



Effect of fuzziness in fuzzy rule-based classifiers defined by strong fuzzy partitions and winner-takes-all inference

Gabriella Casalino¹ · Giovanna Castellano¹ · Ciro Castiello¹ · Corrado Mencar¹

Accepted: 8 April 2022 / Published online: 6 May 2022
© The Author(s) 2022

Abstract

We study the impact of fuzziness on the behavior of Fuzzy Rule-Based Classifiers (FRBCs) defined by trapezoidal fuzzy sets forming Strong Fuzzy Partitions. In particular, if an FRBC selects the class related to the rule with the highest activation (so-called Winner-Takes-All approach), then fuzziness, as quantified by the slope of the membership functions, has no impact in classifying data in regions of the input space where rules dominate. On the other hand, fuzziness affects the behaviour of the FRBC in regions where the confidence in classification is low. As a consequence, in the context of Explainable Artificial Intelligence, fuzziness is profitable in FRBCs only if classification is accompanied by an explanation of the confidence of the provided outputs.

Keywords Fuzziness · Strong fuzzy partition · Fuzzy rule-based classifier · XAI

1 Introduction


Explainable artificial intelligence (XAI) is a blooming research field propelled by the increasing demand of intelligent systems which should provide accurate answers to complex problems as well as some kind of human-oriented added values (explanations in choices, rationale and confidence in decisions, possible alternative strategies, and so on) (Hagras 2018). The field of application of XAI spans several areas, including Industry 4.0 scenarios (Lu 2019; Xu et al. 2018). From the methodological viewpoint, there are many ways to embody explainability in intelligent systems, from opening black-box models (Guidotti et al. 2018) to the development of specific methods (Biran and Cotton 2017).

Fuzzy Logic systems have a great potential in the development of XAI solutions. In fact, they are able to express knowledge in a human-oriented fashion thanks to the adoption of a paradigm enabling the use of natural language terms (Computing with Words) (Zadeh 1999). Such a capability allows to provide the users with readable explanations of the embodied knowledge (represented in a perception-based fashion), and may guarantee also illustrative details concerning the inference process behind certain results (Zadeh 2008).

Nevertheless, attention must be paid to the semantics of the formal objects involved in knowledge representation and reasoning. This is to avoid that explanation is only illusory appearance which does not convey any piece of meaningful information. Therefore, when designing an XAI system, the quality of the underlying model should not be evaluated in terms of predictive accuracy only, but also taking into account the capability of generating meaningful information. This is not an easy task; yet it sheds light on new ways of analyzing existing approaches or devising new ones.

In this paper we focus on Fuzzy Rule-Based Classifiers (FRBCs), which are commonly praised for their ability of representing knowledge in an interpretable form (Gorzalczany and Rudziński 2017; Alonso et al. 2008). In essence, an FRBC is based on a knowledge base represented by a collection of rules. They are easy to be read and understood, provided that they have been designed by taking into account interpretability constraints (Alonso Moral et al. 2021). Given

The authors are members of the INdAM Research group GNCS.

✉ Ciro Castiello
ciro.castiello@uniba.it

Gabriella Casalino
gabriella.casalino@uniba.it

Giovanna Castellano
giovanna.castellano@uniba.it

Corrado Mencar
corrado.mencar@uniba.it

¹ University of Bari Aldo Moro, Bari, Italy

an input sample, a fuzzy inference mechanism is triggered, so that the FRBC returns in output a class label. Therefore, an FRBC usually behaves like many other classifiers—not necessarily based on fuzzy logic—but enables a clear interpretation of the knowledge base through the adoption of linguistic terms that reflect the imprecision of perception-based concepts.

In XAI, an obvious step forward consists in endowing FRBCs with the ability to explain the inferred class for a given object. In literature, powerful methods have been designed to give highly comprehensible explanations by using Natural Language Generation (NLG) techniques (Alonso et al. 2017). Usually, such explanation systems provide a symbolic description representing the reasoning process behind the automatic classification; however, the next question is: how much does the fuzziness of the involved linguistic terms affect such explanation? That is the core of this study.

In Sect. 2 we give an account of Strong Fuzzy Partitions (SFPs), which are widely used in FRBC design. Although FRBCs can be designed in different ways, we restrict our attention to SFPs with trapezoidal fuzzy sets because they enable the design of interpretable classification rules, thus explaining their widespread employment. Also, trapezoidal fuzzy sets can be easily designed so as to satisfy SFP constraints, while being flexible enough to adapt to data. The same Section introduces some properties of trapezoidal fuzzy sets that are instrumental for the arguments reported in Sect. 2.1, where FRBCs are formalized and the impact of fuzziness on the classification function is analyzed. The outcomes of an experimental session are reported in Sect. 3 to give a visual and quantitative account of the theoretical results on some synthetic data. Finally, a concluding section discusses the theoretical results from a methodological point of view.

2 Strong fuzzy partitions with trapezoidal fuzzy sets

Let $X = [l, u] \subset \mathbb{R}$ be a Universe of Discourse and let A_1, A_2, \dots, A_{n+1} be a sequence of normal and convex fuzzy sets defined on X . Such a sequence of fuzzy sets constitutes a Strong Fuzzy Partition (SFP) (Dubois et al. 1995; Loquin and Strauss 2006; Perfilieva 2006) provided that:¹

$$\forall x \in X : \sum_{i=1}^{n+1} A_i(x) = 1 \quad (1)$$

Eq. (1) is often referred as *Ruspini condition* after (Ruspini 1969). The employment of SFPs is quite common in

¹ We will denote by A_i both the fuzzy set and its membership function for the sake of conciseness in our notation.

fuzzy modeling, especially when interpretability is a modeling requirement (Alonso Moral et al. 2021).

The membership function of a trapezoidal fuzzy set is a piece-wise linear function constrained by four parameters $a, b, c, d \in X$. Provided that $a \leq b \leq c \leq d$, a trapezoidal fuzzy set $T[a, b, c, d]$ is defined for each $x \in X$ as follows:

$$T[a, b, c, d](x) = \begin{cases} \frac{x-a}{b-a}, & x \in [a, b[\\ 1 & x \in [b, c[\\ \frac{x-d}{c-d}, & x \in [c, d[\\ 0 & x < a \vee x \geq d \end{cases} \quad (2)$$

It should be observed that a trapezoidal fuzzy set collapses to a triangular fuzzy set whenever $b = c$.

Trapezoidal fuzzy sets are convenient in FRBC design because they can be easily constrained in order to generate an SFP. To produce an SFP composed by trapezoidal fuzzy sets with membership functions $A_i = T[a_i, b_i, c_i, d_i]$ ($i = 1, \dots, n+1$), the following conditions must hold:

$$\begin{cases} a_1 = b_1 = l, \\ a_{i+1} = c_i, & (i = 1, \dots, n) \\ b_{i+1} = d_i, & (i = 1, \dots, n) \\ c_{n+1} = d_{n+1} = u \end{cases} \quad (3)$$

The intersection point between two contiguous fuzzy sets A_i, A_{i+1} is called *cut-point* and it is denoted by t_i ($i = 1, \dots, n$).² It is easy to verify that

$$A_i(t_i) = A_{i+1}(t_i) = 0.5$$

In this way, a sequence t_1, t_2, \dots, t_n of cut-points is defined such that $t_{i-1} \leq t_i$, which can be extended by including $t_0 = l$ and $t_{n+1} = u$. As a consequence, the 0.5-cut of a trapezoidal fuzzy set

$$[A_i]_{0.5} = \{x \in X : A_i(x) \geq 0.5\}$$

coincides with an interval identified by consecutive *cut-points*:

$$[A_i]_{0.5} = [t_{i-1}, t_i] \quad 1 \leq i \leq n+1 \quad (4)$$

A key feature of a trapezoidal fuzzy set is the slope of the left and right boundaries, which are informally identified as the two areas of the domain where the membership degrees are neither 0 nor 1. Formally, the left and right boundaries are the open intervals $]a, b[$ and $]c, d[$ respectively. (One of the two boundaries can be empty for the leftmost and

² We used the term *cut* in our previous papers (Castiello and Mencar 2019; Castiello et al. 2019; Mencar et al. 2013).

rightmost fuzzy sets; both boundaries are empty in the case of a singleton fuzzy set.)

In case of non-empty boundaries, the corresponding slope of a trapezoidal fuzzy set $T[a, b, c, d]$ is

$$s_l = \frac{1}{b - a}$$

for the left boundary and

$$s_r = \frac{1}{c - d}$$

for the right boundary. An interesting property for the sake of our study is that the 0.5-cut of a trapezoidal fuzzy set is unaffected by the left and right slopes *if the cut-points are fixed*. In fact, according to the definition of trapezoidal fuzzy set (2),

$$A_i(t_i) = 0.5 \rightarrow t_i = \frac{c_i + d_i}{2} = \frac{a_{i+1} + b_{i+1}}{2}$$

therefore, by replacing c_i with $c'_i = c_i - k$ and d_i with $d'_i = d_i + k$, for any k that preserves (2), and changing a_{i+1} and b_{i+1} accordingly, the position of the *cut-points* t_i does not change, therefore the 0.5-cut of A_i as in (4) is unaffected.

In the following, we are going to dwell on this concept to analyze the impact of the boundaries on the inference mechanism of an FRBC. As a note of caution, in this work we assume that the cut-points are kept fixed because we are not interested in the ability of an FRBC in adapting to data, but rather on how the fuzziness of the trapezoidal fuzzy sets affects the classification function.

2.1 Classification via fuzzy rules

Let X_1, X_2, \dots, X_m be a collection of Universes of Discourse, each defined as

$$X_j = [l_j, u_j] \subset \mathbb{R}$$

for $j = 1, 2, \dots, m$. For each X_j , an SFP $A_{1,j}, \dots, A_{n_j+1,j}$ of trapezoidal fuzzy sets on X_j is considered. Also, let C be a finite set of class labels.

A rule R is identified by the pair

$$R = (\mathbf{A}, c)$$

where the antecedent is a fuzzy set \mathbf{A} defined over the Cartesian product \mathbf{X} of the aforementioned Universes of Discourse

$$\mathbf{X} = X_1 \times \dots \times X_m,$$

with membership function

$$\mathbf{A}(\mathbf{x}) = \min\{A_{i_1,1}(x_1), \dots, A_{i_m,m}(x_m)\},$$

being $\mathbf{x} = (x_1, \dots, x_m) \in \mathbf{X}$, and $c \in C$ is the consequent of the rule.

An FRBC is defined by a collection

$$S = \{R_1, R_2, \dots, R_r\}$$

of rules $R_k = (\mathbf{A}_k, c_k)$, with the constraint that any couple of rules cannot share the same antecedent, i.e. $\mathbf{A}_{k'} \neq \mathbf{A}_{k''}$ for any $k' \neq k''$. It should be noted that the collection S is determined on the basis of the aforementioned SFP and it does not necessarily coincide with a grid partition (in this sense, the grid partition represents just a special case of the described arrangement). In general, data-driven methods generate a small set of rules based on available training data to avoid combinatorial rule explosion, therefore we can expect to count in S a fewer number of rules with respect to those related to the full combination of possibilities coming from a grid partition. As a consequence of our assumptions, the Ruffini condition is imposed while partitioning each single dimension, but it does not represent a constraint to be verified on the rule antecedents.

The FRBC S is supposed to be applicable to a domain \mathbf{D} such that

$$\mathbf{D} \subseteq \bigcup_k \text{supp } \mathbf{A}_k \tag{5}$$

where $\text{supp } \mathbf{A}_k = \{\mathbf{x} : \mathbf{A}_k(\mathbf{x}) > 0\}$. In this way, we avoid the undesirable case of inputs for which no rules can be applied. The design process of an FRBC should ensure that no data fall outside the support of all rules.³

Given an input $\mathbf{x} \in \mathbf{D}$, the inference function of the FRBC S is carried out as

$$f_S(\mathbf{x}) = c$$

such that $c = c_{k^*}$ and

$$k^* = \arg \max_k \mathbf{A}_k(\mathbf{x})$$

i.e., the class returned by the FRBC is the one related to the rule with highest membership degree for the given input (ties are solved arbitrarily.) This inference rule is also called “Winner-Takes-All” (Angelov and Xiaowei 2008).

³ It may be the case that some classifiers designed from data do not have rules covering some newly-observed data samples: In this situation, the classifier either refuses to classify or it carries out a random classification.

Given a rule $R = (\mathbf{A}, c)$, we define the *region of dominance* of R as

$$[\mathbf{A}]_{0.5}^+ = \{\mathbf{x} : \mathbf{A}(\mathbf{x}) > 0.5\}$$

It is important to notice that the region of dominance of a rule is completely characterized by the cut-points of the underlying SFPs; in fact,

$$[\mathbf{A}]_{0.5}^+ =]t_{i_1-1,1}, t_{i_1,1}[\times \cdots \times]t_{i_m-1,m}, t_{i_m,m}[\tag{6}$$

where $t_{i_j,j} = t_{i_j}$, being t_{i_j} the i_j -th *cut-points* of the SFP defined on X_j for all $i_j > 0$. The validity of (6) can be easily checked by observing that $\mathbf{A}(\mathbf{x}) > 0.5$ if and only if, for each $j = 1, 2, \dots, m$, $A_{i_j,j}(x_j) > 0.5$. This can be achieved when x_j belongs to the 0.5-cut of $A_{i_j,j}$ with the exclusion of the boundary points.

Thanks to the concept of region of dominance, it is possible to establish a useful result concerning the classification function of a FRBC:

Lemma 1 *Let $S = \{R_1, R_2, \dots, R_r\}$ be an FRBC and let \mathbf{D} be an input domain. For any $\mathbf{x} \in \mathbf{D}$, if $\exists k : \mathbf{x} \in [\mathbf{A}_k]_{0.5}^+$, then*

$$f_S(\mathbf{x}) = c_k$$

Proof Since $\mathbf{x} \in [\mathbf{A}_k]_{0.5}^+ = \{\mathbf{x} : \mathbf{A}_k(\mathbf{x}) > 0.5\}$, however chosen a dimension j in $\{1, 2, \dots, m\}$ the following must hold:

$$A_{i_j,j,k}(x_j) > 0.5$$

where $i_j = 1, 2, \dots, n_j$ is the index of the fuzzy set in the j -th dimension, and $k = 1, 2, \dots, r$ is the index of the rule; thus $\mathbf{A}_k = A_{i_1,1,k} \times \cdots \times A_{i_m,m,k}$.

By definition of FRBC, any other rule in S (other than R_k) is characterized by an antecedent which is different from \mathbf{A}_k . Let $R_{k'}$ be the rule in S such that $A_{i_j,j,k} \neq A_{i_j,j,k'}$. The definition of SFP implies that

$$A_{i_j,j,k'}(x_j) < 0.5$$

Therefore, $\mathbf{A}_{k'}(\mathbf{x}) < 0.5$.

As a consequence, the membership degree of \mathbf{x} to R_k is highest among all the rules and therefore $f_S(\mathbf{x}) = c_k$. □

Informally speaking, the previous lemma states that regions of dominance establish subsets of the input domain where only one rule dictates the class label. It is therefore possible to define a subset of the input domain, namely

$$\mathbf{B} = \mathbf{D} \cap \bigcup_k [\mathbf{A}_k]_{0.5}^+$$

where the classification function is determined by one rule only for each input. What is more important for the purpose of our study is the following corollary:

Corollary 1 *The set \mathbf{B} is unaffected by the modifications applied on the slopes of the underlying fuzzy sets, provided that the corresponding cut-points are fixed.*

The corollary follows by observing that \mathbf{B} is included in the union of the regions of dominance of all the rules of an FRBC and, since each region of dominance is defined by the 0.5-cuts of the underlying fuzzy sets, which are not affected by the slopes of the fuzzy sets, then \mathbf{B} is also unaffected by such slopes.

Based on these results, it is possible to affirm that an FRBC S behaves like a crisp classifier in the region \mathbf{B} :

Lemma 2 *Given an FRBC S and a crisp classifier defined as follows:*

$$f'(\mathbf{x}) = \begin{cases} c_k & \text{if } \exists k \text{ s.t. } \forall j : x_j \in]t_{i_j-1,j,k}, t_{i_j,j,k}[\\ \text{undefined} & \text{otherwise} \end{cases}$$

then, $\forall \mathbf{x} \in \mathbf{B} : f_S(\mathbf{x}) = f'(\mathbf{x})$.

Notice that f' does not depend on any of the parameters that define the trapezoidal fuzzy sets underlying the FRBC S , but only on the set of cut-points. Therefore, within \mathbf{B} the classification function of an FRBC is completely unaffected by the slopes of all the trapezoidal fuzzy sets. In other words, it does not benefit from the involved fuzziness.

Outside \mathbf{B} , however, the fuzziness of the fuzzy sets plays a role in determining the confidence of the decision carried out by the FRBC. For each $\mathbf{x} \in \mathbf{U} = \mathbf{D} \setminus \mathbf{B}$, by definition we have

$$\forall k : \mathbf{A}_k(\mathbf{x}) \leq 0.5$$

If $\mathbf{A}_k(\mathbf{x}) = 0.5$, then there may exist another rule $R_{k'}, k \neq k'$, such that $\mathbf{A}_{k'}(\mathbf{x}) = 0.5$. This is verified if rules R_k and $R_{k'}$ share the same fuzzy sets in the antecedent with the exception of one dimension only, say j' , where the fuzzy sets of the two rules intersect. In such a case, if $c_k \neq c_{k'}$ the classification ambiguity can be solved by an arbitrary choice (e.g., random).

If $\mathbf{A}_k(\mathbf{x}) < 0.5$ for all $k = 1, 2, \dots, r$, the classification function can be better analyzed from the viewpoint of Possibility Theory (Dubois and Prade 2015). In fact, the inference schema of an FRBC is compatible with a possibilistic interpretation of the embodied fuzzy rules. Namely, each rule defines the possibility distribution that an object class is c_k provided that the observed features belong to \mathbf{A}_k . We write

$$\pi_k = \mathbf{A}_k(\mathbf{x})$$

to denote the possibility degree that the true class is c_k given the input \mathbf{x} according to the k -th rule. Rules with the same consequent class merge into a single possibility distribution defined by the union of all the antecedent. Formally, for all $c \in C$:

$$\pi_c = \bigcup_{k:c_k=c} \mathbf{A}_k(\mathbf{x}) = \max_{k:c_k=c} \mathbf{A}_k(\mathbf{x})$$

When an input is given, the possibility degree is computed for all class labels, and the class label with the highest possibility degree is chosen. This operation can be justified by introducing the measure of necessity (or certainty): informally speaking, the certainty about a class label is evaluated in terms of impossibility of the other class labels. Formally:

$$\nu_c = 1 - \max\{\pi_{c'} : c' \neq c\}$$

Thus, by selecting the class with the highest possibility degree, it is ensured that the certainty degree is also highest.

If $\mathbf{x} \in \mathbf{B}$ it is easy to verify that the certainty degree of the selected class is higher than 0.5 (it is equal to 1 if \mathbf{x} belongs to the core of the antecedent of a rule; this is a consequence of using SFPs). However, if $\mathbf{x} \in \mathbf{U}$ the analysis deserves some notes of caution. By construction, for any $c \in C$ the possibility degree is $\pi_c \leq 0.5$, therefore $\nu_c \geq 0.5$. In fact, ν_c is evaluated as the 1-complement of a quantity that is smaller than π_c (therefore, smaller than 0.5). In other words, the certainty degree of any class label is higher than its possibility: this is an anomalous result since, in normal situations, certainty is never greater than possibility.⁴

Furthermore, for a given class c it is possible to compute the certainty degree that another class is the true one. This can be simply reckoned by computing the impossibility that c is the true class, i.e. $\nu_{C \setminus \{c\}} = 1 - \pi_c$. In the case that $\mathbf{x} \in \mathbf{U}$, whatever class label c is selected, we obtain $\nu_{C \setminus \{c\}} \geq 0.5$, that is, for any possible class label emitted by the classifier, it is more certain that another class is a true one. Again, this is a situation that should be avoided in classification. It must be observed that this case does not happen if $\mathbf{x} \in \mathbf{B}$.

It is important to notice that the slopes of the trapezoidal fuzzy sets affect the volume of the set \mathbf{U} . Ideally, this volume should be as small as possible, which can be achieved by crisp rules. On the other hand, by using triangular fuzzy sets as a special case of trapezoidal fuzzy sets, the volume of \mathbf{U} is maximized.

3 Numerical results

We tested the impact of the theoretical results shown in the previous section on some synthetic datasets. A granulation method was applied to generate SFPs for each dimension related to the data at hand. Namely, for each dataset we used DC* to generate the *cut-points* and the initial SFP for each dimension (Castiello et al. 2019). DC* is a specific algorithm designed to perform a double clustering process devoted to extract interpretable fuzzy granules of information from data and to express them in form of fuzzy classification rules. A first clustering of data is performed using a prototype-guided algorithm; then the derived prototypes are projected on each dimension and those projections are further clustered by exploiting the capabilities of the A* search algorithm.

We applied DC* to the bi-dimensional synthetic datasets depicted in Figs. 1, 2 which illustrate the cut-point configurations produced by DC* together with the data points. Table 1 sums up the main characteristics of the datasets. As can be observed, the datasets differ in the number of classes and datapoints. Also, the application of DC* produced *cut-points* that in some cases are in agreement with the data distribution, while in some other cases they appear to be less appropriate for discriminating among classes.

Once cut-points have been generated from data, SFPs have been designed in terms of trapezoidal fuzzy sets that are constrained to intersect in correspondence of *cut-points*. This has been accomplished in different ways: three heuristic methods called “Constant Slope” (CS), “Variable Fuzziness” (VF) and “Core Points” (CP) (Mencar et al. 2013) and two data-driven techniques based on Particle Swarm Optimization (PSO) (Castiello and Mencar 2019). The two data-driven techniques, called “Leftmost Slope Constraint” (LSC) and “Constant Slope Constraint” (CSC) aims at optimizing the slopes of trapezoidal fuzzy sets in order to achieve the highest classification accuracy on the dataset.

In Table 2 we show the classification accuracy achieved for each dataset and for each method used for generating the SFPs. We observe a high stability of classification accuracy for any given dataset. The most relevant changes can be observed for datasets SD6 and SD7: the corresponding plots in Fig. 2 show the presence of granules (i.e. boxes bounded by *cut-points*, which correspond to rules if there are enough data) where data pertaining to different classes are mixed. Such cases are related to some DC* results which turned out to be less appropriate in terms of class discrimination; however, in those regions the classification function produces varied outputs while the slopes of the trapezoidal fuzzy sets are modified.

In Fig. 3 we provide a comparison of the SFPs obtained by applying the different methods put in action during the experimental session. The SD2 dataset has been chosen as an example and only one dimension has been considered for

⁴ Intuition helps: what is possible could not be certain, but what is certain must be possible.

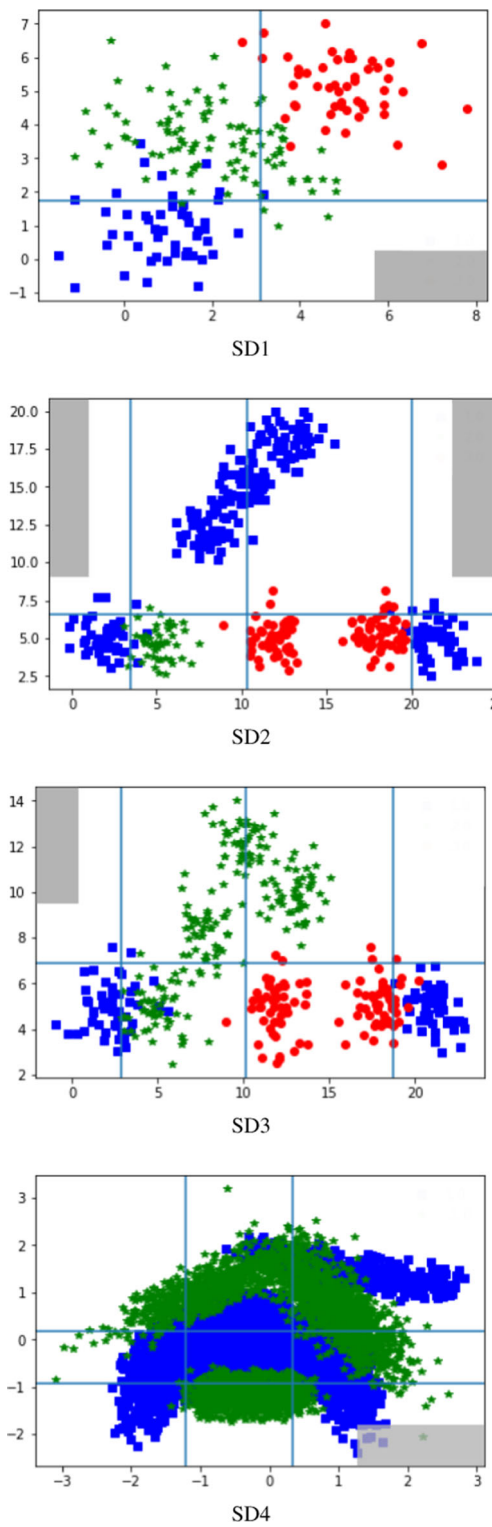


Fig. 1 The datasets SD1-SD4 adopted for the numerical simulation. The shadowed areas correspond to regions outside the support of all rule antecedents. Data points falling in these areas are classified randomly by DC*. Regions delimited by cut-points and without shadowed areas correspond to the regions of dominance of some rules.

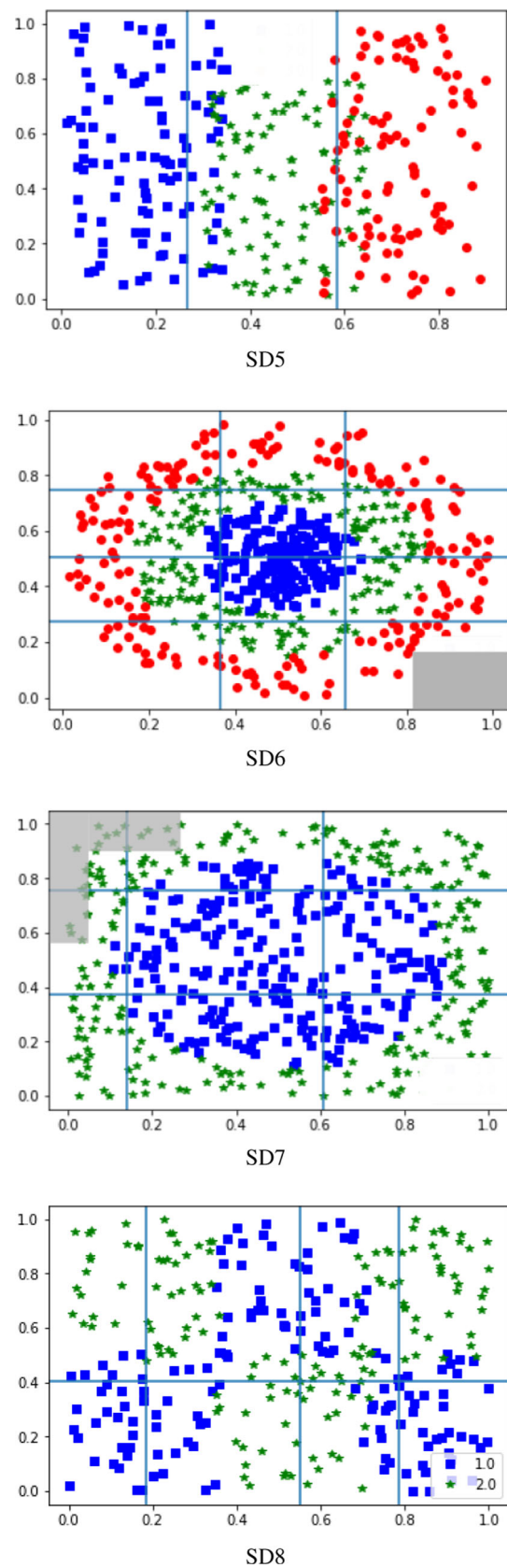


Fig. 2 The datasets SD5-SD8 adopted for the numerical simulation. See also Fig. 1

Table 1 Description of the datasets involved in the experimental session

Dataset	# Input	# Classes	# Samples
SD1	2	3	200
SD2	2	3	400
SD3	2	3	400
SD4	2	2	5300
SD5	2	3	300
SD6	2	3	600
SD7	2	2	500
SD8	2	2	300

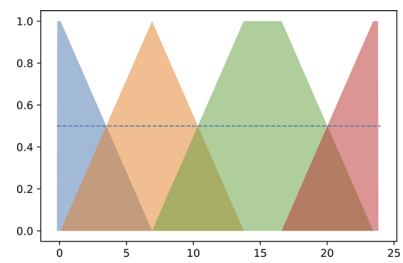
Table 2 Accuracy (%) of fuzzy classifiers embedding the SFPs obtained through the application of different strategies

Dataset	Heuristic methods			PSO methods	
	CS	VF	CP	LSC	CSC
SD1	82.00	82.00	82.00	82.00	82.00
SD2	96.25	96.25	96.25	96.25	96.25
SD3	92.50	92.50	92.50	92.50	92.75
SD4	67.30	67.30	66.32	67.32	67.30
SD5	83.67	83.67	83.67	83.67	83.67
SD6	67.83	67.83	65.83	68.83	68.17
SD7	68.00	68.60	62.40	68.60	68.40
SD8	66.67	66.67	66.67	66.67	66.67

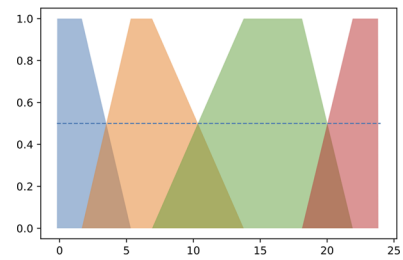
the sake of illustration. It can be observed how the classification results are highly stable through the application of the different methods, in spite of the differences achieved while designing the trapezoidal fuzzy sets involved in the SFPs.

4 Conclusions

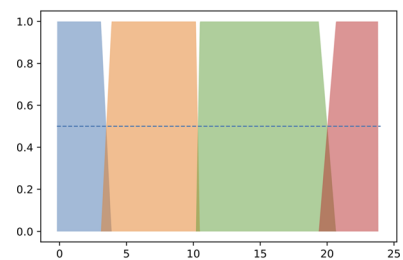
We considered the classification carried out by a FRBC where fuzzy sets in antecedent are aggregated through the min operator and inference is determined by the Winner-Takes-All rule. The theoretical results—supported by the numerical experiments—show that the fuzziness of the linguistic terms involved in an FRBC, as quantified by the slope of the corresponding trapezoidal fuzzy sets, does not affect the classification function in the region where the classifier is more confident (that is, where the degree of certainty of the returned class is greater than 0.5). On the other hand, fuzziness affects the behaviour of an FRBC in a region of the input space where classification is problematic from the possibilistic point of view. However, if an FRBC learning algorithm is capable to capture the hidden relations among data, then most of them will fall in the regions of dominance of some rules, thus reducing the effects of classification outside such safe regions.



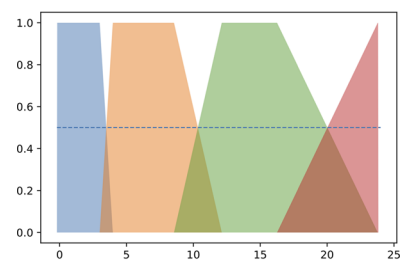
CS



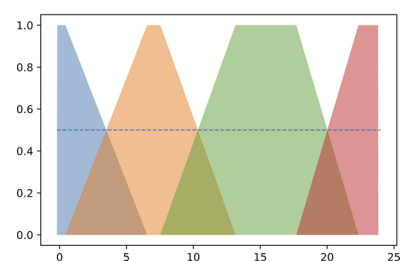
VF



CP



LSC



CSC

Fig. 3 Comparison of computed SFPs for dataset SD2 (first dimension)

All in all, the performance of an FRBC is predominantly determined by the *position* of the fuzzy sets in their domain, which is well captured by the collection of cut-points: when the cut-point positions are modified, the decision boundary of the classifier changes accordingly, thus affecting performance. On the other hand, by changing the *fuzziness* of the membership functions, the impact on the classifier is marginal, provided that SFPs are adopted and a class is selected by choosing the class label of the rule showing the highest membership degree. As an extreme case, which may correspond to the adoption of a grid partition strategy to split the input space, a fuzzy rule-based classifier may act *exactly* as a crisp classifier, thus implying that fuzziness does not play *any role at all* in the classification inference.

What is therefore the role of fuzzy sets in an FRBC? In a Machine Learning perspective, fuzzy sets are useful to fine-tune the decision boundaries in the presence of samples far from the clusters characterizing the regions of dominance of some rule. However, such a result appears to be marginal, since the performance of an FRBC can be improved by injecting more flexibility. For example, SFPs may be put aside, but some care must be taken to preserve interpretability. Moreover, fuzziness can play a relevant role by using different inference schemes, e.g. by allowing the inference of sets of classes, possibly associated with some confidence information.

In the context of XAI, however, the quality of the decision returned by an intelligent system is of utmost importance. In this sense, the fuzziness embodied in the FRBC gives valuable information about the confidence of classification. In particular, the classification function of an FRBC can be enriched by adding to the predicted class label a measure of confidence, i.e. the possibility and certainty degrees expressing the truthfulness of the inferred prediction according to the embodied knowledge base. (Eventually, this additional information can be rendered in legible form through some NLG process.) Finally, the membership degrees, rather than being arbitrarily determined, can be semantically grounded on some data properties (e.g., membership can be defined in terms of similarity with respect to a prototypical sample or interval). In such cases, it is possible to provide a faithful explanation of the reasons behind the decision carried out by an FRBC.

In conclusion, fuzziness may have a reduced role in the inference mechanism of an FRBC, while being relevant in terms of explanation of the produced results. Hence, we believe that these results convey an important message to the designers of fuzzy rule-based classifiers, since it represents a hint concerning the real utility of fuzziness in fuzzy modeling.

Funding Open access funding provided by Università degli Studi di Bari Aldo Moro within the CRUI-CARE Agreement. This work has

been partially supported by Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) under Grant PON ARS01_00141 "CLOSE".

Data availability The datasets and the code generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alonso J, Conde-Clemente P, Trivino G (2017) Linguistic description of complex phenomena With The rLDCP R package. In: Proceedings of the 10th international conference on natural language generation, pp 243–244
- Alonso JM, Magdalena L, Guillaume S (2008) HILK: a new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism. *Int J Intell Syst* 23(7):761–794. <https://doi.org/10.1002/int.20288>
- Alonso Moral JM, Castiello C, Magdalena L, Mencar C (2021) Interpretability constraints and criteria for fuzzy systems. Explainable fuzzy systems: paving the way from interpretable fuzzy systems to explainable AI systems. Springer, Cham, pp 49–89
- Angelov P, Xiaowei Z (2008) Evolving fuzzy-rule-based classifiers from data streams. *IEEE Trans Fuzzy Syst* 16(6):1462–1475. <https://doi.org/10.1109/TFUZZ.2008.925904>
- Biran O, Cotton C (2017) Explanation and justification in machine learning: a survey. *IJCAI Workshop Explain Artif Intell* 8(1):8–13. <https://doi.org/10.1108/13563281111156853>
- Castiello C, Mencar C (2019) Exploiting particle swarm optimization to attune strong fuzzy partitions based on cuts. In: Proceedings of the 11th conference of the european society for fuzzy logic and technology (EUSFLAT 2019), Atlantis Press, pp 430–437, <https://doi.org/10.2991/eusflat-19.2019.60>
- Castiello C, Fanelli AM, Lucarelli M, Mencar C (2019) Interpretable fuzzy partitioning of classified data with variable granularity. *Appl Soft Comput* 74:567–582. <https://doi.org/10.1016/j.asoc.2018.10.040>
- Dubois D, Prade H (2015) Possibility theory and its applications: where do we stand? In: Springer handbook of computational intelligence, Springer, Berlin, pp 31–60, https://doi.org/10.1007/978-3-662-43505-2_3

- Dubois D, Grabisch M, Prade H (1995) Gradual rules and the approximation of control laws. Theoretical aspects of fuzzy control. Wiley, New York, pp 147–181
- Gorzalczany MB, Rudziński F (2017) Interpretable and accurate medical data classification: a multi-objective genetic-fuzzy optimization approach. *Expert Syst Appl* 71:26–39. <https://doi.org/10.1016/j.eswa.2016.11.017>
- Guidotti R, Monreale A, Turini F, Pedreschi D, Giannotti F, Ruggieri S, Turini F, Giannotti F, Pedreschi D (2018) A survey of methods for explaining black box models. *ACM Comput Surv* 51(5):1–42. <https://doi.org/10.1145/3236009> [arXiv:1802.01933](https://arxiv.org/abs/1802.01933)
- Hagras H (2018) Toward human-understandable. Explainable AI. *Computer* 51(9):28–36. <https://doi.org/10.1109/MC.2018.3620965>
- Loquin K, Strauss O (2006) Fuzzy histograms and density estimation. In: *Soft methods for integrated uncertainty modelling*, Springer, pp 45–52
- Lu Y (2019) Artificial intelligence: a survey on evolution, models, applications and future trends. *J Manag Anal* 6(1):1–29. <https://doi.org/10.1080/23270012.2019.1570365>
- Mencar C, Lucarelli M, Castiello C, Fanelli AM (2013) Design of strong fuzzy partitions from cuts. In: *Proceedings of the 8th conference of the european society for fuzzy logic and technology*, Atlantis Press, Paris, France, *Advances in Intelligent Systems Research*, vol 32, pp 424–431, <https://doi.org/10.2991/eusflat.2013.65>, <http://www.atlantis-press.com/php/paper-details.php?id=8427>
- Perfileva I (2006) Fuzzy transforms: theory and applications. *Fuzzy Sets Syst* 157(8):993–1023
- Ruspini EH (1969) A new approach to clustering. *Inf Control* 15(1):22–32
- Xu LD, Xu EL, Li L (2018) Industry 4.0: state of the art and future trends. *Int J Prod Res* 56(8):2941–2962. <https://doi.org/10.1080/00207543.2018.1444806>
- Zadeh LA (1999) From computing with numbers to computing with words. From manipulation of measurements to manipulation of perceptions. *IEEE Trans Circuits Syst I Fundam Theory Appl* 46(1):105–119. <https://doi.org/10.1109/81.739259>
- Zadeh LA (2008) Toward human level machine intelligence: is it achievable? the need for a paradigm shift. *IEEE Comput Intell Mag* 3(3):11–22. <https://doi.org/10.1109/MCI.2008.926583>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.