A logic-based approach for evaluating interpretability of fuzzy rule-based classifiers

Corrado Mencar Ciro Castiello Anna M. Fanelli

Dept. of Informatics, University of Bari
Bari, Italy
Email: {mencar, castiello, fanelli}@di.uniba.it

Abstract— *We describe an automatic approach for evaluating interpretability of fuzzy rule-based classifiers. The approach is based on the logical view of fuzzy rules, which are interpreted as rows in truth tables. These truth tables are subject of a minimization procedure based on a variant of the Quine-McCluskey algorithm. The minimized truth tables are used to build new fuzzy rules, which are compared with the original ones in terms of classification accuracy. If the two sets of rules have similar performances, we deduce that the logical view of rules is applicable to the fuzzy classifier, which is hence considered interpretable. On the other hand, a significant difference in classification ability shows that fuzzy rules cannot be interpreted in logical terms, hence linguistic labelling may not be significant. Two illustrative examples show both the cases.*

Keywords— fuzzy rule-based classifiers, interpretability assessment, logic minimization, Quine McCluskey algorithm

1 Introduction

Interpretability is one of the most important driving forces for the adoption of fuzzy rule-based systems . The reason is clear: several models exist that can perform intelligent tasks (such as prediction, classification, etc.) but fuzzy systems allow for a representation of knowledge that can be easily read and understood by their users. Interpretability, however, is *not* given for granted when fuzzy models are used, especially when they are acquired from data. The main problem is that data-driven design has a great number of degrees of freedom (number of fuzzy sets, their shape, position, etc.) and may end up with fuzzy models that are very accurate but very hard to comprehend. For this reason, interpretability constraints have been defined so as to bind data-driven design in order to derive interpretable fuzzy models [1]. This usually comes to a price that is a lower accuracy with respect to unconstrained design. Furthermore, often interpretability is accounted without taking care of accuracy. This approach has been criticized, since interpretable but inaccurate models are as useless as very accurate but not interpretable ones [2].

Interpretability can be viewed at different levels of a fuzzy model. Interpretability of fuzzy sets involves the possibility of tagging them with linguistic labels. On a higher level, interpretability of partitions concerns the possibility of linguistically describe the domain of data. On the rule-level, which is of our concern in this paper, interpretability is viewed as the ability of explaining relationships between input and output variables in a linguistic form.

An important issue concerns the evaluation of interpretability. Interpretability assessment is important because several fuzzy rule-based models can be derived from the same data. Many of these models could provide for high accuracies but can vary in their interpretability. As a consequence, a tool for interpretability assessment could be helpful in the choice of the final model or, to a greater extent, the design technique. However, interpretability assessment is an ill-posed problem because the definition of interpretability eludes any formal characterization.

In [3], Michalski gives a referential definition of interpretability, the so-called “comprehensibility postulate”, reported in the following.

> 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.

This postulate justifies the use of linguistic values in rule-based fuzzy systems, but that is not enough to guarantee interpretability (for a more extensive discussion of the postulate, see [4]). In [5] Zadeh introduces the notion of *co-intension*, a semantic relation between concepts. Roughly speaking, two concepts are co-intensive if they refer to almost the same entities. In fuzzy rule-based systems, rules are defined by composition of linguistic terms, which are related to two different semantics. The first one is defined by the fuzzy model and the second one is implicitly designated by the linguistic label. By merging the notion of co-intension with the comprehensibility postulate, we derive a formulation of interpretability that can be more helpful for designing assessing techniques: *a rule base is interpretable if the two semantics related to each linguistic term are co-intensive*.

On the basis of this definition, we propose an automatic technique for evaluating interpretability. Our approach evaluates interpretability by assessing the co-intension of the semantics of the rule base of a fuzzy model with the intrinsic semantics designated by linguistic labels. The core of the evaluation technique relies on the propositional view of rules and on logical operations. We expect that, for interpretable knowledge bases, logical operations on rules do not change their semantics and, hence, do not affect accuracy. If this is assumption is violated, we deduce a lack of interpretability.

We focus our research on fuzzy rule-based classifiers, described in Section 2. The proposed approach is then described, by first focusing on its rationale and then on the methodology

ISBN: 978-989-95079-6-8

(Sections 3.1 and 3.2, respectively). Two illustrative examples are reported in Section 4, to show how the proposed approach can be useful in detecting lacks of interpretability for two simple knowledge bases. The paper is concluded with some final remarks.

2 Fuzzy rule-based classifiers

We consider a classifier as a system computing a function of type:

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

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 fuzzy rule-based classifier (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[1] 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 (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);
- $S_j = \{s_{j1}, s_{j2}, \ldots, s_{jm_j+1}\}$ 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 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:

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

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.[2]

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:

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

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 minimum or product.

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

$$\mu_{\lambda_i}(\mathbf{x}) = \frac{\sum_{r=1}^{R} \mu_r(\mathbf{x}) \chi(\lambda_i, \lambda^{(r)})}{\sum_{r=1}^{R} \mu_r(\mathbf{x})}. \tag{7}$$

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

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

3 Interpretability assessment

We assume the availability of an interpretable FRBC, verifying a number of interpretability constraints so that the rule base is described in terms of linguistic values.

Interpretability of FRBC is usually achieved by imposing a number of constraints on the knowledge base. For our purposes, it is necessary that labels denoting linguistic values for each variable can be interpreted as well-distinguished elementary concepts. To achieve this goal, normality, convexity and distinguishability are the basic interpretability constraints to be imposed on the FRBC. Normality is indeed required to represent consistent concepts (i.e. concepts that are fully applicable to at least one object), convexity is necessary to represent elementary concepts (i.e. concepts that cannot be split in two or more sub-concepts) and distinguishability allows the use of different labels to denote different fuzzy sets (for a detailed explanation, the reader is referred to [1]).

3.1 Rationale

The proposed approach for interpretability assessment relies on the formal structure of the FRBC. The rationale behind this approach comes from the observation that the rule base is the linguistic interface of the FRBC to the user. For an interpretable knowledge base, the user should be able to understand the classification rules by simply observing their linguistic representation. All the semantic information (fuzzy sets attached to linguistic values, t-norm used for conjunction, etc.) should be hidden to the user because – this is the key point of interpretability – the semantics of FRBC knowledge should be co-intensive with the user's knowledge, recalled by the linguistic terms.

The proposed approach tries to evaluate how much co-intensive is the semantics of the FRBC knowledge w.r.t. user knowledge. As a further requirement, we want the evaluation

[1] For the sake of clarity, in this paper we use a simplified definition of linguistic variable.

[2] The sequence NOT ANY is not allowed.

ISBN: 978-989-95079-6-8

process to be carried out automatically. To clarify its rationale, we consider the example of two communicating actors, A and B. When A communicates some piece of knowledge to B, he/she uses a linguistic structure that (at least partially) represents his/her knowledge. To be comprehensible, A chooses linguistic terms whose semantics is deemed as co-intensive as possible with B's knowledge. 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 knowledge communication. The actor A can be almost sure of the co-intension with B if they share a similar environment, language, culture, experiences, etc. Thus, co-intension can be achieved if A and B share similar cognitive structures.

To assess interpretability, we exploit the cognitive structures that are shared by users and FRBC. In particular, we observe a strict affinity of a FRBC rule base to logical propositions. Actually, rules are formed so as to resemble propositions, so that they can be understood by users. In consequence of this, FRBC and users share this propositional view of rules. Being like propositions, rules could be transformed by truth-preserving operators without any change of the semantics. This is not completely true since the application of such operators may distort the semantics of rule (which is fuzzy); however, we should expect only small distortions.

The core of our approach is the following: given a rule base of a FRBC, we represent it as a collection of propositions, and then we transform it through a truth-preserving operator, thus obtaining a new set of propositions, that represents a new rule base. We then compare the two rule bases on the basis of their classification ability: if they differ too much, then we conclude that the logical view of rules is wrong. Also, since rules are defined so as to resemble logical propositions, we derive that the semantics of rules (which is responsible of classification) is such that logical view is not possible. This means that the semantics of rules is not co-intensive with user knowledge, since users expect that truth-preserving transformations of propositions do not change (too much) their semantics.

A simple example may further clarify the rationale of our approach. Suppose to have a trivial FRBC with the following two rules:

```
IF cell_size is Large AND
   cell_shape is Irregular THEN malign
IF cell_size is Small THEN benign
```

Any user reading this rule base understands that the size of a cell is enough to discriminate between malign and benign classes. As a consequence, a truth-preserving transformation of the rule base could lead to the following set of rules:

```
IF cell_size is Large THEN malign
IF cell_size is Not Large THEN benign
```

This understanding is legitimate under the Closed World Assumption (CWA), according to which any input satisfies the condition of at least one rule. CWA is generally valid in most fuzzy systems and FRBC in particular, hence it is taken for grant in this work. Open World Assumption (OWA), which leads to non-monotonic reasoning, is out of the scope of this paper.

The two rule bases can be compared in terms of classification errors. If the two classification errors are very similar, we derive that the semantics of the rule base is compatible with the applied transformation and, hence, to the logical view of rules. In this sense, we state that the knowledge base of the FRBC is co-intensive 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, it is not co-intensive with user mind. In other words, it is not interpretable.

We cannot expect that classification errors are always identical, even for very interpretable FRBC, because the law of excluded middle ($A\bar{A} = 0$) is not valid in fuzzy logic but it holds in Boolean logic. On the other hand, we should expect that the violation of excluded middle does not influence too much the inference process, otherwise we should conclude that the FRBC is heavily based on inconsistent knowledge. This latter situation clashes with the property of interpretability.

An issue arises in the choice of the truth-preserving transformation. Actually, several transformations can be considered, but we choose to apply a transformation that minimizes the number of linguistic terms used. This choice has a twofold advantage. First, by eliminating as many terms as possible, we test if logical view of rules is preserved with the minimum required information. Second, if assessment leads to positive results, we could retain the simplified rule base because it is easier to read than the original.

3.2 Methodology

The proposed approach for interpretability assessment is based on a four-stage strategy. It is similar to the approach proposed in [8], however our strategy is able to deal with negative information to achieve more compact representations of the rule bases.

3.2.1 Definition of truth tables

Each rule of the FRBC is seen as a proposition, i.e. a combination of propositional variables that is considered true for a class.

For each class label $\lambda_i \in \Lambda$ and for each rule r, a truth function π_i is defined on the propositional variables defined for the FRBC as:

$$\pi_i(\chi_{11}^{(r)}, \ldots, \chi_{1m_1}^{(r)}, \chi_{n1}^{(r)}, \ldots, \chi_{nm_n}^{(r)}) = \chi(\lambda_i, \lambda^{(r)}), \qquad (9)$$

being $\chi_{jk}^{(r)} = \chi(q_j^{(r)}, q_{jk})$. Inputs $\chi_{jk}^{(r)}$ assume value DC ("don't care") if the corresponding linguistic value is ANY or, in other words, the linguistic variable V_j does not occur in the r-th rule. 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

ISBN: 978-989-95079-6-8

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

Table 1: The truth tables of the simple rulebase

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

Table 2: The minimized truth tables of the simple rulebase

would be expected in the general case of truth function representation. Table 1 shows the truth tables for classes "malign" and "benign" of the previous example.

3.2.2 Minimization

Once each truth table has been built, it can be processed so as to be minimized. The minimization procedure produces a new truth table without modifying the truth function (where it is defined). The new truth table has a number of rows not greater than the original truth table. It also has a number of DC values in its inputs not smaller than in the original truth table. Furthermore, minimization 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 minimization of truth tables [7]. 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 DC that differ in only one input;
2. Find the minimum number of merged rows that cover all rows of the original truth table.

Although computationally expensive, the QMC algorithm can be implemented by an efficient procedure that exploits the peculiar structure of truth tables derived from FRBC rules to perform minimization 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 can be optimized by using heuristic procedures that drive the minimization process without expensive search. Table 2 shows the minimized truth table for the example FRBC.

3.2.3 Reconstruction

After minimization, a new FRBC is built from the rows of the minimized truth table. For each class label $\lambda_i \in \Lambda$ we consider the minimized table associated to the truth function π_i. A rule is built for each row with output equal to 1. It is easy to show that for each j there is at most one k such that $\chi_{jk}^{(r)} \neq DC$. Therefore, the antecedent of the rule can be defined by atoms v_j IS [NOT] $q_j^{(r)}$ where:

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

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

At the end of reconstruction, each rule is formed by a sequence of atoms tied together by conjunction (AND). In this paper we do not deal with more complex representations (such as conjunctions of disjunctions of atoms) that could provide for more compact representations of rules but deserve further investigation that will be object of future research.

3.2.4 Comparison

The two FRBCs, the first with the original rule base and the second with the minimized version, are compared in terms of classification error on the same data.

If the two errors differ too much, we conclude that the original FRBC lacks of interpretability. Its accuracy is mainly due to the semantic definition of linguistic values, which do 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 co-intensive with user knowledge. Attaching natural language terms to such FRBC is useless and potentially misleading.

If the two errors are very similar, 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 semantic of linguistic values is co-intensive with user knowledge. We could retain the simplified FRBC because its interpretability is greater than the original (due to its higher simplicity) while its accuracy is almost the same of the original.

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. Even with small variations of accuracy, some considerations can be drawn on the interpretability of the FRBC, as shown in the next Section.

4 Illustrative examples

In this Section we evaluate two different FRBC derived from the same data set with two different methods. The first method is oriented toward interpretability while the second is more inclined to the design of accurate models, albeit taking into account some kind of readability of the knowledge base.

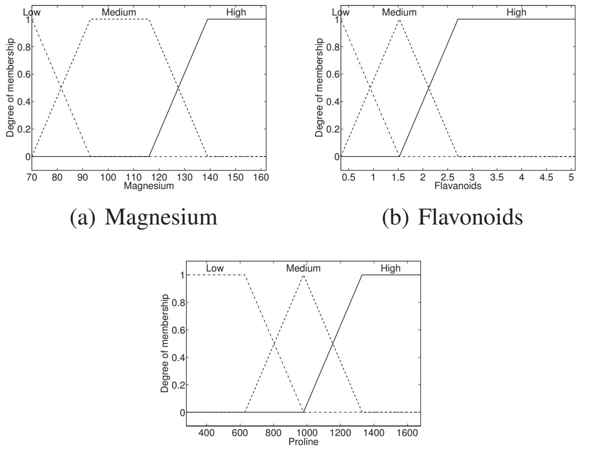
4.1 Interpretable FRBC

We consider a FRBC obtained from the application of HILK [6], a tool for building interpretable fuzzy rule-based systems. HILK allows for the definition of fuzzy rule-based systems from empirical learning, expert knowledge, or both. The resulting systems are considered highly interpretable because fuzzy sets defined for each input satisfy a number of interpretability constraints. Furthermore, the number of fuzzy sets per input, the number of inputs and the number of rules are kept as small as possible since, in principle, the simplest is the knowledge representation, the easier is to read and understand.

ISBN: 978-989-95079-6-8

If Flavonoids is Low THEN Class 1
If Flavonoids is Medium THEN Class 2
If Flavonoids is High AND Proline is Low THEN Class 2
If Flavonoids is High AND Proline is High THEN Class 3
If Magnesium is Medium AND Flavonoids is High AND Proline is Medium THEN Class 3

Figure 1: The rule base of the FRBC considered in the first example



(a) Magnesium

(b) Flavonoids

(c) Proline

Figure 2: The linguistic variables used in the first example

In our experimentation, the FRBC obtained from HILK was acquired from data, in order to classify Wine data, a well-known benchmark dataset, freely available from UCI repository [10]. The knowledge base of the FRBC is reported in fig. 1, while the linguistic variables are shown in fig. 2. The FRBC provided 10.67% of classification error on the entire dataset.

We transformed the rule base of the FRBC into three truth tables – one for each class – minimizing them with QMC algorithm; then we rebuilt the FRBC obtaining the simplified rule base reported in fig. 3.

We observe that the number of rules has been reduced to four, and the linguistic variable "Magnesium" has been removed. From the logical viewpoint, the two rule bases are equivalent since they compute the same truth functions. However, after applying the minimized rule base to the dataset, we obtained 11.24% of classification error, i.e. an increase of $+0.57\%$, corresponding to one more misclassified pattern over 178, which is defined as (Magnesium=162, Flavonoids=2.27, Proline=937). For the misclassified pattern we observe that, according to the definitions of the linguistic variables, Magnesium is high, Proline is medium, and Flavonoids is both medium and high to a significant degree. However, in the original rule base, there is not any classification rule activated when Magnesium is high, Proline is medium and Flavonoids is high. From the logical viewpoint, the truth functions are undefined for this condition. As a consequence, the simplified rule base subsumes this condition in the fourth rule, assigning the pattern to class 3, while the original FRBC arbitrarily assigns the pattern to class 2. The correct classification of the pattern performed by the original FRBC

If Flavonoids is Low THEN Class 1
If Flavonoids is Medium THEN Class 2
If Flavonoids is High AND Proline is Low THEN Class 2
If Flavonoids is High AND Proline is NOT Low THEN Class 3

Figure 3: The rule base of the FRBC in the first example after minimization

If Alc. is High And Fla. is High And A.Ash is Low And OD2 is High And Pro. is High Then class 1
If Alc. is Low And Pro. is Low Then class 2
If Alc is Medium And A.Ash is High And Fla. is Low And OD2 is Low And Pro. is Low Then class 3

Figure 4: The original rule base of 2nd example (Alc. = Alcohol, Fla. = Flavonoids, A. Ash = Alcalinity of Ash, OD2 = OD280/315, Pro. = Proline)

is due to the semantic specification of linguistic values, which does not emerge from the propositional view of rules. In this sense, the original FRBC lacks of interpretability for a specific case. However, due to the reduced increase of classification error, we conclude that the original FRBC is highly interpretable and we can retain the simplified version, which offers a further comprehensible knowledge base.

4.2 Non-interpretable FRBC

As a second example, we consider the FRBC obtained through the approach proposed in [9] aimed at optimizing accuracy by taking into account interpretability. In brief, the approach consists of an iterative procedure for developing the FRBC. The initial model is derived from the data and, subsequently, feature selection and rule-base simplification are applied to reduce the model, while a genetic algorithm is used for parameter optimization.

This approach has been applied to Wine data showing very good classification accuracy (1.1% of classification error[3]). One of the best rule bases derived from data is reported in fig. 4. We observe from fig. 5 that in some cases interpretability constraints are violated. As an example, we observe in fig. 5(b) that the rightmost fuzzy set is highly overlapping with the other fuzzy sets of the variable. Notwithstanding, the three fuzzy sets have different labels, thus recalling distinct concepts.

After simplification, we obtain the very small rule base reported in fig. 6. The reconstructed FRBC shows an unacceptable classification error of about 32%. In consequence of this considerable decay of classification performance, we deduce that classification accuracy of this FRBC is mainly based on the parameters of the fuzzy sets that cannot be expressed in linguistic terms. In consequence of this, we state that the logical view of rules is not applicable to the original rule base. In this sense, we conclude that the FRBC is not interpretable.

5 Conclusion

Assessment of interpretability is not an easy task. Difficulty is mainly due to the blurry definition of interpretability, which requires co-intension with human knowledge. Evaluating interpretability is a subjective task, which could be tedious and even ill-posed, therefore automatic techniques for

[3]We use the results reported in the original paper

ISBN: 978-989-95079-6-8

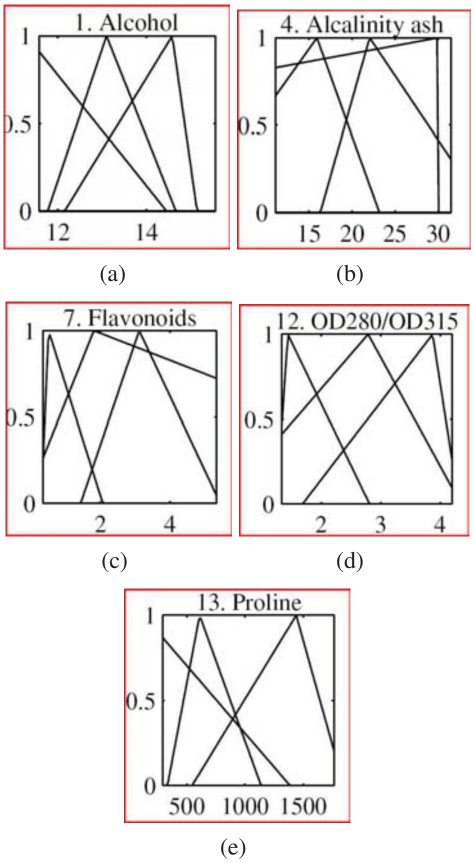


Figure 5: The fuzzy sets used in the second example (for each variable, the leftmost fuzzy set is labelled "Low", the middle one "Medium" and the rightmost "High")

If Alcohol is High Then class 1
If Alcohol is Low Then class 2
If Alcohol is Medium Then class 3

Figure 6: The minimized rule base of 2nd example

interpretability assessment are useful, but they should embody some information on semantic co-intension. In this paper, we have described an approach for automatically evaluating interpretability of rule-based fuzzy classifiers, exploiting the propositional view of rules as a mean to define co-intension. On the basis of this hypothesis, a procedure has been devised, so as to evaluate how much the semantics of fuzzy rules is coherent with their logical view.

The illustrative examples show the effects of the proposed approach for a simple knowledge base. Actually, we were able positively assess the interpretability for a FRBC that verifies several interpretability constraints. Also, we have verified that a FRBC is not interpretable even though it is described in linguistic terms. By analyzing the relative performances of the two FRBCs we immediately observe that interpretability preservation comes to the cost of accuracy reduction and, on the other hand, an accurate description of data (preserving interpretability constraints) would require a great number of rules. This, however, would clash with the Comprehensibility Postulate.

Research on this topic is in progress, especially in the direction of enhancing the efficiency of the minimization algorithm. That will allow an extensive experimentation over a wider class of knowledge bases, in order to promote deeper insights of the proposed technique. Further enhancements could spring from the use of additional information to refine the definition of co-intension in computational terms.

References

[1] C. Mencar. *Theory of Fuzzy Information Granulation: Contributions to Interpretability Issues.* PhD thesis, University of Bari, Italy, 2005

[2] A. Riid, E. Rustern. Interpretability of Fuzzy Systems and Its Application to Process Control. *In: Proc. of IEEE International Fuzzy Systems Conference FUZZ-IEEE 2007*, pp. 1–6, 2007

[3] R.S. Michalski. A theory and methodology of inductive learning. *Technical Report UIUCDCS-R-83-1122*, Department of Computer Science, University of Illinois, Urbana, 1983

[4] 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 (IJUFKS)* 15(5):521 – 537, 2007

[5] L.A. Zadeh. Is there a need for fuzzy logic? *Information Sciences* 178:2751–2779, 2008

[6] 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* 23:761–794, 2008

[7] E.J. McCluskey. Minimization of Boolean functions. *Bell Labs Technical Journal* 35(5):1417–1444,1956

[8] R. Rovatti, R. Guerrieri, G. Baccarani. An enhanced two-level Boolean synthesis methodology for fuzzy rules minimization, *Fuzzy Systems, IEEE Transactions on* 3(3):288–299, 1995

[9] J.A. Roubos, M. Setnes, J. Abonyi. Learning fuzzy classification rules from labeled data. *Information Sciences* 150:77–93,2003

[10] *UCI Machine Learning Repository Content Summary:* http://www.ics.uci.edu/ ˜mlearn/MLSummary.html

ISBN: 978-989-95079-6-8