Information Sciences 181 (2011) 4361–4377

Contents lists available at ScienceDirect

Information Sciences

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

Design of fuzzy rule-based classifiers with semantic cointension

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

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

ARTICLE INFO

Article history:
Available online 3 March 2011

Keywords:
Interpretability
Semantic cointension
Logical minimization
Empirical learning
Fuzzy rule-based classifier

ABSTRACT

In computing with words (CWW), knowledge is linguistically represented and has an explicit semantics defined through fuzzy information granules. The linguistic representation, in turn, naturally bears an implicit semantics that belongs to users reading the knowledge base; hence a necessary condition for achieving interpretability requires that implicit and explicit semantics are cointensive. Interpretability is definitely stringent when knowledge must be acquired from data through inductive learning. Therefore, in this paper we propose a methodology for designing interpretable fuzzy models through semantic cointension. We focus our analysis on fuzzy rule-based classifiers (FRBCs), where we observe that rules resemble logical propositions, thus semantic cointension can be partially regarded as the fulfillment of the "logical view", i.e. the set of basic logical laws that are required in any logical system. The proposed approach is grounded on the employment of a couple of tools: DC*f*, which extracts interpretable classification rules from data, and Espresso, that is capable of fast minimization of Boolean propositions. Our research demonstrates that it is possible to design models that exhibit good classification accuracy combined with high interpretability in the sense of semantic cointension. Also, structural parameters that quantify model complexity show that the derived models are also simple enough to be read and understood.

1. Introduction

In the context of granular computing (GrC), information granules are entities exhibiting functional and descriptive representations of observational data, in according to some level of abstraction [8]. GrC embraces a number of modeling frameworks to deal with representation and processing of information granules. Amongst others, fuzzy set theory is a convenient modeling framework leading to the so-called theory of fuzzy information granulation (TFIG) [48]. In TFIG, information granules are represented as fuzzy relations and form the semantic foundation for the "computing with words" (CWW), a paradigm enabling representation and inference of knowledge in a precisiated natural language [49].

The main effectiveness of CWW consists in providing suitable mechanisms of knowledge representation as well as a computational machinery for human-like reasoning. This endows computers with the abilities to deal with highly complex problems and to share knowledge in a human-centric environment [39]. The contribution of fuzziness mainly resides in granting tools for knowledge manipulation and inference. The use of TFIG approaches for granulating data, therefore, produces information granules that are defined as compositions of fuzzy sets which can be read in natural language by using linguistic variables. This is a key-point in the modelling practice: in order to be understandable, information granules resulting from computing processes should conform with concepts that humans can transpose in their own language. Anyway,

* 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).

doi:10.1016/j.ins.2011.02.014

elements of natural language can be adopted to denote information granules only if their underlying semantics highly matches the semantics related with the information granules. This means that natural language terms convey a semantics depending both on the context and on the covenants shared among all those speaking that language.

This paves the way for a different viewpoint to face the long-debated issue of interpretability assessment in fuzzy modeling. Traditionally, interpretability is translated to a number of interpretability constraints to be imposed on the definition of fuzzy information granules [33]. The definition and the choice of interpretability constraints is an open problem, because of the subjective definition of interpretability; in most cases, interpretability constraints in TFIG concern the structure of information granules (e.g. their shape, their mutual relationships, their number, etc.). Moving from the previously introduced considerations about the role of natural language and semantics in modelling practices, we are able to propose an original approach to address the problem of interpretability, which is based on the semantic cointension [32]. Roughly speaking, cointension can be viewed as a relation between concepts such that two concepts are cointensive if they refer to the same objects. Thus, in our view, a knowledge base is interpretable if its semantics is cointensive with the knowledge a user builds in his/her mind after reading the knowledge representation (expressed in natural language). More properly, we recognize the existence of two semantics coming from a knowledge base described in natural language, i.e. an *explicit* semantics – directly defined by the fuzzy information granules – and an *implicit* semantics – emerging in each user's consciousness in the attempt of understanding the knowledge base. In this setting, we deem a knowledge base as *interpretable* if the two related semantics are cointensive, that is if they refer to the same concepts. The explicit semantics is completely specified by the involved fuzzy sets and operators used for inference, while the implicit semantics is much more difficult to formalize. Nevertheless, common features can be identified for both semantics: by exploiting such common features it is possible, at least partially, to analyze cointension and hence interpretability.

We focus our study to fuzzy rule-based classifiers (FRBCs), where we observe that rules described in natural language resemble logical propositions. As a consequence, the common feature shared by implicit and explicit semantics is represented by the set of basic logical laws that are required in any logical system: we term it "logical view". As a matter of fact, the logical view is supposed to hold in both semantics, being explicitly codified in the inferential structure of knowledge base and implicitly assumed by the user while reading the linguistic rules. The validity of the logical view is then tested through the application of a minimization algorithm on a truth table derived from the knowledge base of a FRBC: by testing the validity of the logical view, we are able to assess the FRBC interpretability in the sense of cointension.

The need of comprehensibility requirements is definitely stringent whenever knowledge is to be shared among human users, and in many human-centric environments models must be acquired from data through *inductive* or *empirical learning* [13]. In such cases, the common practice for deriving interpretable knowledge is to preserve a number of interpretability constraints, which have been extensively proposed in literature especially for the automatic definition of FRBCs from data [33]. However, the introduction of the interpretability constraints inevitably restricts the degrees of freedom in the definition of information granules: that constitutes a bias limiting the model capability to adapt to data. As a consequence, accuracy and interpretability emerge as conflicting features, so that a trade-off is often required [41]. Within such a scenario, FRBCs are usually provided with a knowledge base expressed in natural language, but cointension between the explicit and implicit semantics is ultimately neglected.

The alternative approach we propose in this paper allows to acquire FRBCs from data with the specific aim of controlling interpretability in the sense of semantic cointension. Our research demonstrates that it is possible to decouple the cointensive interpretability of FRBCs from their classification capabilities: in this sense, we can get rid of the interpretability/accuracy trade-off. More properly, the trade-off is translated to a different level where the *complexity* of a FRBC is involved. In other words, the common heuristics pertaining to the structural analyses of a model are related to its intrinsic complexity, while in our stance interpretability has to be considered only from a semantical point view. Since complexity still remains a major feature to ensure readability of a FRBC, we are interested in a structural evaluation as well: we experimented our strategy on a number of real world benchmarks and a comparison with another model developed in literature shows how FRBCs which are interpretable from a cointeinsive point of view are comparable with other models also as respects their structural complexity.

The proposed approach is grounded on the employment of a couple of tools: DC*f* [16] and Espresso [11]. The DC*f* framework has been developed during a previous research activity of ours for extracting interpretable classification rules from data. Espresso is a tool capable to derive fast minimization of Boolean propositions: it has been used as a part of the four-stage strategy of interpretability assessment based on the logical view approach. The combined employment of DC*f* and Espresso enables the realization of a framework designed to build up FRBCs starting from data and to evaluate them in terms of semantic cointension. It is important to observe that the FRBCs produced by means of the DC*f* framework adhere to a number of interpretability constraints (related to the shape and displacement of the involved fuzzy sets) and can be bounded in terms of their complexity (related to the number of the involved fuzzy rules), thus ensuring an a priori control over their structural side. The subsequent evaluation is centred only on the logical view approach.

The following section deals with the problem of interpretability assessment, with special emphasis on the point of view concerning semantic cointension. Section 3 is devoted to the mathematical formalization of FRBCs and the description of the proposed approach for data-driven design. Section 4 reports setup and outcomes of the experimental sessions, with a consequent discussion of the obtained results. Conclusive remarks are drawn in Section 5.

2. Interpretability assessment

The proposed approach for interpretability assessment is oriented toward the evaluation of semantic cointension of a knowledge base, which is almost complementary to the structural approaches that are usually found in literature. Therefore, we briefly present a state of art of structural methods before illustrating the details of the semantic cointension approach.

2.1. Structural approaches

Interpretability assessment represents a major issue in the field of fuzzy knowledge-based system modeling. However, a proper evaluation of interpretability appears to be a controversial problem, since the definition of interpretability eludes any formal characterization. To face this problem the common practice is to 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 (see [33] for a survey), and many approaches have been proposed for interpretability-driven design of fuzzy models based on such constraints [16,17,6,18,36,47].

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 [52]. Some approaches for interpretability evaluation at the level of fuzzy sets and partitions can be found in literature. In [10] 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 [19] 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 [20] an index is defined to preserve the semantic integrity of fuzzy sets when a rule-based fuzzy model undergoes an optimization process through a multi-objective genetic algorithm. Also, the complexity of the rule base is minimized so as to improve its readability.

In many cases, evaluation of interpretability at fuzzy set level is reduced to assessing similarity of fuzzy sets so as to maximize their distinguishability. Several approaches have been proposed to derive fuzzy models that minimize similarity, often through genetic approaches such as genetic algorithms [42,31,45], evolution strategies [27], symbiotic evolution [25], co-evolution [40], multi-objective genetic optimization [26]. Alternatively, distinguishability improvement is realized in separate learning stages, or within penalized learning schemes [18,37,44,22,35,15].

In our research activity we are interested in the analysis of a higher level of modeling, namely the level of knowledge base. In this case, the common practice for interpretability assessment mainly relies on structural considerations about the knowledge base arrangement (involving, for instance, the number of rules or the number of fuzzy sets per rule). The basic assumption is that the lower the number of some of these structural features is kept, the more interpretable the knowledge base can be deemed. As an example, in [24] interpretability is evaluated in terms of number of rules, total rule length and average rule length, while in [30] 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.

In [34] interpretability of a FRBC is evaluated by considering a number of parameters, including the number of labels, the coverage degree of the fuzzy partition and the number of classes divided by the total number of premises. In [3] such an approach 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 [51] 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; penalized learning is used to balance interpretability and accuracy.

In [4] an empirical study has been carried out with the aim of detecting the most adequate interpretability index on the basis of the results of a web poll. In the poll users were requested to judge the interpretability of a number of knowledge bases defined for the same classification problem. The study revealed the extreme subjectivity in evaluating interpretability; as a result, none of the considered indices were conformant to human judgment. In a successive study, the authors propose a flexible index that takes into account user preferences in evaluating interpretability of a fuzzy system [2].

In summary, all the aforementioned structural approaches mainly focus on the *complexity analysis* of a fuzzy model. Certainly, complexity is a relevant factor to be considered whenever we are interested in assessing the *readability* of a fuzzy model. However, interpretability assessment cannot be reduced to the evaluation of readability, but trespasses to a semantic analysis of a knowledge base that escapes from structural assessments. This is also emphasized in [4], where two dimensions have been defined to characterize interpretability measures: description and explanation. Whilst description concerns the structural evaluation of a knowledge base (including, e.g. the shape of fuzzy sets, the number of linguistic variables, number of rules, etc.), explanation regards the semantics conveyed by a fuzzy knowledge base. The work also distinguishes between interpretability of partitions and rules since they call for different evaluation criteria. In this paper we focus on the explanation capabilities of rules, which we cast on the more general quest for semantic cointension.

2.2. Interpretability as semantic cointension

The logical view approach has been recently introduced in a previous paper of ours [32]. The rationale behind this approach for interpretability evaluation comes from the observation that the rule base is the linguistic interface of the fuzzy model to the user. More properly, the adoption of linguistic terms has the effect of concealing from the user the underlying mathematical representation based on information granules, that is the actual model's working engine. This mathematical apparatus is associated with a form of explicit semantics, which is quantitatively appreciable due to its formal counterpart (including the fuzzy sets with their parameters, the t-norm used for expressing conjunction, and so on). On the other hand, the linguistic terms convey also another kind of semantics, evoked by their specific meaning as tools of communication. This implicit semantics is associated with the cognitive process put in action by the user while reading the knowledge base and, as such, is much more elusive of any formal description. Therefore, it appears that implicit semantics can be only qualitatively evaluated.

Our approach for interpretability assessment, rather than relying on structural analysis, is oriented to focus the meaning of interpretability on the effectiveness of the semantics related to a knowledge base. In order to draw a connection between the explicit and the implicit semantics, we refer to the notion of cointension proposed by Zadeh [50]:

> 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.

Since we are going to analyze interpretability in terms of the user's capability of understanding the knowledge base starting from its linguistic representation, we formulate 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.

Due to the different nature of the involved semantics, the task of evaluating the cointension between them is not trivial. In fact, while explicit semantics is supported by a mathematical set-up providing for a formal counterpart, implicit semantics is not liable to formal investigation being, ultimately, a cast of human mind. Yet, there exists something which, on the one hand, is embedded both in human mind and in knowledge bases and, on the other hand, is also subjected to a formal representation. Such an overlap – which represents for sure only a partial intersection between explicit and implicit semantics – is constituted by the *logical view*, meant as the ensemble of laws in formal logics that can be recognized both in the propositional structure of the knowledge base (made up of rules) and in the rational thinking of a human user. As a consequence, the logical view can be exploited as the common ground to analyze both explicit and implicit semantics conveyed to the reader. By doing so, we are able to translate the interpretability evaluation problem into a formal process: being like propositions, rules can be modified by truth-preserving operators and the consequent distortion of their fuzzy semantics can be prefigured to a reduced amount, due to the shared propositional view between the fuzzy model and the user.

In summary, the logical view approach for interpretability assessment is based on a four-stage strategy. In the first step a fuzzy rule base is transformed into several truth tables – one truth table for each class – without any semantics change. This is possible because of the propositional view of the fuzzy rules. In the second step these truth tables are, in turn, minimized by applying truth-preserving operators. The new set of truth tables, obtained by the minimization process, is transformed into propositions in the third step, so to constitute a new rule base, different from the original one. Finally, the two rule bases are compared on the basis of their classification performance. If they do not differ too much, we recognize that the logical view of the original model (which is shared with the human user) is in agreement with the explicit semantics exhibited by the fuzzy rules. In other words, cointension with the user's knowledge is verified and the model can be deemed interpretable.

On the other hand, if the two knowledge bases are characterized by notably different accuracy values, then the logical view of the model is not compatible with the explicit semantics of fuzzy rules, 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 labels would be misleading and the classification capability of the original model only relies on the mathematical configuration of its parameters (without any engagement of comprehensible information).

It is worth noticing that truth table minimization in FRBCs has been previously proposed in two works. In [43] a methodology is presented for the minimization of a given set of fuzzy rules by means of mapping fuzzy relations on Boolean functions and exploiting existing Boolean synthesis algorithms. In [21] logic minimization 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.

The choice of the truth-preserving operators to be applied on the logical propositions deserves a special mention. Actually, several modifications can be conceived to convert the original fuzzy rule base. Among them, the one minimizing the number of the involved linguistic terms appears to be mostly suitable. In fact, by eliminating as many terms as possible, it is possible to verify the preservation of the logical view in the specific condition where only the minimum required information is available. Furthermore, if assessment leads to positive results, the simplified rule base can be retained in place of the original one (it should be preferred by reason of its increased compactness).

As a final remark, we address the common assertion of regarding interpretability and accuracy as conflicting properties [40,24,7,46,13]: roughly speaking, accurate models are usually not interpretable and vice versa. This is the case when structural analysis is involved to assess interpretability so that simplicity (intended as antinomy of complexity) is required to make a knowledge base comprehensible.

It is quite obvious that simple models are more readable than complex ones, but this should not represent a bias to define interpretable models, even when the relationship to be modeled is very complex. If the complexity issue (which can be addressed by a number of methods purposely designed for organizing knowledge) is kept aside from interpretability assessment (in the sense of cointension verification), interpretability and accuracy can be thought as orthogonal characterizations 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.

3. Design methodology

3.1. Fuzzy rule-based classifiers

We consider a classifier as a system computing a function:

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

where $\mathbf{X} = X_1 \times X_2 \times \cdots \times X_n \subseteq \mathbf{R}^n$ is a n-dimensional input space, and $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_c\}$ is a set of class labels. If a dataset:

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

of N pre-classified instances is given, then the classification error:

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

can be computed, 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:

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

being:

- v_j the name of the variable.
- X_j the domain of the variable.
- $Q_j = \{q_{j1}, q_{j2}, \ldots, q_{jm_j}, \text{Any}\}$ is a set of labels denoting linguistic values for the variable (e.g. Small, Medium, Large).
- $S_j = \{s_{j1}, s_{j2}, \ldots, s_{jm_j+1}\}$ 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 $I_j(q_{jk}) = s_{jk}$.

We also 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:

$$\text{IF } v_1 \text{ is [Not] } q_1^{(r)} \text{ AND } \cdots \text{ AND } v_n \text{ is [Not] } q_n^{(r)} \text{ THEN } \lambda^{(r)}, \tag{5}$$

being $q_j^{(r)} \in Q_j$ and $\lambda^{(r)} \in \Lambda$. The 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.[1]

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:

$$\mu_r(\mathbf{x}) = s_1^{(r)}(x_1) \wedge s_2^{(r)}(x_2) \wedge \cdots \wedge s_n^{(r)}(x_n), \tag{6}$$

being $s_j^{(r)} = v_j^{(r)}(I_j(q_j^{(r)}))$, $j = 1, 2, \ldots, n$, $r = 1, 2, \ldots, R$. Function $v_j^{(r)}(t)$ is $1-t$ if "Not" occurs before $q_j^{(r)}$, otherwise it is defined as t. As concerning the operator $\wedge : [0,1]^2 \to [0,1]$, we choose the min function amongst all possible t-norms, since it is required for semantic cointension [12].

The degree of membership of input $\mathbf{x}$ to class λ_i is computed by considering all the rules of the FRBC with λ_i in the consequent part, i.e.:

$$\mu_{\lambda_i}(\mathbf{x}) = \max \mu_r(\mathbf{x}) \chi\left(\lambda_i, \lambda^{(r)}\right). \tag{7}$$

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

$$f_{\text{FRBC}}(\mathbf{x}) = \lambda \Rightarrow \mu_\lambda(\mathbf{x}) = \max_{i=1,2,\ldots,c} \mu_{\lambda_i}(\mathbf{x}). \tag{8}$$

[1] The sequence Not Any is not allowed.

3.2. FRBC learning

A FRBC is acquired from data through a learning process that is based on the double clustering framework (DC*f*) [14]. DC*f* is a general framework that enables interpretable fuzzy information granulation by representing each information granule as the Cartesian product of one-dimensional fuzzy sets that satisfy the following interpretability constraints:

Normality: for each fuzzy set there exists at least one element (prototype) with full membership. Normal fuzzy sets define consistent concepts [38].
Convexity: the degree of membership of an element to a fuzzy set decreases as the distance of the element from the prototype increases. Convexity enables the definition of fuzzy sets representing elementary concepts.
Distinguishability: two fuzzy sets defined on the same domain do not overlap too much. In DC*f*, distinguishability is quantified by the possibility measure, which is constrained to be less than a threshold (typically, 0.5). Distinguishable fuzzy sets represent distinct concepts, that can be safely denoted with different linguistic terms.
Coverage: each element of a domain belongs to at least one fuzzy set with membership degree greater than a threshold (typically, 0.5). Coverage avoids under-represented regions of the domain on the conceptual level.
Bounds: the minimum and maximum values of each domain are prototypes of some fuzzy sets. This avoids inconsistent representations of bound values.

As concerning its working engine, DC*f* is based on two sequential steps:

1. First, data are clustered or quantized in order to derive a set of multidimensional prototypes that represent hidden relationships among data. Such prototypes provide for a compressed view of the dataset.
2. Then, all prototypes are projected onto each dimension and further clustered according to the desired granularity level. The derived information is used to define fuzzy sets that satisfy the aforementioned interpretability constraints. Such fuzzy sets are combined through the Cartesian product to define fuzzy information granules.

DC*f* can be customized according to specific needs. In this work, we use fuzzy C-means for the first step and an incremental combination procedure to generate granules. After data granulation, a FRBC is derived by exploiting the derived granules and the class information available from data. In the following both the involved procedures are briefly described.

3.2.1. Fuzzy c-means

Fuzzy C-means (FCM) is a widely known prototype-based clustering algorithm [9]. Due to its popularity, here we provide only a short presentation. Given a dataset $D_{\mathbf{X}} = \{\mathbf{x}_i | \mathbf{x}_i \in \mathbf{X}, i = 1, 2, \ldots, N\}$, FCM operates through alternate optimization so as to minimize the following objective function:

$$J(\mathbf{V}, \mathbf{U}) = \sum_{i=1}^{N} \sum_{j=1}^{c} u_{ij}^{m} \|\mathbf{x}_i - \mathbf{v}_j\|^2, \qquad (9)$$

where $V = [\mathbf{v}_j]$, $j = 1, 2, \ldots, c$ is the prototype matrix, $\mathbf{v}_j \in \mathbf{X}$ and $U = [u_{ij}]$, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, c$ is the partition matrix that must verify the following conditions:

$$\forall j : 0 < \sum_{i=1}^{N} u_{ij} < N,$$
$$\forall i : \sum_{j=1}^{c} u_{ij} = 1.$$

FCM is widely used because of several points of strength, including $\mathcal{O}(N)$ computational complexity and robustness against prototype initialization. On the other hand, FCM requires the a priori specification of the number of clusters (the c parameter) and the fuzziness coefficient (the m parameter, which we fix to 2 since we are more interested on the prototypes than on the partition matrix). We preferred FCM over vector quantization techniques (such as LVQ1) to avoid the risk of dead units that may appear in competitive learning schemes [29].

After clustering, each derived prototype $\mathbf{v}_j$ is associated to a class $\lambda_j \in \Lambda$ according to its neighborhood (see Fig. 1). More specifically, the dataset $D_{\mathbf{X}}$ is partitioned into c parts, so that an example $\mathbf{x}_i$ belongs to the jth partition iff the distance $\|\mathbf{x}_i - \mathbf{v}_j\|$ is minimal amongst all prototypes. Then, the class labels of each example are used, and the most frequent class label within a partition is attached to the corresponding prototype.

3.2.2. Data granulation

In the second step, each prototype $\mathbf{v}_j = \langle v_{j1}, v_{j2}, \ldots, v_{jn} \rangle$ is projected onto each dimension $k = 1, 2, \ldots, n$ (see Fig. 2). The projections $v_{1k}, v_{2k}, \ldots, v_{ck}$ inherit the prototype class labels $\lambda_{jk} = \lambda_j$ and are used to define the fuzzy sets on X_k.

In this work, we use Gaussian fuzzy sets, which are completely characterized by their center and width. Gaussian fuzzy sets that meet the aforementioned interpretability constraints can be defined through *cuts*. Informally speaking, a cut is the

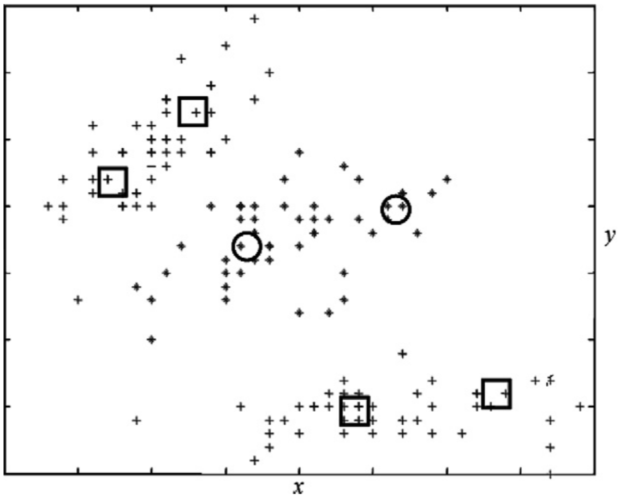

Fig. 1. Data compression: data points belonging to two classes (crosses and stars) are compressed by FCM prototypes (squares and circles).

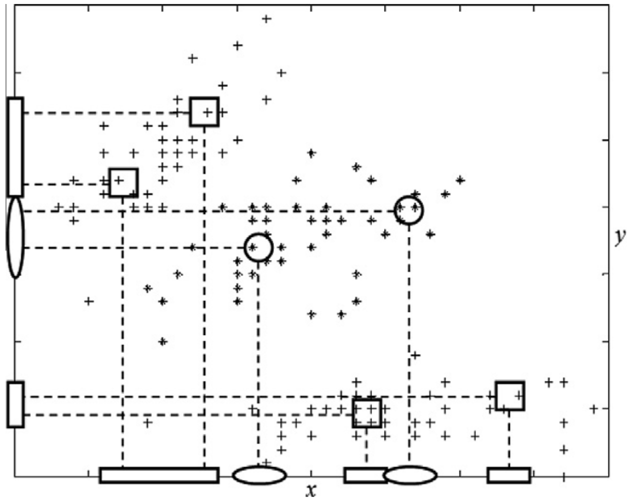

Fig. 2. Data granulation/1: prototypes derived from FCM are projected onto each dimension and then clustered.

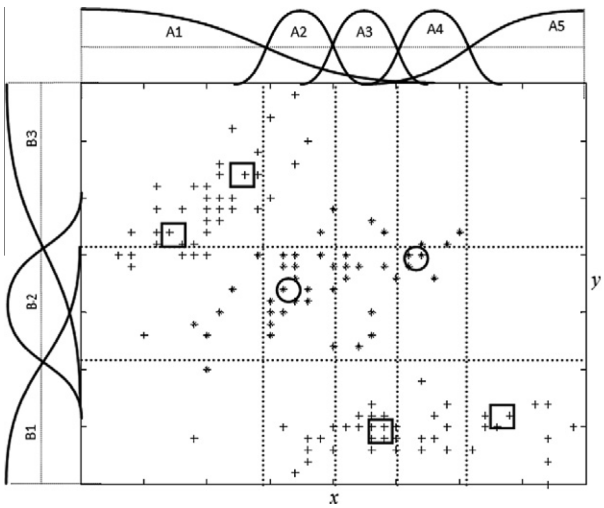

Fig. 3. Data granulation/2: cuts are used to define Gaussian fuzzy sets with interpretability constraints.

midpoint between two projections belonging to different classes. Algorithm 1 formally describes the procedure to compute the cut values. As it can be observed, the computational complexity of the algorithm is $\mathcal{O}(c \log c)$ (due to the sorting operation), thus resulting in a feasible procedure since usually $c \ll N$.

Algorithm 1: Calculate cuts

Require: prototype projections $v_{1k}, v_{2k}, \ldots, v_{ck}$
Require: bound values $m_k = \min X_k$ and $M_k = \max X_k$
Require: class labels $\lambda_{1k}, \lambda_{2k}, \ldots, \lambda_{ck}$

Ensure: sequence of cuts $t_{0k}, t_{1k}, \ldots, t_{mk}, t_{m+1,k}$

SORT the prototype projections and re-order the class labels accordingly
$\lambda_{current} = \lambda_{1k}$
$m \leftarrow 0$
for $j \leftarrow 2$ to c **do**
 if $\lambda_{jk} \neq \lambda_{current}$ **then**
  $m \leftarrow m + 1$
  $t_{mk} \leftarrow \frac{v_{j-1,k} + v_{jk}}{2}$
  $\lambda_{current} = \lambda_{jk}$
 end if
end for
$t_{0k} \leftarrow 2m_k - t_{1k}$
$t_{m+1,k} \leftarrow 2M_k - t_{mk}$ {t_{0k} and $t_{m+1,k}$ are required for the bound constraint}
return the sequence $t_{0k}, t_{1k}, \ldots, t_{mk}, t_{m+1,k}$

Once cuts are available, Gaussian fuzzy sets are defined accordingly onto each dimension (see Fig. 3). More specifically, $m + 1$ fuzzy sets are defined with the following membership function:

$$\mu_{hk}(x) = \exp\left(-\frac{(x - \omega_{hk})^2}{2\sigma_{hk}^2}\right), \tag{10}$$

where

$$\omega_{hk} = \frac{t_{h-1,k} + t_{hk}}{2} \tag{11}$$

and

$$\sigma_{hk} = \frac{t_{hk} - t_{h-1,k}}{\sqrt{-2 \log \epsilon}}, \tag{12}$$

being ϵ the desired distinguishability/coverage degree and h varying from *1* to $m + 1$.

Information granules are defined as the Cartesian products of one-dimensional fuzzy sets (one for each dimension). To avoid combinatorial explosion, only a selection of information granules is selected. Specifically, for each multidimensional prototype $\mathbf{v}_j$ and for each dimension k, the fuzzy set with highest membership degree on the projection v_{jk} is selected, namely:

$$h_{jk}^* = \arg \max_{h=1,2,\ldots,m+1} \mu_{hk}(v_{jk}). \tag{13}$$

The selected information granule for the prototype $\mathbf{v}_j$ is

$$\mathbf{G}_j = \mu_{h_{j1}^*,1} \times \mu_{h_{j2}^*,2} \times \cdots \times \mu_{h_{jn}^*,n}, \tag{14}$$

with class label $\lambda(\mathbf{G}_j) = \lambda_j$. Thus, at most c information granules are selected with time complexity $\mathcal{O}(c)$.

In a similar way, a ranked list of some other information granules is generated as follows. For each data point $\mathbf{x}_i$, an information granule is selected as in (13 and 14) – provided that it is different from those previously generated from the prototypes. Thus, at most N granules are selected with time complexity $\mathcal{O}(N)$. This granule list is ranked on the basis of the frequency of selection: the more frequently a granule is selected, the higher its position in the list. In other words, the ranking is performed on the basis of the numerousity of the data points inside each granule; accordingly, each granule is assigned a class label specifying its position inside the ranking. The ranked list will be used in the following step to generate the FRBC according to an incremental procedure (see Fig. 4).

3.2.3. FRBC generation

The information granules selected from prototypes are used to generate the first version of a FRBC that consists of as many rules as the number of granules. The generation of a rule is straightforward and simply consists in a representation of each information granule according to the rule schema (5) (see Table 1).

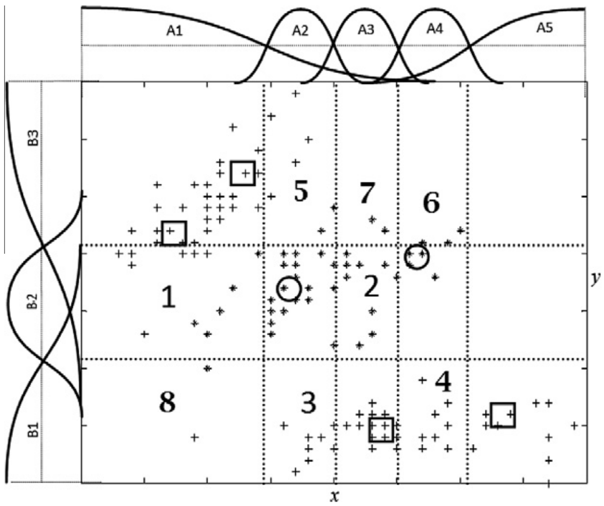

Fig. 4. Data granulation/3: the ranked list of selected information granules.

Table 1
FRBC generation: linguistic rule base representation.

Rule	Premises	Class
R1	IF x IS A_1 AND y IS B_3	Square (S)
R2	IF x IS A_2 AND y IS B_2	Circle (C)
R3	IF x IS A_3 AND y IS B_1	Square (S)
R4	IF x IS A_4 AND y IS B_2	Circle (C)
R5	IF x IS A_5 AND y IS B_1	Square (S)

Table 2
Binary rule base representation (the table should be read as the merging of the truth functions defined for the class labels).

Rule	Premises								Class	
	x					y				
	A_1	A_2	A_3	A_4	A_5	B_1	B_2	B_3	S	C
R1	1	0	0	0	0	0	0	1	1	0
R2	0	1	0	0	0	0	1	0	0	1
R3	0	0	1	0	0	1	0	0	1	0
R4	0	0	0	1	0	0	1	0	0	1
R5	0	0	0	0	1	1	0	0	1	0

However, this is only the first version of the FRBC that, expectedly, needs further refinement by adding new rules derived by the previously defined ranked list of information granules. The refinement procedure is iterative and based on the evaluation of the FRBC in terms of classification accuracy and interpretability (see forth). At each step of the iteration, a rule is added by removing the topmost information granule from the list and representing it in a linguistic form.

The iteration stops when either the maximum allowed number of rules is reached or no further accuracy improvement is observed.

3.3. Interpretability assessment

The assessment of semantic cointension is based on a four-stage strategy: (i) truth table representation, (ii) minimization, (iii) reconstruction and (iv) comparison.

3.3.1. Truth table representation

Each rule of the FRBC is seen as a proposition, i.e. a combination of propositional variables that is considered true for a class. A truth function is then defined for each class label. Each column of the truth table corresponds to a linguistic value used in the FRBC (except the Anys), while each row in the truth table corresponds to a rule of the FRBC. Given a rule and a

Table 3
Reduced DCClass binary rule base representation.

Rule	Premises								Class	
	x					y				
	A_1	A_2	A_3	A_4	A_5	B_1	B_2	B_3	S	C
R1	–	–	–	–	–	–	0	–	1	0
R2	–	–	–	–	–	–	1	–	0	1

linguistic value of a linguistic variable, if the linguistic value appears in the positive form (i.e. without NOT), then the corresponding value in the truth table is set to 1 and all the values corresponding to the remaining linguistic values of the same variable are set to 0. On the other hand, if the linguistic value appears in the negative form (i.e. with NOT), then the value in the truth table is set to 0 and the values corresponding to the remaining linguistic values of the same variable are set to "Don't Care" (or "–"). Finally, if the linguistic value is Any, then all the values of the truth table corresponding to the linguistic values of the variable are set to "Don't Care". The output of the row is equal to "1" if the class label of the truth table matches the class of the rule, "0" otherwise. For instance, Table 2 shows the true table of the rule base expressed in Table 1.

3.3.2. Minimization

Each truth table is minimized so as to reduce the number of rows and to maximize the number of "Don't Care", yet preserving the truth equivalence of the original table. We adopt Espresso as a tool for efficient minimization of Boolean formulas [11]. Espresso uses a number of heuristics to carry out the minimization process (as known, the minimization problem is not solvable in polynomial time [23]). For the sake of clearness a reduction of Table 2 is shown in Table 3.

3.3.3. Reconstruction

After minimization, a new FRBC is built from the rows of the minimized truth table, according to a procedure which performs in a reversed order the steps followed to build up the truth table. Table 4 presents the reconstruction obtained from the values reported in Table 3.

3.3.4. Comparison

The two FRBCs – the original rule base and its minimized version – are compared in terms of a differences of classification index (DC) evaluated on the same data, which can be formulated as follows:

$$\mathrm{DC} = \sum_{i=1}^{N} (1 - \chi(f(\mathbf{x}_i), f'(\mathbf{x}_i))), \tag{15}$$

being $f(\mathbf{x}_i)$ and $f'(\mathbf{x}_i)$ the classification provided for the input instance $\mathbf{x}_i$ by the original knowledge base and the minimized one, respectively. The rationale behind this evaluation is that whenever the two classification responses differ too much, we can conclude that the original FRBC lacks of interpretability. Actually, there is no threshold to be adopted in the error comparison (hence, 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. This leads to the definition of the following index – called Semantic Cointension (SC) – that can be used to compare multiple models:

$$\mathrm{SC} = 1 - \frac{\mathrm{DC}}{N}. \tag{16}$$

On the basis of the SC index it is possible to draw some conclusions about the interpretability of the predictive model under analysis. Low values of SC for an accurate FRBC indicate that its accuracy is mainly due to the semantic definition of linguistic values, which does not correspond to the propositional view of rules. In such cases, the original FRBC can be used for classification as a "gray box", but its linguistic labeling is arbitrary and not cointensive with user knowledge. Actually, attaching natural language terms to this kind of FRBC is useless and potentially misleading. On the other hand, high values of SC indicate that two classification responses are very similar, therefore we conclude that the original FRBC is interpretable: the semantic definition of linguistic values is coherent with the logic operators used in minimization. In this sense, the semantics of linguistic values is cointensive with user knowledge.

Table 4
Reduced DCClass linguistic rule base representation.

Rule	Premises	Class
R1	IF y IS NOT B_2	Square (S)
R2	IF y IS B_2	Circle (C)

It is straightforward to observe that at the end of the process we could adopt the simplified FRBC because its readability is greater than the original (due to its higher structural compactness), while its accuracy is almost the same of the original. In other words, a complexity analysis would reward for sure the minimized version of the FRBC, which can be retained and employed for future references.

4. Experimental analysis

The objective of the experimental analysis is to evaluate the ability of the proposed approach in acquiring fuzzy classification rules that are both accurate and interpretable according to semantic cointension. To this pursuit, we applied the proposed methodology to design FRBCs from a number of benchmark datasets; then we analyzed the resulting models in terms of accuracy, semantic cointension and model complexity.

4.1. Experimental setup

The datasets used for experimentation have been acquired from the UCI repository.[2] They all relate to classification problems with two or more classes:

BUPA: Bupa liver disorders recognition. The problem consists of making a diagnosis on a male patient to predict his liver disorder, based on blood tests and alcohol consumption. The dataset is made up of 345 samples, six attributes and two classes. No missing values are present in the dataset.

GLASS: Glass identification. The glass classification problem concerns the identification of six types of glasses given 214 samples. Nine attributes are used for describing each sample. No missing values are present in the dataset.

WBC: Wisconsin breast cancer. This dataset was obtained from the University of Wisconsin Hospitals, Madison, from Dr. William H. Wolberg. The adopted version consists of 683 samples characterized by nine features and two classes (benign or malignant). No missing values are present in the dataset.

WINE: Wine identification. This dataset collects the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. A number of 178 instances, 13 features and three classes are involved. No missing values are present in the dataset.

IRIS: Iris plants identification. The data set contains three classes of 50 instances each (without missing values), where each class refers to a type of iris plant. One class is linearly separable from the other two. Samples are described by four features.

IONOSPHERE: Identification of free electrons in the ionosphere. Data are collected by using a radar system in Goose Bay, Labrador. The dataset contains 351 instances, 34 features and two classes (good or bad). No missing values are present in the dataset.

To avoid feature dominance in FCM, all datasets have been linearly normalized so that each feature falls in the interval $[0,1]$.

Only one hyper-parameter is involved in the proposed methodology, i.e. the number c of clusters in FCM. This value has been varied in each experimental session ranging from a minimum value coinciding with the number of classes of the dataset at hand, to a maximum value which is three times the number of classes (with the exception of GLASS data).

Finally, in order to get robust results, each experimental session has been run according to stratified 10-fold cross validation sessions, which required to split each dataset into ten parts, in such a way that the proportion of classes in the entire dataset is respected into each part.

4.2. Experimental results

According to the proposed methodology, each session run provides for an incremental sequence of FRBCs $F_1, F_2, \ldots, F_s$. Each FRBC F_i provides for quantitative features such as classification accuracy A_i (i.e. the complement of the classification error on the test set defined in (3)), semantic cointension index SC_i (defined in (16)), number of rules R_i and total number P_i of premises, i.e. atoms that do not have ANY as linguistic values.

The proposed methodology does not end up with a single model, but rather a sequence of models. The objective of the experimentation is to evaluate these models in terms of accuracy and interpretability. To simplify the evaluation, we pick up to three models from each sequence, namely the most accurate (MA), the most interpretable (MI) and the best trade-off (BT). Since we are evaluating the results at the end of the design stage, we measure the performances on the test set. The evaluated models are chosen according to the following criteria:

- The most accurate (MA) FRBC, i.e. the classifier with minimum classification error on the test set:

$$\mathrm{MA} = \arg\max_i A_i.$$

[2] <http://archive.ics.uci.edu/ml/>.

Table 5
Experimental results on BUPA dataset related to the selected most accurate (MA), best trade-off (BT) and most interpretable (MI) models, in terms of the averaged features: A = accuracy; SC = semantic cointension; R = number of rules; P = number of premises. *c* is the number of clusters in FCM (i.e. an upper bound on the number of fuzzy sets per input).

c	MA model				BT model				MI model			
	A (%)	SC (%)	R	P	A (%)	SC (%)	R	P	A (%)	SC (%)	R	P
2	65.3	87.5	9.5	27.8	62.4	98.8	13.0	41.8	61.2	99.1	12.7	40.1
3	65.1	76.7	8.2	22.5	62.5	87.9	8.3	23.8	58.4	89.6	8.8	24.8
4	66.4	83.9	6.2	15.3	63.6	93.7	5.5	13.6	61.3	95.1	5.6	13.6
5	69.3	83.9	9.8	27.5	66.7	91.1	9.1	24.8	59.5	94.2	6.4	14.6
6	71.4	75.5	11.4	34.2	65.9	85.4	6.6	16.9	57.3	91.7	2.6	3.3

Table 6
Experimental results on GLASS dataset (legend in Table 5).

c	MA model				BT model				MI model			
	A (%)	SC (%)	R	P	A (%)	SC (%)	R	P	A (%)	SC (%)	R	P
6	63.0	54.6	27.0	102.9	58.4	67.4	22.6	78.4	41.6	76.1	17.3	52.7
7	64.3	55.1	37.1	162.8	58.9	66.9	45.0	197.4	48.7	70.7	20.5	75.9
8	65.4	54.4	65.8	309.1	62.4	60.8	71.4	344.4	40.5	66.3	23.5	95.9
9	67.7	47.5	129.0	635.9	65.1	59.7	168.3	812.3	47.3	64.3	69.5	256.9

Table 7
Experimental results on WBC dataset (legend in Table. 5).

c	MA model				BT model				MI model			
	A (%)	SC (%)	R	P	A (%)	SC (%)	R	P	A (%)	SC (%)	R	P
2	96.2	96.9	4.5	9.1	95.9	98.2	4.6	9.5	95.1	98.8	3.7	9.7
3	96.3	97.2	4.2	8.5	96.0	98.2	4.6	9.3	95.7	98.4	4.5	8.9
4	96.8	95.7	3.9	7.6	96.2	97.7	4.1	8.3	95.4	98.1	4.2	8.2
5	96.3	97.6	4.4	8.5	96.2	98.4	4.7	9.3	96.2	98.4	4.7	9.3
6	96.3	97.8	4.1	7.6	96.3	98.2	4.5	8.6	96.3	98.2	2.5	8.6

- The most interpretable (MI), i.e. the classifier with highest semantic cointension index:

  $$\mathrm{MI} = \arg\max_i \ \mathrm{SC}_i.$$

- The best trade-off (BT). Although not generally conflicting, the most interpretable FRBC may not coincide with the most accurate. Thus, a different index could be useful in selecting a single model that shows the best trade-off between accuracy and interpretability. The Best Trade-off is defined as the model that exhibits, on the average, the high accuracy and high interpretability. Formally:

  $$BT = \arg\max_i \left((A_i - \overline{A}) + (\mathrm{SC}_i - \overline{\mathrm{SC}}) \right),$$

where $\overline{A} = \frac{1}{s}\sum_i A_i$ and $\overline{\mathrm{SC}} = \frac{1}{s}\sum_i \mathrm{SC}_i$.

Finally, as each experimental session consists of ten runs (due to cross-validation), the average values of each feature for MA, MI and BT models have been considered. These values are shown in Tables 5–10.

4.3. Analysis of the results

A number of considerations can be drawn from the analysis of the experimental results.

4.3.1. Accuracy

The proposed methodology is capable of deriving FRBCs whose accuracy is comparable with those related to classifiers obtained by other fuzzy-based techniques (see, e.g. [1,28]), especially when the models labeled as MA are considered. This applies for all datasets.

When the models labeled as MI are considered, we observe a decay of the classification accuracy on test set. The decay is mainly due to the criterion adopted for selecting the MI model, which does not take into account its accuracy on the test set. Nevertheless, for some datasets (such as IRIS and WBC) the decay is negligible.

Table 8
Experimental results on WINE dataset (legend in Table 5).

c	MA model				BT model				MI model			
	A (%)	SC (%)	R	P	A (%)	SC (%)	R	P	A (%)	SC (%)	R	P
3	97.8	63.6	3.4	5.7	95.0	87.0	3.6	7.5	91.7	89.8	3.6	7.8
4	97.7	76.0	4.3	10.0	96.6	85.8	4.0	9.5	87.2	89.7	3.7	8.1
5	95.4	66.6	3.4	5.9	93.7	83.0	4.1	9.6	91.0	83.6	4.0	9.0
6	96.6	69.4	3.8	8.3	94.9	83.1	4.5	10.3	94.3	83.7	4.5	10.3
7	95.4	63.6	5.2	11.5	93.2	82.7	5.5	13.2	86.6	84.3	4.8	11.1
8	98.3	60.2	5.7	14.2	91.1	79.6	7.5	20.7	87.6	80.1	7.4	21.0
9	95.4	59.9	7.0	23.3	92.0	73.3	8.2	25.2	83.7	77.3	7.2	20.9

Table 9
Experimental results on IRIS dataset (legend in Table 5).

c	MA model				BT model				MI model			
	A (%)	SC (%)	R	P	A (%)	SC (%)	R	P	A (%)	SC (%)	R	P
3	98.0	86.7	3.3	3.6	97.3	99.3	3.6	4.2	95.3	100.0	3.5	4.0
4	93.3	91.3	3.4	3.8	93.3	100.0	3.4	3.8	92.0	100.0	3.3	3.6
5	92.0	90.7	3.7	4.6	91.3	96.0	4.0	5.3	88.7	96.3	3.7	4.7
6	96.7	84.0	3.3	3.6	96.0	99.3	4.1	5.9	94.7	100.0	4.1	5.9
7	96.7	80.7	3.0	3.0	95.3	99.3	6.1	12.4	94.6	100.0	6.4	13.4
8	96.0	77.3	3.0	3.0	96.0	96.7	6.0	11.1	95.3	97.3	6.0	11.1
9	98.0	77.3	3.5	4.3	97.3	96.7	6.6	12.7	96.0	97.3	6.3	12.0

Table 10
Experimental results on IONOSPHERE dataset (legend in Table 5).

c	MA model				BT model				MI model			
	A (%)	SC (%)	R	P	A (%)	SC (%)	R	P	A (%)	SC (%)	R	P
2	85.6	83.7	4.4	10.9	84.0	87.5	4.1	9.6	77.0	90.6	2.6	4.5
3	88.4	77.2	4.6	12.0	84.9	86.9	4.0	9.2	64.8	92.7	2.0	2.2
4	85.8	82.3	3.8	8.5	82.0	89.4	3.6	7.9	69.9	96.0	2.2	2.4
5	86.7	80.0	4.2	10.9	82.2	88.1	3.8	9.1	63.6	93.9	2.0	2.0
6	85.5	75.6	4.7	13.5	80.6	85.3	3.6	8.3	64.4	93.1	2.2	3.3

When the Best Trade-off criterion is adopted we observe that the proposed methodology is able to provide both accurate and semantically cointensive models. The accuracy worsening of BT models with respect to those labeled as MA is below 2% in most cases (with the exception of the GLASS dataset, remarkably difficult to classify, where worsening is about 6%).

As a general remark involving both MA and BT models, we observe that in most cases accuracy increases as the number c of clusters increases. There are some exceptions (see, for instance, Tables 8 and 10) which could explained by referring to model overfitting or cluster interference that may affect the generation of information granules. As a consequence, the task of choosing the correct hyper-parameter is tightly problem-specific and it should be tackled by performing a number of trials.

4.3.2. Semantic cointension

Experimental results put in evidence some remarkable issues concerning the evaluation of semantic cointension. We firstly observe that high values of semantic cointension have been achieved for most datasets (up to 100% for the IRIS problem). More interestingly, the BT models on these datasets show both high accuracy and high semantic cointension. This demonstrates that it is possible to derive efficient predictive models while taking into account cointension with human understanding: this should be a key-point whenever fuzzy inference is adopted to tackle classification problems. In this way, interpretability concerns are addressed escaping the common interpretability/accuracy trade-off related to those kinds of evaluations which are restricted only to structural analysis of fuzzy models.

As concerning the GLASS dataset, the proposed methodology did not produce high SC values for the selected models, even when the MI criterion is applied. This can be explained by referring to the specific classification problem which is ill-defined being described by a small amount of data and an uneven distribution of classes. As a consequence, the derived knowledge bases are extremely fitted into the training set, and any subsequent modification of the rules (even when performed by truth-preserving operators) may sensibly alter their classification behavior over the test set. We may conclude that the GLASS classification problem is not only difficult to learn, but also difficult to explain in terms of propositional rules.

Table 11
Example of a rule base derived for IRIS data, before and after minimization.

Rule	Premises	Class
Before minimization		
R1	IF *S.L.* IS Short AND *S.W.* IS Wide AND *P.L.* IS Short AND *P.W.* IS Slim	Setosa
R2	IF *S.L.* IS Normal AND *S.W.* IS Slim AND *P.L.* IS Normal AND *P.W.* IS Normal	Versicolor
R3	IF *S.L.* IS long AND *S.W.* IS Normal AND *P.L.* IS Long AND *P.W.* IS Wide	Virginica
R4	IF *S.L.* IS short AND *S.W.* IS Normal AND *P.L.* IS Short AND *P.W.* IS Slim	Setosa
R5	IF *S.L.* IS long AND *S.W.* IS Slim AND *P.L.* IS Long AND *P.W.* IS Wide	Virginica
R6	IF *S.L.* IS long AND *S.W.* IS Normal AND *P.L.* IS Normal AND *P.W.* IS Normal	Versicolor
After minimization		
R1′	IF *P.W.* IS Slim	Setosa
R2′	IF *P.W.* IS Normal	Versicolor
R3′	IF *P.W.* IS Wide	Virginica

Note: S.L., sepal length; S.W., sepal width; P.L., petal length; P.W., petal width.

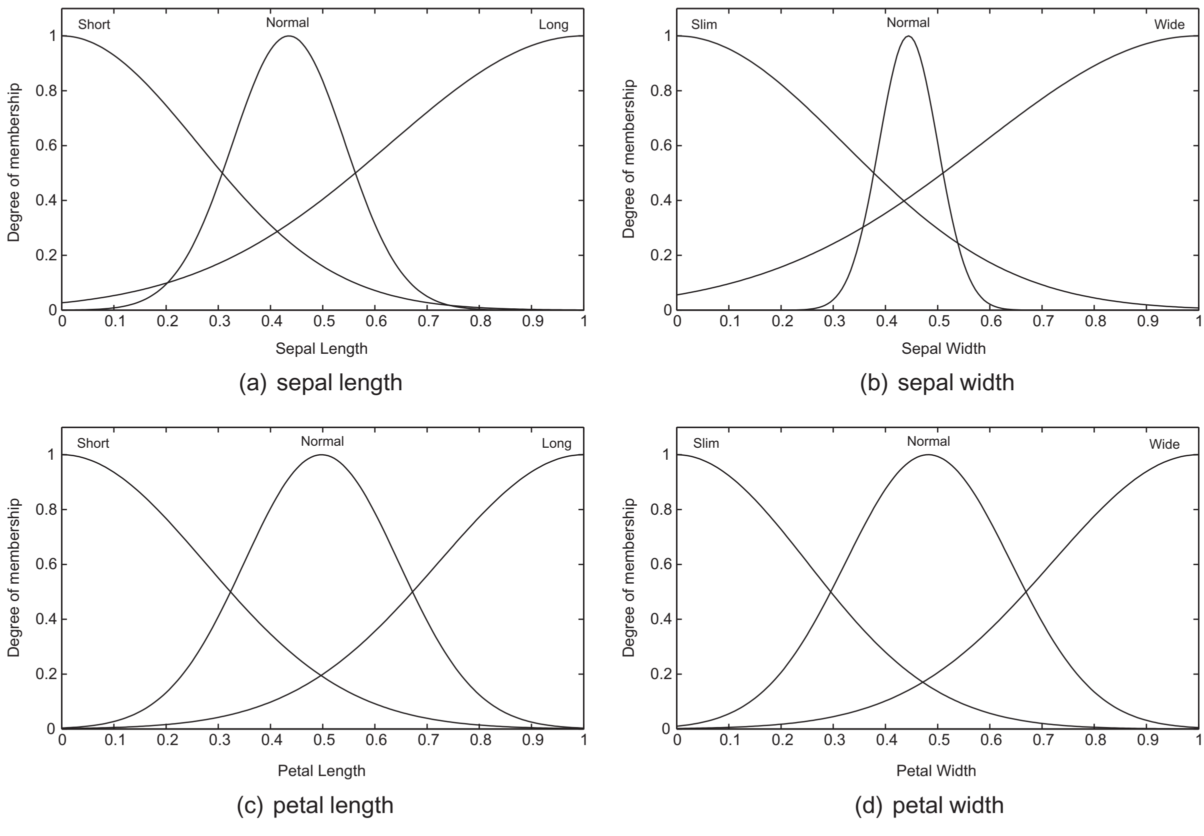

Fig. 5. Linguistic variables for the Iris data.

Table 12
Experimental results of HILK methodology (from [5]).

	Best accuracy			Best trade-off		
	A (%)	R	P	A (%)	R	P
BUPA	64.1	8.6	21.6	57.7	5.4	11.1
WBC	96.5	28.7	90.2	94.2	4.4	7.7
WINE	94.4	12.2	33.3	92.7	6.2	14.5
GLASS	65.8	96.7	382.2	62.6	15.8	44.7

In order to appreciate the effects of the minimization process, we report in Table 11 an example of a rule base derived for the IRIS dataset, and in Fig. 5 a graphical representation of the linguistic variables. Before minimization, the rule base consists of six rules with four atoms each, and its classification accuracy on the test set is equal to 87.4%. The minimization procedure derived a logically equivalent rule base composed by three rules only (one for each class) and a single atom per rule is involved. The semantic cointension (SC) index of the rule bases is 88.2%, i.e. the two rule bases provide the same classification in about nine out of ten cases. We deduce that the original rule base is very interpretable in the sense of the logical view, although for a small part of the samples (namely, for 11.8% of them) the classification cannot be explained by looking at the logical structure of the rules, but requires an inspection on the semantic definition of each linguistic value.

4.3.3. Model complexity

Keeping in mind that structural analyses concerning the complexity of a fuzzy model, besides representing the most widespread mechanisms for comprehensibility evaluation, are relevant for assessing the overall readability of a FRBC, we are interested also in comparing our results with other models, taking into account the structural features. Table 12 reports the accuracy performance (evaluated on test set) and the mean number of rules and premises of the FRBCs obtained by the HILK framework [5]: they have been built with specific regard for interpretability (from a structural point of view) and have been tested on some of the datasets employed in the present work[3]. It can be observed how the models derived by the methodology we have proposed are in line with those reported in literature, both in terms of accuracy and – most notably – in terms of complexity degree.

5. Conclusion

Interpretability is one of the most challenging issues to be addressed when designing intelligent systems based on fuzzy knowledge bases. The challenges arise from the definition of interpretability itself, which appears only partially formalizable, subjective and blur. In this paper we cast the definition of interpretability in the realm of semantic cointension, which complements the usual notion of structural interpretability that is based on a number of constraints on the structure of a fuzzy model.

We have confined our research on fuzzy rule-based classifiers (FRBCs) since they enable an effective way to formalize (at least partially) semantic cointension through the logical view, i.e. the set of logical laws that FRBC rules are supposed to hold since they resemble logical. The paper is focused on the definition of a methodology for designing FRBCs from data with the constraint of semantic cointension. The design methodology is based on the DC*f* framework that is capable of acquiring FRBCs that satisfy a number of interpretability constraints, and on an iterative procedure that incrementally enriches the derived FRBCs by adding rules. An index that quantifies the semantic cointension of the FRBCs is used in combination with accuracy measurements to select the best FRBCs that show both high accuracy and semantic cointension.

The proposed methodology has undergone an experimentation on a number of benchmark datasets. The experimental results show that for most datasets the proposed methodology is capable of providing fuzzy classifiers that are accurate (in comparison to affine models), show high interpretability from the structural viewpoint and, most interestingly, are highly cointensive with the logical view. This enables users in reading and understanding the rules by simply relying on their linguistic representation, which is the basis for any form of knowledge communication.

In our work we focused on semantic cointension of fuzzy rules, without taking into account the complexity of the derived rules. Complexity, indeed, concerns the readability of a knowledge base and is managed through different structural measurements, such as the number of elements (fuzzy sets, linguistic variables, rules, etc.) and other constraints. We chose to not mix semantic cointension evaluation with readability measures so as to highlight the effects of the former in model design. A complete tool for designing interpretable fuzzy rule-based classifiers should, of course, take into account both the dimensions of semantic cointesion and readability to be useful in real scenarios.

The proposed methodology is open to further refinements that, expectedly, will be aimed at increasing the semantic cointension of the derived fuzzy models.

References

[1] J. Abonyi, H. Roubos, R. Babuška, F. Szeifert, Interpretable semi-mechanistic fuzzy models by clustering, OLS and FIS model reduction, in: J. Casillas, O. Cordón, F. Herrera, L. Magdalena (Eds.), Interpretability Issues in Fuzzy Modeling, Springer-Verlag, Heidelberg, 2003, pp. 221–248.
[2] J. 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: WCCI2010 IEEE World Congress on Computational Intelligence, Barcelona, IEEE, 2010, pp. 961–968.
[3] J. Alonso, L. Magdalena, HILK++: an interpretability-guided fuzzy modeling methodology for learning readable and comprehensible fuzzy rule-based classifiers, Soft Computing – A Fusion of Foundations, Methodologies and Applications (2010) 1–22.
[4] 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 (1) (2009) 115–134.
[5] 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.
[6] 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.

[3] We refer to the results reported in [5], where not all the datasets considered in this experimentation have been used.

[7] P. Baranyi, Y. Yam, D. Tikk, R.J. 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.
[8] A. Bargiela, W. Pedrycz, Granular Computing: An Introduction, Kluwer Academic Publishers, Boston, Dordrecht, London, 2003.
[9] J. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, New York, 1981.
[10] A. Botta, B. Lazzerini, F. Marcelloni, D.C. Stefanescu, Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index, Soft Computing 13 (5) (2009) 437–449.
[11] R.K. Brayton, G.D. Hachtel, C. McMullen, A. Sangiovanni-Vincentelli, Logic Minimization Algorithms for VLSI Synthesis, Kluwer Academic Publishers Group, 1984.
[12] R. Cannone, C. Castiello, C. Mencar, A.M. Fanelli, A study on interpretability conditions for fuzzy rule-based classifiers, in: Proceedings of the Ninth International Conference on Intelligent System Design and Applications, Pisa, Italy, 2009, pp. 438–443.
[13] J. Casillas, O. Cordón, F. Herrera, L. Magdalena (Eds.), Interpretability Issues in Fuzzy Modeling, Springer, Germany, 2003.
[14] G. Castellano, A. 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.
[15] G. Castellano, A.M. Fanelli, C. Mencar, A neuro-fuzzy network to generate human understandable knowledge from data, Cognitive Systems Research, Special Issue on Computational Cognitive Modeling 3 (2) (2002) 125–144.
[16] G. Castellano, A.M. Fanelli, C. Mencar, DCf: a double clustering framework for fuzzy information granulation, in: Proceedings of IEEE International Conference on Granular Computing, 25–27 July, 2005, vol. 2, 2005, pp. 397–400.
[17] 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.
[18] 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.
[19] 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.
[20] 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.
[21] A. Gobi, W. Pedrycz, Logic minimization as an efficient means of fuzzy structure discovery, IEEE Transactions on Fuzzy Systems 16 (3) (2008) 553–566.
[22] S. Guillaume, B. Charnomordic, Generating an interpretable family of fuzzy partitions from data, IEEE Transactions on Fuzzy Systems 12 (3) (2004) 324–335.
[23] E. Hemaspaandra, G. Wechsung, The minimization problem for Boolean formulas, in: Annual IEEE Symposium on Foundations of Computer Science, 1997, pp. 575–584.
[24] 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.
[25] M. Jamei, M. Mahfouf, D.A. Linkens, Elicitation and fine tuning of Mamdani-type fuzzy rules using symbiotic evolution, in: Proceedings of European Symposium on Intelligent Technologies, Hybrid Systems and their Implementation on Smart Adaptive Systems (EUNITE 2001), Tenerife, Spain, 2001, pp. 361–367.
[26] F. Jiménez, A. Gómez-Skarmeta, H. Roubos, R. Babuška, A multi-objective evolutionary algorithm for fuzzy modeling, in: Proceedings of the 2001 Conference of the North American Fuzzy Information Processing Society, NAFIPS'01, New York, 2001, pp. 1222–1228.
[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] U. Johansson, R. König, L. Niklasson, Genetically evolved kNN ensembles, in: R. Stahlbock, S.F. Crone, S. Lessmann (Eds.), Data Mining, Annals of Information Systems, vol. 8, Springer, US, 2010, pp. 299–313.
[29] K.-L. Du, Clustering: a neural network approach, Neural Networks 23 (1) (2010) 89–107.
[30] A.A. Marquez, F.A. 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 (FUZZ), 2010, pp. 277–283.
[31] 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.
[32] C. Mencar, C. Castiello, R. Cannone, A. Fanelli, Interpretability assessment of fuzzy knowledge bases: a cointension based approach, International Journal of Approximate Reasoning 52 (4) (2011) 501–518.
[33] C. Mencar, A.M. Fanelli, Interpretability constraints for fuzzy information granulation, Information Sciences 178 (24) (2008) 4585–4618.
[34] D. Nauck, Measuring interpretability in rule-based classification systems, in: Proceedings of the 12th IEEE International Conference on Fuzzy Systems, vol. 1, FUZZ'03, 2003, pp. 196–201.
[35] D. Nauck, R. Kruse, A neuro-fuzzy approach to obtain interpretable fuzzy systems for function approximation, in: Proceedings of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'98), Anchorage, AK, 1998, pp. 1106–1111.
[36] D. Nauck, R. Kruse, Obtaining interpretable fuzzy classification rules from medical data, Artificial Intelligence in Medicine 16 (1999) 149–169.
[37] R. Paiva, A. Dourado, Merging and constrained learning for interpretability in neuro-fuzzy systems, in: Proceedings of European Symposium on Intelligent Technologies, Hybrid Systems and their Implementation on Smart Adaptive Systems (EUNITE 2001), Tenerife, Spain, December 2001, pp. 17–22.
[38] W. Pedrycz, F. Gomide, An Introduction to Fuzzy Sets, Analysis and Design, MIT Press, Cambridge, MA, 1998.
[39] W. Pedrycz, F. Gomide, Fuzzy Systems Engineering: Toward Human-Centric Computing, Wiley and Sons, 2007.
[40] 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, Springer-Verlag, 2003, pp. 119–146.
[41] A. Riid, E. Rustern, Interpretability of fuzzy systems and its application to process control, in: Proceedings of IEEE International Conference on Fuzzy Systems Conference – FUZZ-IEEE 2007, July 2007, pp. 228–233.
[42] 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.
[43] 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.
[44] M. Setnes, R. Babuška, U. Kaymak, H.R. Van Nauta Lemke, Similarity measures in fuzzy rule base simplification, IEEE Transactions on Systems, Man and Cybernetics, Part B – Cybernetics 28 (3) (1998) 376–386.
[45] M. Setnes, R. Babuska, H.B. Verbruggen, Rule-based modeling: precision and transparency, IEEE Transactions on Systems, Man and Cybernetics, Part C – Applications and Reviews 28 (1) (1998) 165–169.
[46] D. Tikk, P. Baranyi, Exact trade-off between approximation accuracy and interpretability: solving the saturation problem for certain FRBs, in: J. Casillas, O. Cordón, F. Herrera, L. Magdalena (Eds.), Interpretability Issues in Fuzzy Modeling, Springer-Verlag, Heidelberg, 2003, pp. 587–604.
[47] J. Valente de Oliveira, Semantic constraints for membership function optimization, IEEE Transactions on Systems, Man and Cybernetics, Part A – Systems and Humans 29 (1) (1999) 128–138.

[48] L. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, Fuzzy Sets and Systems 90 (1997) 111–117.
[49] L.A. Zadeh, Fuzzy logic = computing with words, IEEE Transactions on Fuzzy Systems 4 (1996) 103–111.
[50] L.A. Zadeh, Is there a need for fuzzy logic?, Information Sciences 178 (2008) 2751–2779
[51] S.-M. Zhou, J. 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.
[52] 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.