretability constraints are mainly based on common-sense and there is no agreement on the minimal set of constraints to be used. Finally, there is no guarantee that any set of constraints could be an exhaustive characterisation of interpretability.

By proposing an approach based on interpretability as cointension, we try to cast interpretability, intended as "ability to read and understand", in the realm of semantics. Thus, assessing interpretability mainly concerns a comparison between the semantics of a knowledge base and the semantics of the knowledge acquired by a user after reading and understanding the knowledge base. The semantics of a knowledge base (we call it *explicit semantics*) is completely specified by the involved fuzzy sets and operators used for inference. On the other hand, the semantics acquired by users when reading the knowledge base (we call it *implicit semantics*) is much more difficult to externalise. Nevertheless, common features can be identified for both implicit and explicit semantics. By exploiting such common features it is possible to analyse cointension and hence interpretability.

In assessing interpretability of knowledge bases that are represented in a linguistic form, we identify a common feature shared by implicit and explicit semantics derived from a fuzzy knowledge base which we call "logical view". The logical view is here intended as the set of properties of propositional calculus which are assumed to hold both in implicit and explicit semantics. The validity of the logical view is tested through the application of a minimisation algorithm on the fuzzy knowledge base. By testing the validity of the logical view in the explicit semantics, we are able to assess the interpretability of a fuzzy knowledge base in the sense of cointension. We limit our argumentation to fuzzy rule-based classifiers (FRBCs).

The paper is organised as follows: in the next section, the problem of defining and assessing interpretability is analysed. In Sections 3 and 4 the proposed approach for interpretability assessment is described in detail. Successively, in Section 5 we report an experimentation concerning ten knowledge bases for the same classification problem (Wine data) obtained through a system that preserves a number of interpretability constraints. In the conclusive part of the paper we address points of strength and weakness of the proposed approach, along with some remarks on future developments. Additionally, for a self-contained discussion the paper is completed with an appendix containing the description of the minimisation algorithm.

2. The problem of interpretability assessment

Interpretability assessment should be regarded as a major issue in the field of fuzzy knowledge-based system modelling. However, a proper evaluation of interpretability appears to be a controversial problem, since the definition of interpretability eludes any formal characterisation.

2.1. Interpretability definition

In [5], an attempt to provide a definition of interpretability is made, in the form of the following "Comprehensibility Postulate" (CP):

The results of computer induction should be symbolic descriptions of given entities, semantically and structurally similar to those a human expert might produce observing the same entities. Components of these descriptions should be comprehensible as single "chunks" of information, directly interpretable in natural language, and should relate quantitative and qualitative concepts in an integrated fashion.

It should be observed that the above postulate has been formulated in the general area of machine learning. Nevertheless, the assertion made by Michalski has important consequences in knowledge-based fuzzy modelling and CWW [6].

The key point of the CP is the human-centrality of the results of a computer induction process. According to the CP, results of computer induction should be described symbolically. Symbols are necessary to communicate information and knowledge, hence pure numerical methods, including neural networks, are not suited for meeting interpretability unless an intepretability-oriented post-processing of resulting knowledge is performed [7,8].

In its essence, the CP anticipates what has been successively called "Granular Computing" [9]. According to the CP, indeed, symbols should represent chunks of information – a synonym of information granule – that is a group of data tied together by some kind of semantic relationships (proximity, similarity, etc.) [10].

Aggregating data into information granules is a fundamental cognitive activity of human beings. As pointed out by Zadeh, human mental activities can be classified into: (a) granulation; (b) organisation; (c) causation [10]. Information granulation, therefore, represents a fundamental activity in computer induction, and the resulting information granules should be tagged with appropriate symbols. In order to be interpretable, such symbols should be directly interpretable in natural language. This does not come down to the fact that symbols should be chosen from a natural language vocabulary, but deeper implications are involved. In particular, the CP requires the *interpretation* of symbols to be in natural language. This is a requirement on the *semantics* of symbols and relations between them, i.e. on the information granules they denote. Therefore, in order to be understandable, information granules resulting from computing processes should conform with concepts a human can conceive. Furthermore, natural language terms convey an underlying semantics (which depends also on the context), that is shared among all human beings speaking that language. As a consequence, a symbol coming from natural language can be used to denote an information granule only if the underlying semantics of the symbol highly matches with the semantics characterised by the information granule.

Models conforming to the CWW paradigm embody a knowledge base that is defined by composition of linguistic terms. This terms composition hides the model's working engine which is purely mathematical (fuzzy sets, t-norms, s-norms, etc.), so it can be associated with a form of explicit semantics. The behaviour of the model is determined by inference, which is carried out on the basis of such semantics and of an inferential apparatus. Moreover, the linguistic labels convey another kind of semantics, evoked by the specific meaning of the involved communicative terms, which is associated with the cognitive processes performed by the user while reading the rules.

Taking into account the previous considerations, we propose a different point of view for the interpretability definition, which is based on the concept of cointension. The notion of cointension has been introduced in [4] as a semantic relation between concepts:

In the context of modeling, cointension is a measure of proximity of the input/output relations of the object of modeling and the model. A model is cointensive if its proximity is high.

We use cointension to relate the explicit semantics defined by information granules with the underlying implicit semantics held by symbols. The rationale behind these ideas comes from the observation that a knowledge base represents the linguistic interface of a CWW model to the user. For an interpretable knowledge base, the user should be able to understand the behaviour of the model by simply observing its linguistic representation. In this sense, we propose the following definition of interpretability:

A knowledge base is interpretable if the explicit semantics embedded in the model is cointensive with the implicit semantics inferred by the user while reading the rules.

The above definition is grounded on a relationship between the knowledge base of a model and the knowledge held by a user. However, there is a second fundamental point of view, where the model is seen as a representation of the input/output relationship sampled in a dataset. We argue that this latter viewpoint generally relates cointension with accuracy, while our peculiar acceptation of cointension (in the way it is specified in the previous definition) is actually translated in what we call "interpretability". Purely numerical models only focus on accuracy, but CWW models should take into account both viewpoints, that are complementary and share the explicit semantics, i.e. the model in its mathematical expression. A good CWW model should be cointensive both in terms of interpretability and accuracy. This is not a trivial assertion, since many interpretability-oriented models take into account only the first viewpoint. Such an approach is criticisable [11], since interpretable but inaccurate models are as useless as very accurate but incomprehensible models.

2.2. Interpretability assessment

Our approach for interpretability assessment in the realm of fuzzy modelling relies on the notion of cointension, which helps to deal with the very core of the matter. We intend to face the question: "How to define interpretability criteria related

to the evaluation of a fuzzy rule-based model?" or, equivalently, "Given a collection of fuzzy rule-based models, how to rank all of them in order to assist the choice for the most interpretable one(s)?".

Some usual answers for those questions rely on considerations about the basic structure of the involved fuzzy rule-based models. This leads to the formulation of a number of interpretability constraints, i.e. a set of properties (both crisp and fuzzy) that must be fulfilled. Many approaches have been proposed for interpretability-driven design of fuzzy models based on such constraints [12–17]. Interpretability constraints can be stratified at different levels of modelling: they can be found for fuzzy sets, for linguistic variables, for fuzzy partitions and for entire knowledge bases [18]. At the lowest levels, interpretability constraints are required to legitimate the attachment of linguistic labels to fuzzy sets so as to form semantically sound linguistic variables [19]. These constraints usually include normality, convexity, distinguishability, coverage, etc. (see [20] for a survey of interpretability constraints and their motivations).

Interpretability constraints define the structural characteristics of a fuzzy rule-based model. Hence they can be used to evaluate interpretability by verifying if (or to what degree) such constraints are valid for a model. Some approaches for interpretability evaluation at the level of fuzzy sets and partitions can be found in literature. In [21] an index is proposed to evaluate ordering of fuzzy sets in partitions. This index is used within a multi-objective genetic algorithm to restrict the search space of fuzzy rule-based models. In [22,23] the proposed interpretability index sums up different sub-indexes, each evaluating a specific structural feature, such as distinguishability, coverage and position of fuzzy sets. In [24] an index is defined to preserve the semantic integrity of fuzzy sets when a rule-based fuzzy model undergoes an optimisation process through a multi-objective genetic algorithm. Also, the complexity of the rule base is minimised so as to improve its readability.

In many cases, evaluation of interpretability at fuzzy set level is reduced to assess similarity of fuzzy sets so as to maximise their distinguishability. Several approaches have been proposed to derive fuzzy models that minimise similarity, often through genetic approaches such as Genetic Algorithms [25,26], Evolution Strategies [27], Symbiotic Evolution [28], Coevolution [29,30], Multi-Objective Genetic Optimisation [31]. Alternatively, distinguishability improvement is realised in separate learning stages, often after some data-driven procedure like clustering, in which similar fuzzy sets are usually merged together [32,33]. When the distinguishability constraints must be included in less time consuming learning procedures – like in neural learning – similarity measures are no longer used. In [32], after a merging stage that uses similarity, fine tuning is achieved by simply imposing some heuristic constraints on centres and width of membership functions. In [34], a distance function between fuzzy sets (restricted to the Gaussian shape) is used in the regularisation part of a RBF[1] cost function. In [35] a complex distance function is used to merge fuzzy sets. In [13,36,37] the possibility measure is adopted to evaluate distinguishability.

From the perspective of a higher level of modelling, which is the level of knowledge base, interpretability is often related with such values as the number of rules or the number of fuzzy sets per rule, with systems regarded as interpretable as long as those features are low in cardinality. As an example, in [38] interpretability is evaluated in terms of number of rules, total rule length and average rule length, while in [39] rules affect interpretability according to three quantities: (i) the total number of rules, (ii) the average number of firing rules, and (iii) the number of weighted rules. Different approaches try to evaluate interpretability both at the lowest and highest levels. In [40] interpretability of a FRBC is evaluated by combining the complexity of a classifier (measured as the number of classes divided by the total number of premises), the number of labels and the coverage degree of the fuzzy partition. In [41] the previous index is further improved, by taking into account six structural features: i. number of rules, ii. total number of premises, iii. number of rules using one input, iv. number of rules using two inputs, v. number of rules using three or more inputs, vi. total number of labels defined per input. In [42] interpretability is evaluated on Takagi–Sugeno–Kang models, and it is expressed as the ability of each rule to define a linear local model of a non-linear function; penalised learning is used to balance interpretability and accuracy. In [43] a number of FRBCs have been evaluated in terms of interpretability by several volunteer users: the experimental results suggest that simpler FRBCs are preferred over complex ones. In a successive study, the authors propose a flexible index that takes into account user preferences in evaluating interpretability of a fuzzy system [44].

The assertion that interpretability and accuracy are conflicting properties is quite often encountered in the literature (see, e.g. [29,38,45–49] or papers in [50]): roughly speaking, accurate models are usually not interpretable and vice-versa. This conflict is mainly due to the inclusion of simplicity (intended as antinomy of complexity) as a requirement for interpretability. Simplicity is indeed required to make a knowledge base comprehensible because of our limited ability to store information in our brain's short-term memory [51]. However, simplicity imposes a heavy bias on models, that could be unable to accurately represent the input/output relationship. We partially subscribe to this point of view: it is obvious that simple models are more comprehensible than complex ones. However, this should not preclude to define interpretable models, even when the relationship to be modelled is very complex. A number of means are usually employed for organising knowledge, dealing with its complexity and making it comprehensible (including hierarchies, graphs, abstraction, analogies and so on). All of them represent solutions to cope with complexity by exploiting our ability to organise knowledge in complex cognitive structures. In this sense, interpretability and accuracy can be thought as orthogonal characterisations for a fuzzy model. We argue that interpretable and accurate models are possible when conceived on the basis of suitable solutions for dealing with complexity, even if their study is rarely explored in the field of interpretability research and resides out of the main scopes of this paper.

[1] Radial Basis Function network.

Constraint-based approaches for evaluating interpretability (referred to as structural approaches in the introductory section) are mainly based on "common sense". We consider this kind of analysis to be somewhat reductive: a couple of arguments can be brought forward in this respect. From an ontological point of view, the previously mentioned measures appear to be more suitable in addressing complexity of a fuzzy model rather than interpretability; in turn, complexity could be related to the specific problem described by the fuzzy model. In other words, from the structural analysis of a knowledge base we should expect to gain information concerning the complexity of the underlying problem (with complicated problems associated with complex knowledge base structures), but no additional hint about the interpretability of the fuzzy model. From an epistemic point of view, establishing a connection between the comprehensibility of a fuzzy rule base and its basic structure measurements is somewhat troublesome. In fact, to a certain extent this kind of evaluation could collapse to a mere thresholding process, where a model is regarded as interpretable when, say, the number of rules (or fuzzy sets) is kept under some reasonable value. In this way, it is straightforward to observe that the interpretability assessment problem can be thought as an automatic procedure, which could be accomplished by any machine without referring to human judgement. Even the problem of ranking a pool of fuzzy rule-based systems in terms of their interpretability simply converges to a single, uncontestable and predictable solution. This appears to be a strong contrast with the inherent difficulty and the human-centric character of the interpretability property. Complexity evaluation plays its important role to discriminate among several models provided that they have been previously qualified as interpretable in the sense of cointension between explicit and implicit semantics.

As opposite to structural approaches, our definition of interpretability relies on a verified cointension between the explicit semantics expressed by the mathematical model of the knowledge base and the implicit semantics inferred by the user. It is noteworthy to underline how this definition does not leave aside the hermeneutic role of the human discernment in assessing the interpretability of a knowledge base. However, the task of evaluating the cointension between a fuzzy model and the user's knowledge is not a trivial one, since we are willing to draw a comparison between a mathematical apparatus (which can be deeply explored in its formal description) and a cast of human mind (which does not lend itself to investigation, other than qualitative appraisal). In other words, management of implicit semantics is not compatible with a formal investigation.

Out of necessity, we directed our work toward a formalisation of the user's knowledge by considering the overlap between the human cognitive structures and an ideally interpretable fuzzy rule-based model. Such an overlap can be reasonably approximated by the *logical view*, which is the propositional structure of the rules in the knowledge base, responding to the laws of formal logics both for the fuzzy rule-based inference and the user thinking. Once the overlap has been identified, it is possible to make it explicit and to verify if it is applicable to a specific fuzzy rule-based model.

3. The proposed approach – Rationale

To clarify the rationale of the proposed approach for assessing interpretability cointension, we firstly refer to an explanatory example involving two communicating (human) actors, A and B. In particular, we can think of A as a scholar communicating some pieces of information to B by adopting a linguistic structure that is supposed to represent her own knowledge. The actor B, on the other hand, stands as an apprentice whose goal is to achieve some expertise by interpreting the information coming from A. In order to be comprehensible, A chooses linguistic terms whose semantics is deemed as cointensive as possible with B's knowledge. However, it is not necessary that the semantics of linguistic terms in A's mind perfectly match the semantics in B's mind: a high overlapping of semantics is enough for enabling knowledge communication. In any case, there are some premises which should be taken for grant in order to prepare the way for cointension: both the scholar and the apprentice must share similar environment, language, culture, experiences, etc. Therefore, cointension can be achieved if A and B share similar cognitive structures. To assess the scholar's comprehensibility we turn to the apprentice's capability. In other words, the actor B should prove to be able to practice in a suitable way what she learnt.

The illustrative scenario described above highlights a number of issues which can be likewise reported in the context of the interpretability assessment. Specifically, some key points must be taken into account:

(1) some premises are necessary in order to make cointension feasible;
(2) cointension between different semantics is evaluated by comparing partial overlapping of the cognitive structures involved;
(3) interpretability assessment relies on accuracy estimation, by verifying the actual comprehension of the communicated information.

Keeping in mind the aforementioned key points, we proceed to discuss the interpretability of fuzzy knowledge bases, drawing an analogy with the comprehensibility evaluation task involving human actors.

As concerning the premises of cointension, we observe that a fuzzy knowledge base is composed by linguistic rules. For our purposes, it is important that labels denoting linguistic values for each variable can be interpreted as well-distinguished elementary concepts. The absence of this requisite, in fact, would undermine the possibility of a hermeneutic reading of the rules by the human user, thus compromising the evaluation of cointension. The fulfilment of this kind of cointensive premises is achieved by imposing basic interpretability constraints over the fuzzy knowledge base under study at the lowest levels of fuzzy sets and linguistic variables. The minimal set of such constraints should include normality, convexity and

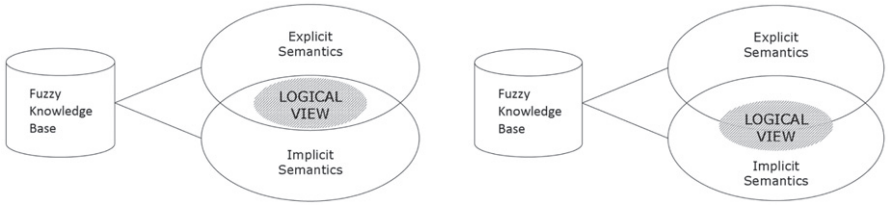
distinguishability. Normality ensures that at least one element inside the range has full membership, thus standing as the prototype for the concept described by the fuzzy set. Convexity requires that membership degrees decrease as the distance of elements from prototypes increases. This property is necessary for expressing elementary concepts, whose semantics is related to the similarity of elements to prototypes that fulfil the concept. Finally, distinguishability ensures enough distinctness among fuzzy sets, so that different linguistic labels can be attached to them. In this way, it can be avoided the misleading labelling of highly overlapping fuzzy sets, which would produce incoherent results (because the involved fuzzy sets would express almost the same concept).

As concerning the second key point, it should be underlined that our approach for interpretability evaluation relies on determining cointension between the explicit semantics embedded into a fuzzy knowledge base and the implicit semantics inferred by the user. However, differently from a context where only human actors are involved, it is somewhat troublesome to identify the bases for evaluating this kind of cointension, due to the inherent different natures of the formal mathematical models and the human cognitive representations. Nevertheless we note a strict affinity of a fuzzy rule base to logical propositions. Actually, rules are constructed so as to resemble propositions, in order to be understood by users. As a consequence, we recognise the logical view of rules as a cognitive structure shared by users and fuzzy knowledge bases. By doing so, we are able to translate the interpretability evaluation problem into a formal process, and we can exploit the logical view as the common ground to analyse both the implicit and the explicit semantics. Being like propositions, rules can be transformed by using logical operators without any change of the semantics. Actually, some distortions may be expected in the process: these are due to the adoption of fuzzy sets to define the explicit semantics of the rules. However, such distortion should be limited to ensure preservation of the shared logical view between the fuzzy knowledge base and the user. After the application of the truth-preserving operators, we obtain a knowledge base which may be different from the original one. Anyway, the preservation of the logical view lets us review the novel rule-based model as a suitable candidate for evaluating cointension with the explicit fuzzy knowledge base semantics, in place of the original knowledge base. If the logical view holds for a specific fuzzy knowledge base, indeed, the semantics of the transformed knowledge base must be almost equivalent to the original one. If this is not verified, we deduce that the operators employed during the minimisation process do not preserve the explicit semantics which, in turn, cannot be represented in terms of logical propositions. This means that logical view is not held by the fuzzy knowledge base and, hence, the model is not interpretable from our cointensive point of view.

The third key point concerns the interpretability assessment, which can be performed in terms of semantic equivalence between the original knowledge base and its transformed version. This is the moment to rely on accuracy: similarly to the apprentice who is called to put in practice the information acquired from the scholar's lesson, the newly obtained knowledge base has to be tested for the sake of comparison with the accuracy capability of the original fuzzy rule base. The test results will indicate if the logical view is applicable to the semantics of rules, thus providing a response to the question of the interpretability assessment.

We compare the two rule bases on the basis of their accuracy: if they do not differ to much, we recognise that the logical view of rules is correct. With reference to Fig. 1(a), this means that the logical view represents a proper intersection between the explicit and implicit semantics related to the fuzzy rule base. In this sense, cointension with the user's knowledge is verified and the fuzzy knowledge base can be deemed interpretable. On the other hand, if the two rule bases are characterised by notably different accuracy values, then the logical view of the fuzzy knowledge base is not compatible with the explicit semantics of fuzzy rules. Again, such a situation can be observed in Fig. 1(b), where the logical view cannot be intended as a proper intersection of explicit and implicit semantics. Therefore, the knowledge base is not cointensive with user's knowledge and it can be deemed as not interpretable. This means that any attempt at reading the linguistic rules would be misleading: the fuzzy knowledge base is indeed a kind of "grey box", whose accuracy capability only relies on the mathematical configuration of its parameters.

At the end of the evaluation process we can draw some conclusions about the accuracy and the interpretability of a fuzzy knowledge base. The plot in Fig. 2 is intended to correlate accuracy vs. interpretability of a fuzzy knowledge base. More properly, the y-axis denotes the accuracy component (in terms of error values of the original knowledge bases) and the x-axis denotes the interpretability component (in terms of variation of error values registered for the transformed knowledge bases with respect to the original ones). Of course, the correlation reported in the plot assumes that the error values of the



(a) Fuzzy knowledge base deemed as interpretable

(b) Fuzzy knowledge base deemed as not interpretable

Fig. 1. The results of the interpretability assessment process based on the cointension analysis between explicit and implicit semantics: (a) interpretable fuzzy knowledge base and (b) not interpretable fuzzy knowledge base.

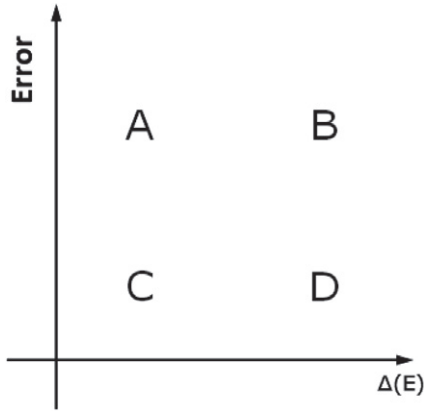


Fig. 2. Categorisation of fuzzy knowledge bases at the end of the interpretability evaluation process.

original and the reformulated fuzzy knowledge bases have been derived by testing over the same datasets. The original fuzzy knowledge base, whose interpretability is under examination, may be situated into one of the regions *A*, *B*, *C*, *D*. According to our definition of interpretability, the following populations can be defined for each one of the regions:

- inaccurate but interpretable fuzzy knowledge bases are included in region *A*;
- inaccurate and not interpretable fuzzy knowledge bases are included in region *B*;
- accurate and interpretable fuzzy knowledge bases are included in region *C*;
- accurate but not interpretable fuzzy knowledge bases are included in region *D*.

Obviously, the regions are not precisely bound but are determined by a fuzzy trade-off between accuracy and interpretability. It can be argued that an interpretable fuzzy knowledge base is not accurate of necessity: in principle, it could be possible to organise in a comprehensible fashion some pieces of erroneous information. Moreover, it should be observed that no mention has been made of the complexity of a fuzzy knowledge base while evaluating its interpretability. This is in agreement with the previously underlined assumption that some of the commonly accounted features (such as the cardinality of a fuzzy rule base, the number of involved fuzzy sets and so on) are not suitable for assessing interpretability as cointension and they should be related to the complexity of the underlying classification problem. Further post-processing could be applied to organise complexity in a readable form; however, this topic is outside of the scope of this paper.

4. The proposed approach – Formalisation

In this section, the entire interpretability assessment process is formalised. Specifically, we develop a four-stage strategy that allows the interpretability evaluation of a fuzzy rule base in terms of cointension. We restrict our argumentation to fuzzy rule-based classifiers (FRBCs) whose rules, described in natural language, resemble logical propositions so that they can be understood by users. As a consequence, the propositional view of rules stands as the cognitive structure shared by the user and the FRBC so as to apply the logical view in order to analyse the cointension of both explicit and implicit semantics.

Before describing the four steps of our strategy, let us introduce the formalisation of FRBCs. We define a classifier as a system computing the following function:

$$f : \mathbf{X} \longrightarrow \Lambda, \tag{1}$$

where $\mathbf{X} \subseteq \mathbf{R}^n$ is an n-dimensional input space, and $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_c\}$ is a set of class labels. If a dataset D of pre-classified data is given, i.e.

$$D = \{(\mathbf{x}_i, l_i) | \mathbf{x}_i \in \mathbf{X},\ l_i \in \Lambda,\ i = 1, 2, \ldots, N\}, \tag{2}$$

then the classification error can be computed as:

$$E(f, D) = \frac{1}{N} \sum_{i=1}^{N} (1 - \chi(l_i, f(\mathbf{x}_i))), \tag{3}$$

being $\chi(a, b) = 1$ iff $a = b$ and 0 otherwise.

A FRBC is a system that carries out classification (1) through inference on a knowledge base. The knowledge base includes the definition of a linguistic variable for each input. Thus, for each $j = 1, 2, \ldots, n$, linguistic variables are defined as[2]:

$$V_j = (v_j, X_j, Q_j, S_j, I_j), \tag{4}$$

[2] For the sake of simplicity we do not take into account the generative grammar for the linguistic values, as originally proposed by Zadeh in [52].

being:

- v_j the name of the variable;
- X_j the domain of the variable (it is assumed that $\mathbf{X} = X_1 \times X_2 \times \cdots \times X_n$);
- $Q_j = \{q_{j1}, q_{j2}, \ldots, q_{jm_j}, \text{ANY}\}$ is a set of labels denoting linguistic values for the variable (e.g. SMALL, MEDIUM, LARGE, ANY);
- $S_j = \{s_{j1}, s_{j2}, \ldots, s_{jm_j+1}\}$ is a set of fuzzy sets on X_j, $s_{jk} : X_j \to [0, 1]$;
- I_j associates each linguistic value q_{jk} to a fuzzy set s_{jk}. We will assume that $I_j(q_{jk}) = s_{jk}$ and $I_j(\text{ANY}) = s_{jm_j+1}$.

As asserted in the previous section, we require that the normality, convexity and distinguishability constraints must be imposed on the FRBC, as a necessary condition for the successive cointension evaluation process. The three constraints can be respectively formalised as:

$$\forall j \forall s \in S_j : \max_{x \in X_j} s(x) = 1, \tag{5a}$$

$$\forall j \forall s \in S_j : x < y < z \to s(y) \geq \min\{s(x), s(z)\}, \tag{5b}$$

$$\forall j \forall s', s'' \in S_j : \frac{|s' \cap s''|}{|s' \cup s''|} \leq \sigma \tag{5c}$$

(usually, $\sigma = 0.5$). Additionally, we assume that each linguistic variable contains the linguistic value "ANY" associated to a special fuzzy set $s \in S_j$ such that $s(x) = 1, \forall x \in X_j$.

The knowledge base of a FRBC is defined by a set of R rules. Each rule can be represented by the schema:

$$\begin{aligned} &\text{IF } v_1 \text{ IS } [\text{NOT}]\ q_1^{(r)} \text{ AND } \cdots \text{ AND } v_n \text{ IS } [\text{NOT}]\ q_n^{(r)} \\ &\text{THEN } \lambda^{(r)}, \end{aligned} \tag{6}$$

being $q_j^{(r)} \in Q_j$ and $\lambda^{(r)} \in \Lambda$. Symbol NOT is optional for each linguistic value. If for some j, $q_j^{(r)} = \text{ANY}$, then the corresponding atom "v_j IS ANY" can be removed from the representation of the rule.[3]

Inference is carried out as follows. When an input $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is available, the strength of each rule is calculated as:

$$\mu_r(\mathbf{x}) = s_1^{(r)}(x_1) \otimes s_2^{(r)}(x_2) \otimes \cdots \otimes s_n^{(r)}(x_n), \tag{7}$$

being $s_j^{(r)} = \nu_j^{(r)}(I_j(q_j^{(r)}))$, $j = 1, 2, \ldots, n$, $r = 1, 2, \ldots, R$. Function $\nu_j^{(r)}(t)$ is $1 - t$ if NOT occurs before $q_j^{(r)}$, otherwise it is defined as t. The operator $\otimes : [0, 1]^2 \to [0, 1]$ is usually a t-norm, such as the minimum or product functions.

The degree of membership of input $\mathbf{x}$ to class λ_i is computed by considering all the rules of the FRBC as:

$$\mu_{\lambda_i}(\mathbf{x}) = \frac{\sum_{r=1}^{R} \mu_r(\mathbf{x}) \chi(\lambda_i, \lambda^{(r)})}{\sum_{r=1}^{R} \mu_r(\mathbf{x})}. \tag{8}$$

Finally, since just one class label has to be assigned to the input $\mathbf{x}$, the FRBC assigns the class label with highest membership (ties are solved arbitrarily):

$$f_{FRBC}(\mathbf{x}) = \lambda \Rightarrow \mu_\lambda(\mathbf{x}) = \max_{i=1,2,\ldots,c} \mu_{\lambda_i}(\mathbf{x}). \tag{9}$$

4.1. Definition of the truth tables

In the first step of our strategy, a FRBC is transformed in several true tables (one for each class) without any semantic change. This transformation is justified under the Closed World Assumption (CWA), which is almost always postulated when dealing with fuzzy predictive models, since they must provide an output for any submitted instance. According to CWA, a rule base expresses complete knowledge, i.e. information enclosed in the rule base is necessary and sufficient to perform classification; therefore, for each input instance at least one rule antecedent is verified. We also assume that the rule base is consistent, hence for each input instance the rules fired with maximum strength assign the input to the same class. Under these requirements, the rule base can be rewritten in the canonical disjunctive form, which can be represented into as many truth tables as the number of classes.

Each rule of a FRBC is therefore seen as a proposition, i.e. a combination of propositional variables that is considered true for a class. For each class label $\lambda_i \in \Lambda$, a truth function π_i is defined on the propositional variables defined for the FRBC such that, for each rule r:

$$\pi_i\left(\chi_{11}^{(r)}, \ldots, \chi_{1m_1}^{(r)}, \ldots, \chi_{n1}^{(r)}, \ldots, \chi_{nm_n}^{(r)}\right) = \chi\left(\lambda_i, \lambda^{(r)}\right). \tag{10}$$

[3] The sequence NOT ANY is not allowed.

Inputs $\chi_{jk}^{(r)}$ may assume values 0, 1 or X ("don't care"). Their definition depends on the value of the linguistic variable V_j in the rth rule. In particular, for each input j:

(1) If the atom "v_j IS $q_j^{(r)}$" occurs in r-th rule, then $\chi_{jk}^{(r)} = \chi(q_j^{(r)}, q_{jk})$ for each $k = 1, 2, \ldots, m_j$. This case makes the assignment of a linguistic value to a linguistic variable exclusive, i.e. whenever a linguistic value is assigned, all other linguistic values are not assigned;
(2) If the atom "v_j IS NOT $q_j^{(r)}$" occurs in the r-th rule, then $\chi_{jk}^{(r)} = 0$ if $q_j^{(r)} = q_{jk}$ and $\chi_{jk}^{(r)} = X$ otherwise. This case specifies that whenever a linguistic value is *not* assigned, then any other value might be assigned;
(3) If the atom "v_j IS ANY" occurs in the r-th rule, then for all $k = 1, 2, \ldots, m_j : \chi_{jk}^{(r)} = X$. This case defines the meaning of ANY: any linguistic value is admissible.

For any other combination of inputs, the output of π_i is undefined, i.e. any truth value is possible (again, this condition is usually referred as "don't care").

Each truth function π_i can be represented as a truth table, which enumerates any combination of assignments to the propositional variables of the FRBC and associates the value of π_i to each combination. Combinations associated to undefined values of π_i are not included in the table. The number of rows of each truth table matches the number of rules of the FRBC. This prevents the combinatorial explosion of rows that would be expected in the general case of truth function representation.

4.2. Minimisation of true tables

Once each truth table has been built, it can be processed so as to be minimised. The minimisation procedure produces a new truth table without modifying the truth function. The new truth table has a number of rows not greater than the original truth table. It also has a number of X values in its inputs not smaller than in the original truth table. Furthermore, minimisation guarantees that any further simplification (in terms of rows and inputs) provides for a truth function different from the original.

The Quine-McCluskey (QMC) algorithm represents an effective mechanism for minimisation of truth tables [53]. It is mainly based on the distributive property, which simplifies propositions according to the law: $ABC + A\bar{B}C \equiv AC$.

The QMC algorithm works in two stages:

(1) Merge rows with output 1 or X that differ in only one input;
(2) Find the minimum number of merged rows that cover all rows of the original truth table.

To overcome the computational cost of QMC we designed an efficient procedure that exploits the peculiar structure of truth tables derived from FRBC rules to perform minimisation quickly. A specific implementation of the first stage of QMC avoids the generation of all input combinations, thus saving time for merging. This can be achieved because the number of rows of the truth tables representing rule bases is often very small. Also, the second stage has been optimised by using heuristic procedures that drive the minimisation process without expensive search. Finally, the algorithm is forced to avoid rows that represent multiple assignments to each linguistic variable. A sketch of our procedure is presented in Appendix A.

4.3. Reconstruction of the fuzzy rule base

After minimisation, a new FRBC is built from the rows of the minimised truth table. For each class label $\lambda_i \in \Lambda$ we consider the minimised table associated to the truth function π_i. A rule is built for each row with output equal to 1. By definition of the truth table (see Section 4.1) and the minimisation algorithm (see Appendix A), for each j there is at most one k such that $\chi_{jk}^{(r)} \neq X$.

The antecedent of the rule can be defined by atoms v_j IS [NOT] $q_j^{(r)}$ where:

- $q_j^{(r)} = q_{jk}$ if $\chi_{jk}^{(r)} \neq X$;
- NOT occurs if $\chi_{jk}^{(r)} = 0$ and it does not occur if $\chi_{jk}^{(r)} = 1$;
- $q_j^{(r)} =$ ANY if $\forall k : \chi_{jk}^{(r)} = X$.

Atoms with ANY are removed to improve the readability of the rule. The consequent of the rule is λ_i.

4.4. Comparison

The two FRBCs, the first with the original rule base and the second with the minimised version, are compared in terms of classification error on the same data. In particular we register the differences in classification for each data. If the original

Table 1
The truth tables of the simple rule base.

Small	Large	Regular	Irregular	Malign	Benign
0	1	0	1	1	0
1	0	X	X	0	1

Table 2
The minimised truth tables of the simple rule base.

Small	Large	Regular	Irregular	Malign	Benign
X	1	X	X	1	0
X	0	X	X	0	1

and the minimised FRBCs are cointensive, then we expect them to produce (nearly) the same classification for every data. However, a certain amount of difference in classification is almost unavoidable in most cases. We establish the results of interpretability evaluation by analysing such a classification difference:

- If the classifications of the two FRBCs differ too much, we conclude that the original FRBC lacks of interpretability. Its accuracy is mainly due to the explicit semantics of linguistic values, which does not correspond to the propositional view of rules. The FRBC can be used for classification as a "grey box", but its labelling is arbitrary and not cointensive with user knowledge. Attaching natural language terms to such FRBC is useless and potentially misleading.
- If the two classifications are very similar, we conclude that the original FRBC is interpretable in the sense of cointension with the logical view. The explicit semantics of linguistic values is coherent with the logic operators used in minimisation. In this sense, the rule base is cointensive with user knowledge. We could retain the simplified FRBC because it is more readable than the original one, while its accuracy is almost the same.

There is no threshold to decide if a FRBC is interpretable or not, but rather a continuous spectrum of possibilities. Interpretability – as expected – is a matter of degree, and the degree of interpretability is, in our approach, inversely proportional to the difference of accuracies.

4.5. A clarifying example

For the sake of clarity the proposed strategy is applied to a simple example. Let us consider a trivial FRBC shown in (11).

IF cell_size is Large AND cell_shape is Irregular THEN malign
AND (11)
IF cell_size is Small THEN benign

Under CWA we expect that, for each input instance, only one rule is fired with maximum strength (since the two rules refer to different classes). As a consequence, the rule base in (11) can be rewritten as:

cell_size is Large AND cell_shape is Irregular AND malign
OR (12)
cell_size is Small AND benign

In Table 1 the truth tables for classes "malign" and "benign" are shown, while in Table 2 the truth tables resulting from the minimisation process are reported. The minimised truth tables lead to the following representation:

cell_size is Large AND malign
OR (13)
cell_size is Small AND benign

that is equivalent (under CWA) to:

IF cell_size is Large THEN malign
AND (14)
IF cell_size is Small THEN benign

Such a result derives from the assumption of close world and may appear somewhat controversial. Yet, we believe that the CWA is strictly related to the common inductive practice and embodies a very usual convention among humans: to assume that the information available about some aspects of reality is all the possible information about it. Actually, the stance

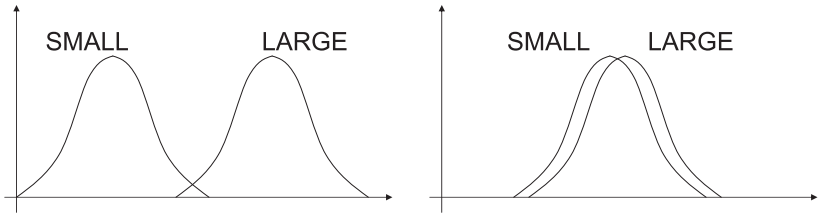


Fig. 3. The "small" and "large" concepts involved in the fuzzy rule base (11).

assuming that induction is characterised as closed-world reasoning from the available data (followed by an inductive jump) is supported also by a number of studies proposed in literature (as an example, the interested reader may refer to [54]).

The rule bases (11) and (14) can be compared in terms of classification errors. If the two classification errors are very similar, we state that the semantics of the original rule base is compatible with the applied transformation and, hence, with the logical view of rules. In this sense, we state that the knowledge base of the FRBC is cointensive with user knowledge. On the other hand, if the classification errors differ too much, then the semantics of the knowledge base cannot be represented with a set of propositions. Hence, the knowledge base is not cointensive with user mind representation: in other words, it is not interpretable.

As an exemplification, the rule base (11) may be prescribed by a human expert. In this case, the embedded information is likely to be well-founded regarding the involved input features, and the adoption of meaningful linguistic terms (large/small, malign/benign) should ensure the validity of the semantic interpretation of rules. If the concepts "large" and "small" were formalised into fuzzy sets, they would be depicted as in Fig. 3(a), being characterised — as expected — by a suitable amount of distinguishability. The minimised version (14) of the fuzzy rule base, therefore, would produce classification results in agreement with the original version (11), thus proving its interpretability in terms of logical view. On the other hand, the fuzzy rule base (11) could be obtained by means of an uncostrained automatic procedure, namely an algorithm that is able to derive a linguistic knowledge base starting from a clustering process of sample data. In this case, the soundness of the embedded information is not guaranteed and the labels "large" and "small" may be associated to a couple of concepts whose mathematical formalisation does not reflect the underlying expected semantics. Such a circumstance is depicted in Fig. 3(b), where "large" and "small" are represented by highly overlapping fuzzy sets. The performance of the original fuzzy rule base may be accurate, yet its final minimised version would assume a quasi-incoherent form due to the synonymy of the large/small concepts involved in the rule base (14). That would be the case of a poorly cointensive original knowledge base.

As concerning the adopted truth-preserving transformation, the choice of minimising the truth table as a truth-preserving transformation has a twofold advantage: by eliminating as many terms as possible we test whether the logical view of rules is preserved with the minimum required information; furthermore, if the interpretability assessment process provides positive results, we can retain the simplified rule base since we shall prefer its compactness.

Truth table minimisation in FRBCs has been previously proposed in a couple of works. In [55] a methodology is presented for the minimisation of a given set of fuzzy rules by means of mapping fuzzy relations on Boolean functions and exploiting existing Boolean synthesis algorithms. In [56] logic minimisation is used to discover the structure of neuro-fuzzy systems, which are successively tuned to improve their accuracy capabilities. Our approach is different in scope as we are dealing with interpretability assessment whilst the two works focus on simplification and refinement of fuzzy rule-based systems. Furthermore, there are technical differences as we manage positive as well as negative information to achieve higher compactness of the reduced rule base.

5. Experimental results

The effectiveness of the proposed approach for interpretability assessment is evaluated on a number of FRBCs designed from data, so as to satisfy several interpretability constraints. Of course the FRBCs taken into account are evaluated to assess their interpretability in terms of cointension.

5.1. The classification problem

The dataset used in this paper is the result of a chemical analysis of wines produced in a region of Italy, but derived from three different cultivars. The analysis determined the quantities of 13 constituents (alcohol, malic acid, ash, alcalinity of ash, magnesium, total phenols, flavanoids, nonflavanoid phenols, proanthocyanins, colour intensity, hue, OD280/OD315 of diluted wines and proline) found in each of these types of wines. The values for each attribute are integer or real numbers, and there are no missing values. The number of instances stored in this dataset is 178. The task performed is the classification of wines into the three different classes corresponding to cultivars. The dataset is freely available on the Internet [57].

In order to evaluate the effectiveness of the proposed approach, we used some FRBCs designed for the same classification problem (described above) by HILK methodology [41], which is based on a cooperation framework proposed in [58]. In

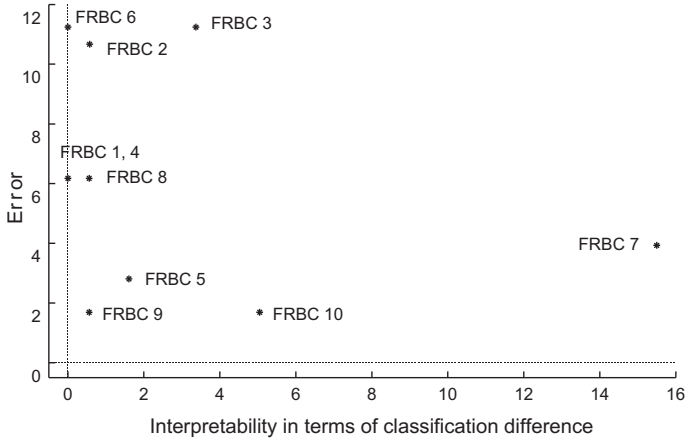


Fig. 4. Interpretability vs. accuracy of the considered FRBCs.

brief, the HILK methodology is based on a three-steps process, namely: (i) Partition design; (ii) Rule base definition and integration; (iii) Knowledge base improvement. A free software tool (distributed under the terms of the GNU General Public License) for generating FRBCs through HILK is Knowledge Base Configuration Tool (KBCT) [59,60].

We used 10 classifiers obtained by means of KBCT through different settings of the learning process. All of them are presented in [43] and they have been kindly provided to us by the authors of that work.[4] The classifiers satisfy a number of interpretability constraints, such as:

- Justifiable number of elements: the number of fuzzy sets per variable is limited to the range 7 ± 2 (following the results of Miller [51]).
- Distinguishability: fuzzy sets overlap at most at their 0.5-cut.
- Coverage: each domain value belongs to at least one fuzzy set.
- Normality and convexity: assured by using triangular fuzzy sets.

All the FRBCs have been processed so as to respond to the rule schema (6). A test has been successfully made to verify that this process does not change the accuracy of the classifiers. We will refer to the resulting FRBCs as the original ones.

5.2. Interpretability assessment of FRBCs and results discussion

Interpretability assessment followed the methodology outlined in Section 4. First of all, each rule base of an original FRBC is transformed into three true tables (related to the three classes of the wine classification task), which undergo the minimisation process. Successively, the minimised true tables are reconverted into rule base so as to obtain a new FRBC. We will refer to the minimised FRBCs as the reduced ones.

For both original and reduced FRBCs, the classification error on the entire dataset has been evaluated. Furthermore, for each FRBC the different responses between the original version and the reduced one on the same data sample have been recorded. Then, the percentage of the dataset instances differently classified has been calculated, which we call classification difference.

The experimental results are summarised in Table 3, collecting some pieces of information related to the original and reduced FRBCs. In particular, for each fuzzy rule base the table reports: the number of the involved inputs (among the 13 wine constituents), the number of rules, the number of conditions (namely, the total numbers of atoms occurring in the rules), the accuracy performance and the classification difference. It should be observed that since the classification difference encompasses *all* the differently classified instances — namely both the correctly and the wrongly classified instances —, it may not correspond with the gap between the error values of the original and reduced FRBCs (this is the case, for instance, of FRBC 3 and FRBC 7).

To highlight the point of view of the cointension-based interpretability evaluation, in Fig. 4 we drew the relationship between the accuracy (y-axis) of each original FRBC and its interpretability (x-axis). Interpretability is quantified in terms of the classification difference between the original and the reduced FRBCs. We observe that several FRBCs can be tagged as interpretable in cointension sense, since the classification difference is very small. In practice, all FRBCs, with the exception of FRBC 3, FRBC 7 and FRBC 10, can be considered interpretable: the registered classification difference is below 2%, corresponding to three – at most – differently classified examples. We should expect some discrepancy in classification because

[4] The FRBCs obtained by means of KBCT in general do not adhere to the rule schema (6) reported on page 508, therefore in some cases their rules have been subjected to a reformulation. As an example, a rule of the following form: (i) IF *A* IS MEDIUM OR LARGE THEN *B* is rewritten as the combination of the rules: (i) IF *A* IS MEDIUM THEN *B*; (ii) IF *A* IS LARGE THEN *B*.

Table 3
Experimental results.

		#Input	#Rule	#Conditions	Error (%)	Class. Diff. (%)
FRBC1	Original	6	24	50	6.18	0
	Reduced	6	24	50	6.18	
FRBC2	Original	3	5	9	10.67	0.56
	Reduced	2	4	6	11.24	
FRBC3	Original	6	8	16	11.24	3.37
	Reduced	5	8	15	11.24	
FRBC4	Original	8	32	74	6.18	0
	Reduced	8	32	74	6.18	
FRBC5	Original	4	5	10	2.84	1.7
	Reduced	3	5	9	4.54	
FRBC6	Original	3	5	10	11.3	0
	Reduced	3	5	10	11.3	
FRBC7	Original	5	11	25	4	15.43
	Reduced	3	8	15	13.71	
FRBC8	Original	7	30	111	6.18	0.56
	Reduced	7	25	86	6.74	
FRBC9	Original	9	32	94	1.69	0.56
	Reduced	9	24	72	2.26	
FRBC10	Original	7	6	15	1.7	5.11
	Reduced	3	5	9	6.82	

the inferential apparatus of a FRBC does not completely adhere to the formal logics. The use of fuzzy set theory, in fact, does not verify the law of excluded middle, and the defuzzification process has no logical counterparts. This should be considered as a point of strength of fuzzy systems because they are capable of dealing with partial inconsistency or knowledge incompleteness. However, when interpretability is a major concern, these features should have an exceptional rather usual character. In other words, the main character of an interpretable FRBC is logical, with some exceptional cases tolerated.

From the figure we observe that the FRBCs exhibiting high interpretability are quite different in terms of accuracy. This confirms our assumption of orthogonality between interpretability and accuracy. The FRBC which can be reasonably tagged as "best" both in accuracy and interpretability is FRBC 9, whose rule base is shown in Table 4. The information reported in the table concerns the explicit representation of the fuzzy rules both for the original and the reduced version of the knowledge bases. Rules in the tables are sorted by class and then by the number of conditions per rule. An additional table – Table 5 – has been included to provide a legend for the previously reported information.

We also observe that FRBC 10 has the same accuracy of FRBC 9 but its interpretability is lower because of the high classification gap. Interestingly, the rule base reported in Table 6 (corresponding to FRBC 10) is very simple, even when compared with the rule base corresponding to FRBC 9. Therefore, the structural approaches would have considered FRBC 10 more interpretable than FRBC 9. Again, this is a confirmation that complexity is not a valid indicator of semantic cointension.

The original version of FRBC 9 numbers 32 rules and 2–5 conditions per rule. The reduced version obtained after minimisation includes only 24 rules, but the accuracy performance is almost unchanged. Hence, we could retain the reduced version as the final classification model. This FRBC is not as simple as other classifiers we used in the experimentation. Nevertheless, we should adopt it as the final model because it is the sole classifier exhibiting high accuracy while being capable to communicate its embedded knowledge in a logical form.

5.3. An example of cointension violation

A deeper analysis of FRBC 7 may be insightful. From Table 3 we observe how this particular rule base, while exhibiting good classification performance, presents also a relevant gap when compared with the accuracy value of its reduced version. Therefore, the minimisation process highlights in this case a very low degree of cointension: the applied truth-preserving transformation remarkably decreases the classification ability of the FRBC.

An explanation of this phenomenon can be given by observing Table 7. It can be noted how the original rule 5 is replaced by the rule 4 in the reduced version of the knowledge base. This is in agreement with the logical observation that the condition 'fls IS vh' is present only in rule 5 (among the original rules) and therefore it is distinctive enough to characterise instances of class 2 (under the CWA). Accordingly, the minimisation process led to the definition of the rule 4 in the reduced version of the rule base, where instances are labelled as belonging to class 2 if the single condition 'fls IS vh' is verified. Therefore, it is straightforward to observe the logical equivalence preserved at the end of minimisation. However, we have also verified that

Table 4
Rule base of FRBC 9.

Class	Original		Reduced	
	Rule	Conditions	Conditions	Rule
c1	1	flv IS vl	flv IS vl	1
	2	flv IS l ∧ ci IS h	flv IS l ∧ ci IS h	2
	3	flv IS l ∧ ci IS vh	flv IS l ∧ ci IS vh	3
	4	flv IS l ∧ ci IS l ∧ od IS vl	flv IS l ∧ ci IS l ∧ od IS vl	4
	5	flv IS l ∧ ci IS m ∧ hue IS vl	flv IS l ∧ ci IS m ∧ hue IS NOT m	5
	6	flv IS l ∧ ci IS m ∧ hue IS l		
c2	7	flv IS l ∧ ci IS vl	flv IS l ∧ ci IS vl	6
	8	flv IS m ∧ prl IS vl	flv IS m ∧ prl IS vl	7
	9	alc IS l ∧ flv IS h	alc IS l ∧ flv IS h	8
	10	alc IS vl ∧ flv IS m ∧ prl IS l	alc IS vl ∧ flv IS m ∧ prl IS l	9
	11	alc IS l ∧ flv IS m ∧ prl IS l	alc IS l ∧ flv IS m ∧ prl IS l	10
	12	flv IS l ∧ ci IS m ∧ hue IS m	flv IS l ∧ ci IS m ∧ hue IS m	11
	13	mgn IS vh ∧ flv IS m ∧ prl IS m	mgn IS vh ∧ flv IS m ∧ prl IS m	12
	14	flv IS l ∧ ci IS l ∧ od IS l	flv IS l ∧ ci IS l ∧ od IS NOT vl	13
	15	flv IS l ∧ ci IS l ∧ od IS m		
	16	flv IS l ∧ ci IS l ∧ od IS h		
	17	flv IS l ∧ ci IS l ∧ od IS vh		
	18	alc IS m ∧ tp IS l ∧ flv IS m ∧ prl IS l	alc IS m ∧ tp IS l ∧ flv IS m ∧ prl IS l	14
	19	alc IS m ∧ ash IS l ∧ tp IS m ∧ flv IS m ∧ prl IS l	alc IS m ∧ ash IS l ∧ tp IS m ∧ flv IS m ∧ prl IS l	15
	20	alc IS m ∧ ash IS m ∧ tp IS m ∧ flv IS m ∧ prl IS l	alc IS m ∧ ash IS m ∧ tp IS m ∧ flv IS m ∧ prl IS l	16
c3	21	flv IS m ∧ prl IS h	flv IS m ∧ prl IS h	17
	22	flv IS m ∧ prl IS vh	flv IS m ∧ prl IS vh	18
	23	alc IS m ∧ flv IS h	alc IS NOT l ∧ flv IS h	19
	24	alc IS h ∧ flv IS h		
	25	alc IS vh ∧ flv IS h		
	26	alc IS h ∧ flv IS m ∧ prl IS l	alc IS h ∧ flv IS m ∧ prl IS l	20
	27	mgn IS l ∧ flv IS m ∧ prl IS m	mgn IS NOT vh ∧ flv IS m ∧ prl IS m	21
	28	mgn IS m ∧ flv IS m ∧ prl IS m		
	29	mgn IS h ∧ flv IS m ∧ prl IS m		
	30	alc IS m ∧ tp IS h ∧ flv IS m ∧ prl IS l	alc IS m ∧ tp IS h ∧ flv IS m ∧ prl IS l	22
	31	alc IS m ∧ ash IS h ∧ tp IS m ∧ flv IS m ∧ prl IS l	alc IS m ∧ ash IS h ∧ tp IS m ∧ flv IS m ∧ prl IS l	23
	32	alc IS m ∧ ash IS vh ∧ tp IS m ∧ flv IS m ∧ prl IS l	alc IS m ∧ ash IS vh ∧ tp IS m ∧ flv IS m ∧ prl IS l	24

Table 5
Legend.

Input	Fuzzy set	Class
alc = Alcohol	vl = Very low	c1 = Class 1
ma = Malic acid	l = Low	c2 = Class 2
ash = Ash	m = Medium	c3 = Class 3
aoa = Alcalinity of ash	h = High	
mgn = Magnesium	vh = Very high	
tp = Total phenols		
flv = Flavanoids		
np = Nonflavanoid phenols		
pra = Proanthocyanins		
ci = Color intensity		
hue = Hue		
od = OD280 OD315 of diluted wines		
prl = Proline		

Table 6
Rule base of FRBC 10.

Class	Original		Reduced	
	Rule	Conditions	Conditions	Rule
c1	1	flv IS l	flv IS l	1
	2	flv IS m ∧ ci IS h ∧ hue IS l	flv IS m ∧ ci IS h	2
c2	3	flv IS m ∧ ci IS l ∧ od IS h	flv IS m ∧ ci IS l	3
	4	alc IS l ∧ flv IS h ∧ prl IS l	flv IS h ∧ prl IS l	4
c3	5	flv IS h ∧ prl IS h	flv IS h ∧ prl IS NOT l	5
	6	mgn IS l ∧ flv IS h ∧ prl IS m		

the increased classification error characterising the reduced rule base is due to a number of instances which are misclassified because of the novel generated rule 4. As an example, instances belonging to class 3, which were correctly classified by the original rule base thanks to the rule 10, are then classified as belonging to class 2 just for the increased value of fire strength produced by the rule 4. We argue that the logical view of the original rule base is defective and, therefore, FRBC 7 cannot be intended as interpretable. The correct classification of more than 95% of instances in the dataset is mainly due to the specific

Table 7
Rule base of FRBC 7.

Class	Original		Reduced	
	Rule	Conditions	Conditions	Rule
c1	1	flv IS m ∧ ci IS h	flv IS m ∧ ci IS h	1
	2	flv IS m ∧ ci IS vh	flv IS m ∧ ci IS vh	2
	3	flv IS l ∧ ci IS m	flv IS l ∧ ci IS NOT vl	3
	4	flv IS l ∧ ci IS h		
c2	5	alc IS l ∧ flv IS vh	flv IS vh	4
	6	flv IS l ∧ ci IS vl	flv IS NOT h ∧ ci IS vl	5
	7	flv IS m ∧ ci IS vl		
	8	alc IS l ∧ flv IS h ∧ prl IS m	flv IS h ∧ prl IS m	6
	9	flv IS m ∧ ci IS l ∧ prl IS l	ci IS l ∧ prl IS l	7
c3	10	flv IS h ∧ prl IS vh	flv IS h ∧ prl IS NOT m	8
	11	mgn IS m ∧ flv IS h ∧ prl IS h		

configuration of fuzzy sets parameters, which does not emerge from the propositional view of rules. In other words, the only responsible of good accuracy performance is the FRBC explicit semantics, but it does not find correspondence in the implicit semantics formulated by the user: being purely a configuration of mathematical parameters, it cannot be communicated through the rule base. The model embedded in FRBC 7 must be considered as a "grey box" and the attachment of linguistic labels to fuzzy sets is potentially misleading.

Interestingly, from Table 7 we further observe that the original FRBC 7 is quite compact, with only 11 rules and 2–3 conditions per rule. This kind of remark stands as a confirmation of the rationale for differentiating our approach from standard mechanisms of interpretability assessment based on structural measures, such as number of rules and condition per rules. In fact, the illustrative FRBC 7 would have been tagged as highly interpretable by structural approaches because of its compactness, while we show how such a rule base fails in communicating correctly its embedded knowledge to the user. In this sense, our approach overcomes structural assessment methods in quantifying the interpretability of a FRBC when it is considered in its basic meaning, i.e. the ability of *understanding* knowledge.

6. Conclusions

The research we have conducted in this paper sets a novel direction in interpretability assessment of fuzzy knowledge bases. The point of departure for our work consisted in considering interpretability as a form of cointension of a fuzzy rule-based model with respect to user knowledge, involving the explicit semantics of rules and the implicit semantics embedded by their linguistic representations. Although ill-posed, the problem of evaluating cointension has been approximated by means of the logical view, which refers to the set of logical properties a user assumes to hold when reading a rule base. We exploited the logical view to transform the rule base into a different one that is logically equivalent to the former; then we evaluate if the retained logical equivalence corresponds also to semantic equivalence.

Experimental results performed on a number of fuzzy classifiers showed that some of them, although verifying a number of interpretability constraints (so that they could be tagged as interpretable after a structural evaluation) actually differ in terms of cointension. This makes possible to choose the most accurate fuzzy rule base, among a pool of candidates sharing the same interpretability evaluation. Eventually, the selected FRBC may exhibit some noticeable structural complexity. However, our position is that such complexity is sometimes necessary to match the complexity of the underlying problem. In those cases, alternative forms of representation can be used to organise a readable knowledge, thus managing this complexity surplus.

We recognise that the proposed approach represents the first step towards cointension based interpretability assessment. Other forms of assessment can be tailored to specific models. In this paper we have focused on FRBCs, but similar approaches could be applied to other rule-based models, like fuzzy graphs, case-based reasoning models, etc. Furthermore, the proposed approach can be adopted to evaluate the appropriateness of some design choices in FRBCs and similar models, such as the t-norm used for aggregating fuzzy sets, the t-conorm used for aggregating fuzzy rules, the defuzzification operator and so on. That will be matter of future investigation.

Appendix A. Minimisation algorithm

The algorithm used in this paper for truth table minimisation is based on the well-known Quine-McCluskey (QMC) method, but it is specifically tailored to exploit the structure of truth tables as derived from FRBCs to reduce the computational complexity.

Like QMC, the algorithm is based on two steps: reduction and minimal cover. The first step is aimed at reducing the number of rows in the truth table by merging rows that satisfy the distributive property $ABC + A\bar{B}C \equiv AC$. The reduction process is carried out on the ON-set and the DC-set. The ON-set is the set of rows corresponding to output 1 (ON) of the truth function, while the DC-set is defined by all the rows with output Don't Care (DC).

Since in our context the truth tables are generated from FRBCs, we observe that the ON-set is quite small in cardinality, as it is limited by the number of rules that assign the same class. Similarily, the OFF-set (i.e. the set of rows corresponding to output 0) is small, since the sum of cardinalities of the ON-set and the OFF-set matches the number of rules of the FRBC. We deduce that the cardinality of the DC-set is very high, being defined by all the possible bit combinations not representing rules (refer to Section 4 for the formalism):

$$2^{\sum_{j=1}^{n} m_j} - R.$$

A naïve implementation of the QMC algorithm requires an explicit representation of both the ON-set and the DC-set, thus yielding an exponential amount of space to represent the truth table. To avoid such a combinatorial explosion, we do not explicitly represent the DC-set, but just the ON-set and the OFF-set: we deduce that a row belongs to the DC-set if it does not appear neither in the ON-set nor in the OFF-set.

The reduction procedure works as follows. A row is extracted from the ON-set, and a number of generalised configurations is generated. Then, each configuration is tested so as to verify if it does not cover any row in the OFF-set. If covering is not verified, then the configuration is added to the ON-set. If none of the generated configurations is added to the ON-set, then the original row is put in the PRIME-set, which is the set of prime implicants used for the successive step of the QMC algorithm. The entire reduction process is repeated until the ON-set is empty. Formally, the reduction procedure can be described as follows:

Require: ON-set, OFF-set, $\langle m_1, m_2, \ldots, m_n \rangle$
Ensure: PRIME-set
  PRIME-set $\leftarrow \emptyset$
  while ON-set $\neq \emptyset$ **do**
    Extract a row r from the ON-set
    $G \leftarrow Generalize(r, \langle m_1, m_2, \ldots, m_n \rangle)$
    $G' \leftarrow \emptyset$
    for all $r' \in G$ **do**
      if $Not(Covers(r', \text{OFF-set}))$ **then**
        $G' \leftarrow G' \cup \{r'\}$
      end if
    end for
    if $G' = \emptyset$ **then**
      PRIME-set $\leftarrow$ PRIME-set $\cup\, r$
    else
      ON-set $\leftarrow$ ON-set $\cup\, G'$
    end if
  end while

The cover test verifies if a row covers at least one row in the OFF-set. A row covers another row if they match the same bits according to a pairwise comparison. Two bits match if either one is Don't Care or they are equal. The cover test algorithm can be hence formalised as follows:

Require: r, OFF-set $\{r = \langle r_1, r_2, \ldots, r_m \rangle, r_i \in \{0, 1, X\}\}$
Ensure: *true* iff r covers at least one row in the OFF-set
  COVER $\leftarrow$ *false*
  for all $s \in$ OFF-set **do**
    $\{s = \langle s_1, s_2, \ldots, s_m \rangle, s_i \in \{0, 1, X\}\}$
    $c \leftarrow$ *true*
    for $j = 1$ to m **do**
      if $r_j \neq X$ and $s_j \neq X$ and $r_j \neq s_j$ **then**
        $c \leftarrow$ *false*
      end if
    end for
    COVER $\leftarrow$ *COVER* $\vee\, c$
  end for
  return *COVER*

An issue arises in the generation of the generalised configurations. Here we exploit the feature of FRBC rules according to which at most one linguistic value is associated to each variable (see schema (6) reported on page 508). We need to partition the row to generalise into blocks, i.e. sequences of bits representing the assignment of the linguistic value to each variable, according to relation (10) on page 508. For each block of m_i bits, m_i generalised configurations are generated by forcing all bits except one to X. The total number of generalised configurations is hence $\sum m_i$. The procedure for generating generalised configurations is formalised in the following.

Require: $r, \langle m_1, m_2, \ldots, m_n \rangle$
Ensure: G the set of generalised configurations of rows.
  $G \leftarrow \emptyset$
  $k \leftarrow 1$
  for $i = 1$ to n **do**
    for $j = k$ to $k + m_i$ **do**
      $s \leftarrow r$
      for $h = k$ to $k + m_i$ **do**
        if $h \neq j$ **then**
          $s_h \leftarrow X$
        end if
      end for
      $G \leftarrow G \cup s$
    end for
    $k \leftarrow k + m_i$
  end for

The above procedure ensures that, for every generalised configuration and for each block, there is at most one bit different from X. This enables the reconstruction as shown in Section 4.3.

The final step of the minimisation algorithm is to select the minimal number of rows among the prime implicants that cover the whole ON-set. We use a simple heuristic to find the minimal cover. For each prime implicant in the PRIME-set, we select the implicant covering the greatest number of rows of the ON-set. Both the selected implicant and the covered rows are deleted from the corresponding sets. The process is iterated until the ON-set is empty; all the selected prime implicants form the desired minimal cover.

References

[1] L.A. Zadeh, Fuzzy logic = computing with words, IEEE Transactions on Fuzzy Systems 4 (1996) 103–111.
[2] 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) (1999) 105–119.
[3] L.A. Zadeh, Toward human level machine intelligence – Is it achievable? The need for a paradigm shift, Computational Intelligence Magazine, IEEE 3 (3) (2008) 11–22.
[4] L.A. Zadeh, Is there a need for fuzzy logic?, Information Sciences 178 (2008) 2751–2779.
[5] R.S. Michalski, A theory and methodology of inductive learning, in: R.S. Michalski, T.J. Carbonell, T.M. Mitchell (Eds.), Machine Learning: An Artificial Intelligence Approach, TIOGA Publishing Co., Palo Alto, CA, USA, 1983, pp. 83–134.
[6] C. Mencar, G. Castellano, A.M. Fanelli, On the role of interpretability in fuzzy data mining, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 15 (5) (2007) 521–537.
[7] U. Johansson, L. Niklasson, R. Köning, Accuracy vs. comprehensibility in data mining models, in: Proceedings of the Seventh International Conference on Information Fusion, Stockholm, Sweden, 2004, pp. 295–300.
[8] Z.-H. Zhou, Comprehensibility of data mining algorithms, in: J. Wang (Ed.), Encyclopedia of Data Warehousing and Mining, IGI, Hershey, PA, 2005, pp. 190–195.
[9] A. Bargiela, W. Pedrycz, Granular Computing: An Introduction, Kluwer Academic Publishers, Boston, Dordrecht, London, 2003.
[10] L.A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, Fuzzy Sets and Systems 90 (1997) 111–127.
[11] A. Riid, E. Rustern, Interpretability of fuzzy systems and its application to process control, in: Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2007), 2007, pp. 1–6.
[12] S. Altug, M.-Y. Chow, H.J. Trussel, Heuristic constraints enforcement for training of and rule extraction from a fuzzy/neural architecture – Part II: Implementation and application, IEEE Transactions on Fuzzy Systems 7 (2) (1999) 151–159.
[13] J. Espinosa, J. Vandewalle, Constructing fuzzy models with linguistic integrity from numerical data – AFRELI algorithm, IEEE Transactions on Fuzzy Systems 8 (5) (2000) 591–600.
[14] M.-Y. Chow, S. Altug, H.J. Trussel, Heuristic constraints enforcement for training of and knowledge extraction from a fuzzy/neural architecture – Part I: Foundation, IEEE Transactions on Fuzzy Systems 7 (2) (1999) 143–150.
[15] D. Nauck, R. Kruse, Obtaining interpretable fuzzy classification rules from medical data, Artificial Intelligence in Medicine 16 (1999) 149–169.
[16] J. Valente de Oliveira, Semantic constraints for membership function optimization, IEEE Transactions on Systems, Man and Cybernetics, Part A 29 (1) (1999) 128–138.
[17] G. Castellano, A.M. Fanelli, C. Mencar, Dcf: a double clustering framework for fuzzy information granulation, in: 2005 IEEE International Conference on Granular Computing, vol. 2, 2005, pp. 397–400. doi:10.1109/GRC.2005.1547320.
[18] S.-M. Zhou, J.Q. Gan, Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modelling, Fuzzy Sets and Systems 159 (23) (2008) 3091–3131.
[19] W. Pedrycz, J. de Oliveira, Optimization of fuzzy models, IEEE Transactions on Systems, Man and Cybernetics, Part B 26 (4) (1996) 627–636.
[20] C. Mencar, A.M. Fanelli, Interpretability constraints for fuzzy information granulation, Information Sciences 178 (24) (2008) 4585–4618, doi:10.1016/j.ins.2008.08.015.
[21] A. Botta, B. Lazzerini, F. Marcelloni, D. Stefanescu, Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index, Soft Computing 13 (5) (2009) 437–449.
[22] P. Fazendeiro, J.V. de Oliveira, A working hypothesis on the semantics/accuracy synergy, in: Joint EUSFLAT-LFA 2005, Barcelona, Spain, 2005, pp. 266–271.
[23] P. Fazendeiro, J.V. de Oliveira, W. Pedrycz, A multiobjective design of a patient and anaesthetist-friendly neuromuscular blockade controller, IEEE Transactions on Biomedical Engineering 54 (9) (2007) 1667–1678.
[24] M. Gacto, R. Alcala, F. Herrera, Integration of an index to preserve the semantic interpretability in the multiobjective evolutionary rule selection and tuning of linguistic fuzzy systems, IEEE Transactions on Fuzzy Systems 18 (3) (2010) 515–531.
[25] H. Roubos, M. Setnes, Compact and transparent fuzzy models and classifiers through iterative complexity reduction, IEEE Transactions on Fuzzy Systems 9 (4) (2001) 516–524.

[26] P. Meesad, G.G. Yen, Quantitative measures of the accuracy, comprehensibility, and completeness of a fuzzy expert system, in: Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ '02), Honolulu, Hawaii, 2002, pp. 284–289.
[27] Y. Jin, Fuzzy modeling of high-dimensional systems: complexity reduction and interpretability improvement, IEEE Transactions on Fuzzy Systems 8 (2) (2000) 212–221.
[28] M. Jamei, M. Mahfouf, D.A. Linkens, Elicitation and fine tuning of fuzzy control rules using symbiotic evolution, Fuzzy Sets and Systems 147 (1) (2004) 57–74.
[29] C.-A. Peña-Reyes, M. Sipper, Fuzzy CoCo: balancing accuracy and interpretability of fuzzy models by means of coevolution, in: J. Casillas, O. Cordon, F. Herrera, L. Magdalena (Eds.), Accuracy Improvements in Linguistic Fuzzy Modeling, Studies in Fuzziness and Soft Computing, vol. 129, Springer-Verlag, 2003, pp. 119–146.
[30] C.-A. Peña-Reyes, M. Sipper, Fuzzy CoCo: a cooperative-coevolutionary approach to fuzzy modeling, IEEE Transactions on Fuzzy Systems 9 (5) (2001) 727–737.
[31] F. Jiménez, A. Gómez-Skarmeta, H. Roubos, R. Babuška, A multi-objective evolutionary algorithm for fuzzy modeling, in: Proceedings of the NAFIPS'01, New York, 2001, pp. 1222–1228.
[32] R. Paiva, A. Dourado, Merging and constrained learning for interpretability in neuro-fuzzy systems, in: Proceedings of the to First International Workshop on Hybrid Methods for Adaptive Systems, Tenerife, Spain, 2001.
[33] M. Setnes, R. Babuška, U. Kaymak, H.R. Van Nauta Lemke, Similarity measures in fuzzy rule base simplification, IEEE Transactions on Systems, Man and Cybernetics, Part B 28 (3) (1998) 376–386.
[34] Y. Jin, B. Sendhoff, Extracting interpretable fuzzy rules from RBF networks, Neural Processing Letters 17 (2003) 149–164.
[35] S. Guillaume, B. Charnomordic, Generating an interpretable family of fuzzy partitions from data, IEEE Transactions on Fuzzy Systems 12 (3) (2004) 324–335.
[36] D. Nauck, R. Kruse, A neuro-fuzzy approach to obtain interpretable fuzzy systems for function approximation, in: Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'98), Anchorage (AK), 1998, pp. 1106–1111.
[37] G. Castellano, A.M. Fanelli, C. Mencar, A neuro-fuzzy network to generate human understandable knowledge from data, Cognitive Systems Research 3 (2) (2002) 125–144, Special Issue on Computational Cognitive Modeling.
[38] H. Ishibuchi, Y. Nojima, Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning, International Journal of Approximate Reasoning 44 (1) (2007) 4–31.
[39] A. Marquez, F. Marquez, A. Peregrin, A multi-objective evolutionary algorithm with an interpretability improvement mechanism for linguistic fuzzy systems with adaptive defuzzification, in: Proceedings of 2010 IEEE International Conference on Fuzzy Systems, 2010, pp. 277–283.
[40] D. Nauck, Measuring interpretability in rule-based classification systems, in: Proceedings of The 12th IEEE International Conference on Fuzzy Systems FUZZ '03, vol. 1, 2003, pp. 196–201.
[41] J.M. Alonso, L. Magdalena, S. Guillaume, HILK: a new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism, International Journal of Intelligent Systems 23 (2008) 761–794.
[42] S.-M. Zhou, J.Q. Gan, Extracting Takagi–Sugeno fuzzy rules with interpretable submodels via regularization of linguistic modifiers, IEEE Transactions on Knowledge and Data Engineering 21 (8) (2009) 1191–1204.
[43] J.M. Alonso, L. Magdalena, G. González-Rodríguez, Looking for a good fuzzy system interpretability index: an experimental approach, International Journal of Approximate Reasoning 51 (2009) 115–134.
[44] J.M. Alonso, L. Magdalena, Combining user's preferences and quality criteria into a new index for guiding the design of fuzzy systems with a good interpretability-accuracy trade-off, in: Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2010), 2010, pp. 961–968.
[45] A. Riid, E. Rustern, Interpretability improvement of fuzzy systems: reducing the number of unique singletons in zeroth order Takagi–Sugeno systems, in: Proceedings of 2010 IEEE International Conference on Fuzzy Systems, 2010, pp. 2013–2018.
[46] M. Antonelli, P. Ducange, B. Lazzerini, F. Marcelloni, Learning concurrently partition granularities and rule bases of mamdani fuzzy systems in a multi-objective evolutionary framework, International Journal of Approximate Reasoning 50 (7) (2009) 1066–1080.
[47] P. Pulkkinen, H. Koivisto, Fuzzy classifier identification using decision tree and multiobjective evolutionary algorithms, International Journal of Approximate Reasoning 48 (2) (2008) 526–543.
[48] P. Baranyi, Y. Yam, D. Tikk, R. Patton, Trade-off between approximation accuracy and complexity: TS controller design via HOSVD based complexity minimization, in: J. Casillas, O. Cordón, F. Herrera, L. Magdalena (Eds.), Interpretability Issues in Fuzzy Modeling, Springer-Verlag, Heidelberg, 2003, pp. 249–277.
[49] D. Tikk, P. Baranyi, Exact trade-off between approximation accuracy and interpretability: solving the saturation problem for certain FRBSs, in: [50], 2003, pp. 587–604.
[50] J. Casillas, O. Cordon, F. Herrera, L. Magdalena (Eds.), Interpretability Issues in Fuzzy ModelingStudies in Fuzziness and Soft Computing, vol. 128, Springer-Verlag, Germany, 2003.
[51] G.A. Miller, The magical number seven, plus or minus two: some limits on our capacity for processing information, The Psychological Review 63 (1956) 81–97.
[52] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning – Part 1, Information Sciences 8 (1975) 199–249.
[53] E.J. McCluskey, Minimization of boolean functions, Bell Labs Technical Journal 35 (5) (1956) 1417–1444.
[54] G. Núñez, U. Cortés, J. Larrosa, Non-monotonic characterization of induction and its application to inductive learning, International Journal of Intelligent Systems 10 (10) (1995) 895–927.
[55] R. Rovatti, R. Guerrieri, G. Baccarani, An enhanced two-level boolean synthesis methodology for fuzzy rules minimization, IEEE Transactions on Fuzzy Systems 3 (3) (1995) 288–299.
[56] A.F. Gobi, W. Pedrycz, Logic minimization as an efficient means of fuzzy structure discovery, IEEE Transactions on Fuzzy Systems 16 (3) (2008) 553–566.
[57] A. Frank, A. Asuncion, UCI machine learning repository, 2010. URL: <http://archive.ics.uci.edu/ml>.
[58] S. Guillaume, L. Magdalena, Expert guided integration of induced knowledge into a fuzzy knowledge base, Soft Computing 10 (9) (2006) 773–784.
[59] J.M. Alonso, L. Magdalena, S. Guillaume, KBCT: a knowledge management tool for fuzzy inference systems, free software under GPL license, 2003. URL: <http://www.mat.upm.es/projects/advocate/kbct.htm>.
[60] J.M. Alonso, L. Magdalena, S. Guillaume, Kbct: a knowledge extraction and representation tool for fuzzy logic based systems, in: Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2004), vol. 2, 2004, pp. 989–994.