LFoC-Interpretability of Linguistic Rule Based Systems and its Applications To Solve Regression Problems

Cat Ho Nguyen  
Institute of Information technology, VAST  
Hoang Quoc Viet str. 18, Cau Giay,  
Ha Noi, Vietnam  
Email: ncatpho@gmail.com

Thai Son Tran  
Institute of Information technology, VAST  
Hoang Quoc Viet str. 18, Cau Giay,  
Ha Noi, Vietnam  
Email: trn_thaison@yahoo.com

Van Thong Hoang  
Information Technology Faculty,  
University of Transport and Communications at Ha Noi, Viet Nam  
Email: thonghoangvan@yahoo.com

Van Long Nguyen  
Information Technology Faculty,  
University of Transport and Communications at Ha Noi, Viet Nam.  
Email: nvlongdt@yahoo.com.vn

Abstract

The low-interpretability of fuzzy rule based systems (FRBSs) is required to based on the inherent semantics a word-set of a variable, written shortly as LFoC. LFoCs are fundamental semantic basis and the interpretability of their fuzzy representation is examined based on how an interpretation-assignment of fuzzy sets to their words can convey discovered key properties of LFoCs to their fuzzy set representation. The study proposes two additional constraints imposed on the desired interpretation to maintain discovered features of LFoCs to their fuzzy set representation. Thus, FRBSs that are developed in a formalism based on the fuzzy set theory to simulate the human capabilities by handling fuzzy sets assigned to words instead of the words themselves. By this, the interpretability of FRBSs is a crucial problem to analyse the relation between FRBSs and human users who describe the FKBs in terms of their own linguistic words and develop the ARMs based on the linguistic experience domain knowledge in their head. In the existing approaches to the FRBS interpretability, there is no a formal basis to connect the formalism based on which FRBSs are able to handle their elementary computational objects as well as their above mentioned components to the “formalism” based on which human experts handle their own words, domain knowledges and their reasoning methods acting immediately on their words. Such two formalisms are completely different and the lack such a formal connection seems to be the reason resulting in too many approaches to examine this topic existing in the literature since 90’s of the last century, e.g. the studies by Ishibuchi et al. [23-26], and the interpretability has been intensively studied since then and now becomes an essential characteristic of FRBSs, especially for user-centric applications.

Keywords: fuzzy rule based systems, order-based semantics of words, hedge algebras, low-interpretability, regression problems

1. Introduction

Inspired by human capabilities in using their knowledge expressed in terms of sentences of their natural languages to do reasoning and making decisions in the reality, fuzzy rule based systems (FRBSs) are proposed to comprise mainly two key components, namely a fuzzy knowledge base (FKB) and an approximate reasoning method (ARM) acting on fuzzy sets. Therefore, FRBSs that are developed in a formalism based on the fuzzy set theory to simulate the human capabilities by handling fuzzy sets assigned to words instead of the words themselves. By this, the interpretability of FRBSs is a crucial problem to analyse the relation between FRBSs and human users who describe the FKBs in terms of their own linguistic words and develop the ARMs based on the linguistic experience domain knowledge in their head. In the existing approaches to the FRBS interpretability, there is no a formal basis to connect the formalism based on which FRBSs are able to handle their elementary computational objects as well as their above mentioned components to the “formalism” based on which human experts handle their own words, domain knowledges and their reasoning methods acting immediately on their words. Such two formalisms are completely different and the lack such a formal connection seems to be the reason resulting in too many approaches to examine this topic existing in the literature since 90’s of the last century, e.g. the studies by Ishibuchi et al. [23-26], and the interpretability has been intensively studied since then and now becomes an essential characteristic of FRBSs, especially for user-centric applications.
In addition, since linguistic words of variables play as bridge to link real semantics conveyed by human words to their fuzzy set based semantics, the lack such a formal connection leads to a situation that the existing approaches cannot link the real meaning conveyed by human words with the fuzzy sets designed for the words appearing in FRBSs. For instance, various studies related to the low-interpretability in the sense of Zhou and Gan [50] mainly focus to how one can design the fuzzy set shapes assigned to certain words of a variable and how reasonably arrange them in fuzzy partition to meet human user intuition about the word semantics without any formal linkage with the inherent semantics of words. In this tendency, they utilize the following techniques to construct fuzzy sets and fuzzy partitions:

1. **Tuning fuzzy set parameters** is applied by Pedrycz and Oliveira [45], Cordón and Herrera [14], Oliveira [42], Babuska [9], Botzewhim et al. [11,12], Pulkkinen et al. [46], Antonelli et al. [6-8] and Gacto et al. [17].

2. **Proposing similarity measure and/or metrics** between fuzzy sets is studied in Setnes et al. [75], Guillaume and Charnomordic [21,22 ], and Antonelli et al. [8].

3. **Maintaining a reasonable distinguishability between fuzzy sets** is examined in Oliveira [39,40], Jin [28,29], Mencar et al. [31], Pulkkinen and Koivisto [46]; or using the core of fuzzy sets to measure the degree of their distinguishability is examined, e.g. in Espinosa and Vandewalle [17], or, more generally, using similarity measure which expresses also the degree of their distinguishability is examined, e.g. in Espinosa and Vandewalle [17].

4. **Constraints imposed on fuzzy sets or fuzzy partitions** to maintain their semantics are proposed and examined in many studies, which are scattered in time from 1990’s up to now. An interesting survey on such constraints is made by Mencar and Fanelli [32], where one can find 35 constraints imposed on FRBSs among which there are 16 constraints related to the low level interpretability, i.e. to fuzzy sets and fuzzy partitions, and the remaining related to the rule level and the fuzzy systems themselves.

5. **Various distinct interpretability indexes**, from low level interpretability to high level, defined by simple or more complicated definitions using mainly numeric indices or also sometimes fuzzy indices, are discussed and examined in almost researches related to designing genetic FRBSs. A taxonomy of such indices can be found in the study by Alonso et al. [5] that looks for a good interpretability index, a fuzzy index. In last ten year this problem is still more intensively investigated, see for example the studies by Mencar [31], Ishibuchi and Nojima [27], Alonso et al. [3,4], Fazendeiro and de Oliveira [18], Botta et al. [10], Gacto et al. [19].

6. An interesting interpretability index is proposed and examined by Antonelli et al. [6-8] in which they introduce a similarity measure between two fuzzy partitions defined based on the whole triangle fuzzy partitions, the one is a current partition and the other is a uniformed one being regarded as a most interpretable partition, to maintain the semantics of fuzzy partitions is examined.

These continuous studies of a so complex concept lead, of course, to several overviews on the interpretability of FRBSs, including those by Mencar and Fanelli [32], Zhou and Gan [50], Gacto et al. [20] and Cordón [16]. The survey [16] analyzes that in the 90s of the last century most researchers pay mainly attention to the accuracy of the FRBSs, but less attention to their interpretability. However, since then the FRBS interpretability more and more attracts the researcher interest which leads to studies trying to solve the accuracy-interpretability tradeoff problem in designing FRBSs. The survey by Mencar and Fanelli [32] summarizes studies of the interpretability of FRBSs and exposes 35 constraints imposed on various different components of FRBSs, e.g. constraints on their fuzzy sets, which are their elementary elements, or on the fuzzy partitions of the variable universes, whose fuzzy sets of each variable form the so-called Frame of Cognitions (FoCs); or on the granularities representing FoCs, and so on.

Hedge algebras are math-models of the word-domains of variables which are order-based structures generated from the inherent semantics of the words. Since human beings are able to recognize reality and make their proper decisions in their daily life, their words in variable domains and, hence, the words of their respective hedge algebras, can translate the real world semantics assigned by human beings to the words into the designed computational objects. This is a practical and methodological basis to establish another approach to the interpretability of FRBSs introduced in [41] based on application of the interpretation concept of formalized languages and formalized theories. Restricted to the interpretability of computational representations of LFoCs which are word-sets humans designers used to develop their FRBSs, in this approach their interpretability is guaranteed by discovering structural properties of LFoCs and considering them as constraints imposed on the interpretation assignment of the words in LFoCs to their desired computational objects. To answer the question whether this approach can enhance the performance of the designed FRBSs, this approach is applied to genetically design LFoC-interpretable FRBSs using tringle-semantics of words to solve
regression problems. The proposed method in that paper to design LFoC-interpretable FRBSs is shown to be statistically outperform the counterpart methods mentioned there.

The above approach opens a novel way to examine the interpretability of FRBSs and to design LFoC-interpretable FRBSs and, hence, it is required many research efforts to demonstrate its advantages in the design of FRBSs, especially, in the context of uncertainty environment. Therefore, in this study we give an additional discussion on this approach and contribute to show its effectiveness by designing LFoC-interpretable FRBSs using trapezoid semantics of words to solve also regression problems.

Fuzzy set shapes play also a significant role in representing the semantics of words. Trapezoids applied instead of triangles to construct fuzzy systems are examined in a number of researches, e.g. Botzheim et al. [11,12], Paetz [44], Wong et al. [49] and Mencar et al. [33]. However, as discussed previously, they are constructed and assigned to the words under consideration without a formal relation with the inherent semantics of words. Nguyen et al. [40] develop an enlargement of a given ordinary hedge algebras (HAs), called enlarged HAs (EnHAs), to establish a math-basis to produce trapezoid-semantics of words from the words of variables, considering the semantics of any individual word in the context of its entire variable, to follows the principle that “words determine their computational semantics”.

The rest of the paper is organized as follows. Section 2 gives necessary knowledge of semantics core of words and the enlarged hedge algebras considered as a formalized basis to handle the semantics core of words and large regions of fuzzy sets. Section 3 is devoted to study the (semantics-based) low-interpretability problem based on the new concept of the interpretability, some specific features of the word semantics and two related interpretation constraints on computational/fuzzy set representations of LFoCs and the interpretability of trapezoid MGr-representation of LFoCs. Section 4 describes a proposed method to design LRBSs interpretable in the new sense with the trapezoid fuzzy set semantics of words to solve regression problems. There, it is applied to develop main algorithm and two modified versions to produce LRBSs using newly interpretable triangle/trapezoid MGr-representation of LFoCs to solve regression problems. Simulation studies of the proposed algorithms are executed in Section 5 to expose the meaning of interpretable trapezoid MGr-representations of LFoCs. The study is closed with some conclusions and discussion given in Section 6.

2. Enlarged hedge algebras – a formalized foundation for modeling the semantics core of words and for producing the core regions of the fuzzy set based word semantics

To establish formalized formalism in the HA-approach to produce trapezoid fuzzy sets from LFoCs, the study [40] develops so-called enlarged hedge algebras (EnHAs), in which the semantics cores of words can be modelled, and establish a formalism, in which interval-semantics of the semantics cores of the words represent just the cores of their trapezoid fuzzy sets to be constructed. In this section, we summarize some necessary facts of the EnHAs convenient for reference in the sequel.

The concept of semantics core of words is proposed and examined inspired by the observation that when representing the semantics of a word \( x \) of a variable \( X \) with the universe \( U \) by a fuzzy set on \( U \), denoted for short by \( FS(x) \), the core region of \( FS(x) \), denoted by \( \text{Core}(FS(x)) = \{ (u, \mu_{FS}(u)) : u \in U & \mu_{FS}(u) = 1 \} \), should represent a certain particular semantic aspect of \( x \), called the core of its semantics or, shortly, its semantics core and denoted by \( \text{Core}(x) \). Similarly, the set \( \text{Sem}(x) = \{ (u, \mu_{FS}(u)) : u \in U \} \) can be interpreted as a model of the inherent semantics of the word \( x \), denoted by also just \( x \), for simplicity. Intuitively, it is observed that, for any \( x, y \) of \( X \), the following inequalities hold:

\[
\text{C1) } \text{Core}(x) \subseteq \text{Sem}(x);
\]

\[
\text{C2) } x \leq y \Rightarrow \text{Core}(x) \leq \text{Core}(y) & x \leq \text{Core}(y) & \text{Core}(x) \leq y.
\]

It is interesting that the semantics core of words can be formalized in terms of the semantic order relation of the ordinary HAs. As usual, a linguistic variable \( X \) is simply the one whose values are natural words which can be expressed as strings of hedges and atomic words, e.g. “rather very young”. To model the order-based semantics of word-domains of variables, an ordinary HA of \( X \) is a structure \( \mathcal{A}X = (X, G, C, H, \leq) \), where \( X = \text{Dom}(X) \) is the underlying set of \( \mathcal{A}X \), \( G = \{ c, c' \} \) consists of two atomic words or word-generators; \( C = \{ 0, W, 1 \} \) is the set of three specific words, the least word, the “neutral” and the greatest one in \( X \), respectively; \( H \) is the set of the hedges of \( X \) and \( \leq \) is the semantic order relation on \( X \). So, every element of \( X \) is represented as a string \( h_n \ldots h_{c} \), where \( h_i \in H \) and \( c \in G \), and hence \( X = H(G) \cup C \), in which \( H(G) = H(c) \cup H(c') \), where, for every \( x, H(x) = \{ \sigma : c \in G, \sigma = h_n \ldots h_{a} \in H^a \} \) and, for \( \sigma = \varnothing, \sigma = x \).
Hence, in this structure, for any hedge that, for every natural language, one have to adopt a hypothesis $h$, for every word $x$, for every word $y$. Accordingly, Nguyen et al. (KBS-14) propose an enlargement of every given linear ordinary HA, $AX = (X, \<, \> \leq, \leq)$, to a so-called enhanced HA, denoted by $AX_{en} = (X_{en}, \<, \> \leq, \leq)$, where $\forall h \in H$ and $h \neq H, X_{en} = C \cup H(h) = C \cup \{h_0, \ldots, h_m : e \in G, h_i \in H\}$, where $i = 1$ to $m$. Hence, $X = C \cup H(G) \subseteq X_{en} = C \cup H(h)$.

(A5th) Axioms for the semantics core of terms: For $\forall \forall x, y \in X_{en}$ and $y \neq y$,
(i) $h'hx = hx$, for $\forall h' \in H_{en}$ i.e. $hx$ is always a fixed point.
(ii) For $\forall x \in X, hx = x$ if and only if $x$ is constant, otherwise, $hx$ and $x$ are incomparable.

In order to describe the order-based structure of the EnHAs for reference in the sequel, we recall here some facts examined in Nguyen et al. [40]. First, we recall some notations. For the set $X$ of $AX$, we have $X_0 = \{x \in X : |x| = k\}$ and $X_{k+} = \{x \in X : |x| \leq k\}$, while, for the set $X_{en}$ of its EnHA, $AX_{en}$, we put $X_{en,h} = X_0 \cup \{h : e \in G, h \in H\}$ and $X_{en,0} = \{x \in X_{en} : |x| \leq k\}$. Then, assume that $H = \{h : 1 \leq j \leq p\}$ is the set of the positive ones, and $H = \{h : -1 \leq j \leq -q\}$ is the set of the negative ones, the following properties hold in every EnHA, noting that $H_{en}(hx) = \{h_0x\}$:

(Pr1) $0 \leq H_{en}(c) \leq \{x \in \mathbb{R} : h<c, x\}$
(Pr2) The sets $H_{en}(hx), j \in [-q, p]$, noting that $H_{en}(x) = \bigcup_{j=-q}^{p} H_{en}(h_jx)$, are linearly ordered and we have

(i) $\text{sign}(hx) = +1$ (i.e. $hx \geq x$) $\Rightarrow H_{en}(hx) \leq \ldots \leq H_{en}(h_0x) \leq H_{en}(h_1x) \leq \ldots \leq H_{en}(h_px)$
(ii) $\text{sign}(hx) = -1$ (i.e. $hx \leq x$) $\Rightarrow H_{en}(hx) \leq \ldots \leq H_{en}(h_0x) \leq H_{en}(h_1x) \leq \ldots \leq H_{en}(h_px)$

(Pr3) Semantically, the core $hx$ is included in $x$, we have: $\forall x, y \in X_{en}, x < y \iff x < h_0y \iff h_0x < y \iff hx < h_0y$.

(Pr4) All semantics cores of the words, $hx, x \in X$ are linearly ordered.

(Pr5) The set $X_{en,h} = X_0 \cup \{h : u \in X_{en,1}\} \subseteq X_{en}$ is also linearly ordered.

The above properties contribute to describe the semantic structure of word-domains involving the fuzziness of the words but in terms of We shall present shortly how the words in this structure can be numerically quantified in the HA-approach that leads to very close concepts: interval-valued quantifying mappings, fuzziness measure and fuzziness intervals of words. To produce interval core regions of fuzzy sets from the word-domain of a variable $X$ modeled by the underlying set $X_{en}$ of the EnHA $AX_{en}$, we should deal with Interval-Valued semantically Quantifying Mappings (IVQMs) of the form $f : X_{en} \rightarrow \mathbb{I}(\{0,1\})$, where $\mathbb{I}(\{0,1\})$ is the set of all subintervals of $[0,1]$, which are left-closed and right-open, except when their left end-points are 1, they are also right closed. Naturally, it is required that the desired IVQMs must convey the word semantics of variables, i.e. they should maintain essential semantic order relationships in their EnHAs. Therefore, Nguyen et al. [40] propose the following definition of the IVQMs:

Definition 2.1 [40]. An $f : X_{en} \rightarrow \mathbb{I}(\{0,1\})$ is said to be an IVQM of $AX_{en}$ provided that the following conditions hold:

(IVQM1) $f$ preserves the order of the words in $X$, i.e. it is an isomorphism of $(X_{en}, \leq)$ in the category of the ordered sets. That is that ($\forall x, y \in X_{en}(x \leq y \Rightarrow f(x) \leq f(y))$).

(IVQM2) $\forall x \in X_{en} f(hx) \leq f(x)$, i.e. $f$ preserves the functionality of the hedge $h_0$ ($\forall x \in X_{en}: \{h_0x, x \in X\}$) and $f(X) = \{f(x) : x \in Y\}, \forall Y \subseteq X_{en}$.

These conditions are most general and aimed to ensure that $f$ may properly convey the order-based semantics of the words to the interval-semantic. We will see that the $f$-values may model actually the semantics cores of words.

Fuzziness measures of variables and fuzziness intervals of their linguistic words

In the HA-approach, IVQMs are utilized to define the concept of fuzziness measure of variables. Nguyen
C.H and Nguyen V.L. [37] argue that the set $H(x)$ and, in this study, the set $H_{al}(x)$ is a suitable model of the fuzziness of $x$, noticing in general that $H_{al}(x) = \{a\}$, for $\forall a \in C$, means that $a$‘s are non-fuzzy. Then, the size of the image of $H_{al}(x)$ under a given IVQM $f$, $(H_{al}(x))$, can be taken to be the fuzziness measure of the word $x$, denoted by $f_{m}(x)$, as $f$, by its definition, can maintain the order-based semantics of words. Since in the HA-approach, words in principle should determine their numeric semantics, fuzziness intervals and fuzziness measure of the words of $X$ can closely be defined as follows:

**F1) Fuzziness intervals of the words:** For every word $x \in X_{en}$ as the IVQM $f$ preserves the order relation and $f(X_{en})$ is dense in $[0, 1]$, the image $f(H_{al}(x))$ must be dense in a certain sub-interval of $[0, 1]$ and, hence, the closure of $f(H_{al}(x))$, denoted by $\text{Cl}(f(H_{al}(x)))$, defines an interval, called the fuzziness interval of $x$ and denoted by $\mathfrak{I}(x)$. Consequently, we have $\mathfrak{I}(h_{0}) = f(h_{0})$ in $\mathfrak{P}([0, 1])$ as $H_{al}(h_{0}) = \{h_{0}\}$, a singleton set.

**F2) Fuzziness measure of $X$:** The fuzziness measure $f_{m}(x)$ of $X$ is defined to be the length of $\mathfrak{I}(x)$, $f_{m}(x) = |\mathfrak{I}(x)|$.

So, the fuzziness models and the fuzziness intervals of the words are generated from the algebraic structure of the word-domain of $X$ and, hence, the fuzziness models of the words have their own structure as presented in Fig. 1 and 2, whose words are of the length not greater than 2.

Relying on the definition of IVQMs and the structure of $AX_{en}$, especially of the sets appearing in (Pr1), whose set-union is equal to $X_{en}$, and the structure of the sets appearing in either (i) or (ii) of (Pr2), whose set-union is equal to $H_{al}(x)$, Nguyen et al. [40] have proved that:

- The fuzziness intervals of the words $0, c^{-}, W, c^{+}$ and $1$ form a binary partition of $[0, 1]$ and we have $\mathfrak{I}(0) \cup \mathfrak{I}(c^{-}) \cup \mathfrak{I}(W) \cup \mathfrak{I}(c^{+}) \cup \mathfrak{I}(1) = [0, 1]$ (1)

- The order of the fuzziness intervals of the words $h_{x}, -q \leq j \leq p$, is in accordance with the order of their fuzziness measure $f_{m}$ of $X$. It is completely defined if the numeric values of $f_{m}(c^{-})$, for $c \in C$, $f_{m}(c^{+})$, and $\mu(h)$, for $h \in H$ (does not contain $h_{0}$), called the independent fuzziness parameters of $X$, are given. They are constrained by (fm1) and (fm4).

**Interval-valued quantifying mappings induced by fuzziness measures**

In turn, a fuzziness measure of $X$ given by its independent fuzziness parameters does uniquely determine its IVQM, which is called $f_{m}$-induced IVQM, denoted by $\nu_{f_{m}}$. Note that the interval-values of $\nu_{f_{m}}$ can exactly be determined when their left endpoints of the intervals, $L(f_{m}(h_{0}(x)))$, can be computed. For instance, let us compute $\nu_{f_{m}}(c^{+})$, where its EnHA $AX_{en}$ with $H_{en} = \{L, h_{0}, V\}$
\(\mathcal{X}_{\text{en}}\) is given by \(H_{\text{en}} = \{L, h_0, V\} \). Since \([h_0, L] = 3\), we must consider the structure of the fuzziness intervals of \(X_{\text{en}}(j)\), which is shown in Fig. 3. By (1), \(v_{fm}(x) = 3(h(x)\mathcal{L}_c)\) and \(Lt(f(3(h(x)\mathcal{L}_c)))\) can be calculated as follows:

\[
Lt(f(3(h(x)\mathcal{L}_c)) = \sum \mu(h) + \mu(h_0W) + \mu(VLC) + \mu(WVc) + \mu(Vc)
\]

In general case, it can be verified that:

\[
Lt(v_{fm}(x)) = \sum y \in X_{\mathcal{L}+} f(h(y)) + \sum y \in X_{\mathcal{L}+} f(y)
\]

Although (3) is explicitly understood, it is not useful for computing \(v_{fm}\)-values of a set of several words, e.g. the set of all the words of an attribute of an application. The following theorem establishes a recursive procedure to compute \(v_{fm}\)-values of an even infinite set of words of a variable \(X\).

**Theorem 2.1** (Nguyen et al. [40]). Let be given an EnHA \(\mathcal{A}_{\mathcal{X}_{\text{en}}}\) with \(H = \{h_j : 1 \leq j \leq p\}\) and \(H = \{h_j : -1 \leq j \leq -q\}\), we put \(\beta = \sum_{j=-1}^{p} \mu(h_j)\) and \(\alpha = \sum_{j=-1}^{q} \mu(h_j)\). So, \(\alpha + \beta + \mu(h_0) = 1\). The IVQM \(v_{fm}\) induced by \(fm\) can be calculated by the following recursive procedure:

1. For the terms of length 1:
   \(Lt(v_{fm}(0)) = 0; \)
   \(Lt(v_{fm}(c^-)) = fm(0) + \beta fm(c^-); \)
   \(Lt(v_{fm}(c^+)) = fm(0) + fm(c^-) + \alpha fm(c^+)\) and
   \(Lt(v_{fm}(1)) = 1 - fm(1).\)

2. For \(y = h_x, x \in X\): Putting \(\sigma(x) = \frac{1}{2}(1 + \text{sign}(h_x)\mu(h_0) + \mu(h, V)(\beta - \alpha))\), we have
   \(Lt(v_{fm}(h_x)) = \)
   \(Lt(v_{fm}(0)) + \text{sign}(h_x)fm(x)(\frac{1}{2}(1 + \text{sign}(h_x))\mu(h_0) + \sum_{i=\text{sign}(h_x)}^{\mu(h_0)}\mu(h_i) - \sigma(x)\mu(h_0));\)

**Example 2.1** Computing the IVQS-value of a word, say \(x = \mathcal{L}_c\) of an EnHA \(\mathcal{A}_{\mathcal{X}_{\text{en}}}\) of a variable \(X\) defined by \(G = \{c, c^\prime\}, C = \{0, W, 1\}, H_{\text{en}} = \{h_0, V, L\} \cup \{I\}, H = \{I\}, H = \{V\}\) and so \(p = q = 1\). Assume that the numeric values of the fuzziness parameters given by a domain expert are: \(fm(0) = fm(W) = fm(1) = 0.01, \)
\(fm(c^-) = 0.47, \mu(h_0) = 0.1, \mu(L) = 0.5, \mu(V) = 0.4.\) For \(x = 3\), we have \(X_3 = \{0, V, c^\prime, \mathcal{L}_c, Vc, LC, Vc^\prime, c^\prime, LLC, LC, \mathcal{L}_c, \mathcal{L}c^\prime, LLLc, c^\prime, LVc, Vc, Vc^\prime, 1\}.

1. Applying (3) to compute the IVQS-value of the word \(\mathcal{L}_c\), we have
   \(Lt(v_{fm}(\mathcal{L}_c)) = (fm(0) + fm(h_0W) + fm(h_0c^\prime) + fm(0) + fm(h_0W) + \{fm(Vc) + fm(Lc) + fm(0) + fm(Vc)\})\)
   Using formula (fm3) we obtain:
   \(Lt(v_{fm}(\mathcal{L}_c)) = \{0.01 + 0.02 + 0.05 + 0.25 + 0.01 + 0.08 + 0.1 + 0.125 + 0.1 + 0.094\} = 0.614\)

2. We may also apply Th. 2.1: Noting that \(\alpha = \sum_{j=-1}^{q} \mu(h_j) = \mu(L) = 0.5, \beta = \sum_{j=1}^{p} \mu(h_j) = \mu(V) = 0.4, \mu(h_0) = 0.1\) and \(\text{sign}(\mathcal{L}_c) = -1 = \text{sign}(VLC^+),\) we have:
   \(\sigma(c^+) = \frac{1}{2}(1 + \text{sign}(\mathcal{L}_c)\mu(h_0) + \text{sign}(VLC^+)\mu(L) - \alpha)]\)
   \(\frac{1}{2}(1 + -1)0.1 + -10.4 - 0.5)\) = 0.4 and, thus,
   \(Lt(v_{fm}(c^+)) = Lt(v_{fm}(W)) + fm(W) + \sigma fm(c^+)\)
   \(= fm(0) + fm(c^-) + fm(W) + \alpha fm(c^+)\)
   \(= 0.01 + 0.5 + 0.01 + 0.5\times0.47 = 0.755.\)

So,

\(Lt(v_{fm}(\mathcal{L}_c)) = \)
\(Lt(v_{fm}(c^+)) + \text{sign}(\mathcal{L}_c)\mu(h_0) + \mu(L) - \sigma(c^+)\mu(L)\)
\(= 0.755 + (-1)\times0.47\times\frac{1}{2}(1 + -1)\times0.1 + 0.5 - 0.4\times0.5\) = 0.614.

3. An additional discussion on the low-interpretability of LRBSs using trapezoid semantics

Since the HA-approach can establish a formalized basis to allow FRBSs handling immediately words, in this approach the terminology “LRBS” is used instead of FRBSs to emphasize the fundamental role of words and LFoCs. The study [41] can be considered as the first attempt to examine the relation between LRBSs considered as formalized (symbolic) expressions with their own real world semantics described in terms of human user words and their computational/fuzzy set semantics designed by human designer. Since human user/designer linguistic words, regarded as objects handled by human users that convey real world semantics, and computational objects handled by LRBSs are completely different, it is crucial that how one can translate the semantics of words into their respective computational objects, noting the fact that individual objects (words and computational objects) themselves in general are nothing but their relationships. So, similarly as in formalized logics, mathematics and programing languages, the concept of computational interpretation of formalized expressions
is very essential to properly represent their semantics in terms of desired computational objects.

One way to solve this question proposed in [41] is required that the interpretation assignment should transform the (structural) semantics of the words of LFoCs of the designed LRBSs to their respective computational objects handled by the LRBSs. It results in discovering structural relationships of the words of LFoCs and considering them as constraints imposed on the constructed interpretation assignment. Hedge algebras are developed since 1990 to model the order-the constructed interpretation assignment. Hedge algebras are developed since 1990 to model the order.

3.1. Some additional constraints imposed on computational interpretation

In almost studies of FRBSs, the number of fuzzy sets and, hence, the number of words of every variable used for designing FRBSs is usually assumed to be 7 ± 2. In practice, human users have right to use any words present in their natural languages to formulate their piece of knowledge accumulated from their reality, even these words are still not present in their existing knowledge. From this observation it follows that the semantics of every word of a variable must be defined in the context of the whole word-domain of the variable, that is it depends even on words they do not appear in the designed FRBSs, because otherwise when a human expert expands his domain knowledge by formulating a new additional piece of his knowledge which involves a new word being still not present in his existing old knowledge, his old domain knowledge will be changed caused by the change of the word semantics. This situation leads to proposing the following constraint:

Constraint 5 (Invariability of the computational semantics of LFoC with respect to LFoC augmentation): Any computational representation of the semantics of an LFoC should be able to maintain the current computational semantics of the existing words of when it is augmented by some new additional words.

Evidently, representation of the words of an LFoC by a unique fuzzy partition, called a single-granularity representation (SGr-representation), cannot meet this constraint as, in general, when a new fuzzy set is added in the partition at least two its adjacent fuzzy sets in the SGr-representation are changed, see Fig. 4. This results in the change of the fuzzy set semantics their words handled by the designed LRBS.

Now, we consider the following situation, recalling that a human designer of LRBSs based on the HA-approach should always start with the word-sets, LFoCs, and with his method of designing the computational semantics of the words in these word-sets. Assume that a LFoC of a variable X having two hedges R and V, where R and V stand respectively for “rather” and “very”, of a designed LRBS consists of five words in \{V_small, small, medium, large, V_large\} and assume the elements 0 and 1 to denote the least and the greatest words of X, which are understood as the absolutely small and the absolutely large, respectively, for instance. The lack of the words “R_small” and “R_large” in the case causes by the fact that it does not appear in the LKB of the designed LRBS. In the existing designing method of FRBSs, this word is not under consideration and hence
its respective fuzzy set is not necessary to design and not present in the SGr-structure of $\mathcal{F}$, as otherwise one has to solve the another question that which additional words of $X$ should be added to $\mathcal{F}$. According to our knowledge, there are no any methods proposed to solve this question in literature. In addition, as discussed above, the knowledge of a human expert is taken shape in the context of all the words of a variable $X$ existing in its total domain language, even in its total natural language. Thus, in the HA-approach, any LKB should be considered as being open and the semantics of every word of an $X$ appearing in the LKB must always be defined in the context of the whole language. In the same time, it is not practical to require that all words of $X$ should together be under consideration and assigned to computational semantics. Such an LFoC of $X$ is too large and in practice the size of an LFoC depends on the application demand. Therefore, to design a proper computational semantics of the words, the LFoCs under consideration must meet the following constraint:

**Constraint 6.** Every set LFoC $\mathcal{F}$ of linguistic words of any variable $X$ with the set $H$ of hedges of a designed LRBS should be $full$, i.e. it should satisfy the following conditions:

(i) $\{0, c^-, W, c^+, 1\} \subseteq \mathcal{F}$;
(ii) $hx \in \mathcal{F} \Rightarrow (\forall h' \in H)(h'x \in \mathcal{F})$ (either all words $hx, h \in H$, should together belong to $\mathcal{F}$, or they all do not);
(iii) $x \in \mathcal{F}$ & $x = hx' & h \in H \Rightarrow x' \in \mathcal{F}$ (closed with respect to the taking of its ancestors).

That is, $\mathcal{F}$ satisfies these conditions iff assuming $\kappa$ is the maximal length of the word lengths of $\mathcal{F}$, it must be of the form $X_{\kappa,0}$, the set of all words of $X$ whose lengths are not greater than $\kappa$.

It is worth emphasizing again that in practice it is possible that only a part of the words of $\mathcal{F}$ explicitly appears in the LKB of the designed LRBS, but their computational semantics must be defined in the entire LFoC $\mathcal{F}$. That is that the computational semantics of the words of $\mathcal{F}$ should be present in the computational representation of $\mathcal{F}$.

**Definition 3.1.** (i) A computational representation of an LFoC of a variable is said to be VI-interpretable if it satisfies six constraints, Constraints 1 – 6. If a computational representation satisfies only four constraints, Constraints 1 – 4, it is said to be IV-interpretable.

(ii) An LRBS is said to be LFoC-interpretable if all computational representations of its LFoCs are 6-interpretable.

3.2. An LFoC-interpretable trapezoid representation of LFoCs

The study [41] examined triangle MGr-representations of LFoCs which are IV-interpretable. In this study, we will show that these MGr-representations are also VI-interpretable and examine MGr-representations with trapezoid fuzzy sets instead of triangle fuzzy sets.

Let us consider a variable $X$ with a set $C$ of constants and a set $H = \{L, V\}$ of hedges “little” and “very”, respectively. There are two algebraic structures associated with $X$: an HA $\mathcal{AX}$ and its enlargement, EnHA $\mathcal{AX}_en$. Let be given an LFoC $\mathcal{F}_X$ of $X$ which is in the form $\mathcal{F}_X = X_{\kappa,0} \subseteq X$ of the algebra $\mathcal{AX}$. Since, as discussed above, SGr-representation of an LFoC cannot be interpretable as it cannot maintain the generality-specificity relation of words, we have to deal only with multi-granularity representation (MGr-representations) of LFoCs.

First of all, for a given LFoC $\mathcal{F}_X$ of $X$, we describe a procedure to produce a trapezoid MGr-representation of $\mathcal{F}_X$ from the semantic structure of the LFoC $\mathcal{F}_X$. Let be given values of the independent fuzziness parameters of $\mathcal{AX}_en$ of $X$, denoted by a vector $\pi = (\mu(h_0), \mu(L), fm(0), fm(c), fm(W), fm(1))$. Assume that $X_k = \{x \in X: |x| = k\}$ and $\mathcal{F}_k = X_{\kappa,0} \cup \{0, W, 1\}$ and set $\mathcal{F}_k = X_{\kappa,0}$, the set of the words of specificity degree $k$. The fuzzy sets designed for the words of $\mathcal{F}_k$, will be taken to form a fuzzy partition of the MGr-representation of $\mathcal{F}_X$, however, its is not a strong partition. To overcomes this shortcoming, we extend $\mathcal{F}_k$ to $\mathcal{F}_{k,\kappa} = \mathcal{F}_k \cup \{0, W, 1\}$, where $0_k$ and $1_k$ indicate, respectively, the least and the greatest words of specificity degree $k$, $k = 0, \ldots, \kappa$. We introduce the following procedure for automatically producing trapezoid word semantics from $\mathcal{F}_{k,\kappa}$:

**Procedure PrMGr($\pi(\mathcal{F}_{k,\kappa})$):**

Inputs: $\circ$ An LFoC $\mathcal{F}_X$ of $X$ and the HA and its EnHA of $X$ to represent the context-dependent semantics of $\mathcal{F}_X$;
$\circ$ A given vector of the independent fuzziness parameters of $\mathcal{AX}_en$ of $X$, denoted by a vector $\pi = (\mu(h_0), \mu(L), fm(0), fm(c), fm(W), fm(1))$.

**Step 1:** Construct $k$-level specificity $\mathcal{F}_{k,\kappa}$ of $\mathcal{F}_X$, $k = 0, 1, \ldots, \kappa$.
- Partition $\mathcal{F}_{k,\kappa}$ into $k + 1$ levels, $\mathcal{F}_{k,\kappa} = \{x \in \mathcal{F}_{k,\kappa}: |x| = k\}$, called $k$-specificity level or, shortly, $k$-level of $\mathcal{F}_{k,\kappa}$, $k = 0, 1, \ldots, \kappa$, where for $k = 0, \mathcal{F}_{k,\kappa} = C = \{0, W, 1\}$, the set of the constants of the EnHA.
- For $k > 0$, in order to construct the $k$-level fuzzy partition of the desired MG-representation of $\mathcal{F}_{k,\kappa}$ to
represent the semantics of the \( k \)-level \( \overline{S}_{k,k} \), two artificial least and greatest words of the \( k \)-level, denoted respectively by \( 0_k \) and \( 1_k \), are added to \( \overline{S}_{k,k} \), denoted also by \( \overline{S}_{k,k} \), for simplicity.

Step 2: Compute the core regions of the words of each \( k \)-level \( \overline{S}_{k,k} \), \( k = 0, 1, \ldots, \kappa \)
- For \( k = 0, 1, \ldots, \kappa \), compute the core region (interval) of every word, using the \( fm \)-induced IVQM of the EnHA of \( \mathcal{X} \): \( v_{fm}(x) = \mathcal{Z}(h_{0x}), x \in \overline{S}_{k,k} \) (see Section 2).
- Locate the core regions in \([0, 1] \) (their fuzzy sets attain the value 1 on these core regions). This ensures that the order of the numeric cores and the core regions is in accordance with the order of their respective words of \( \overline{S}_{k,k} \).

Step 3: Construct trapezoid fuzzy partitions representing the semantics of the \( k \)-level \( \overline{S}_{k,k} \), \( k = 0, 1, \ldots, \kappa \)
- Construct small bases of the trapezoid membership functions of the words of \( \mathcal{X} \) so that their cores are the respective core regions;
- For each constructed small base, the respective large base is the interval defined by the right end-point of the left adjacent core region and the left end-point of the right adjacent region of the small base.
- Construct the trapezoid membership functions based on their constructed small and large bases so that the trapezoids on every level of the constructed MGr-representation form a strong fuzzy partition.

Output: - The fuzziness intervals of \( \overline{S}_{k} \), \( \mathcal{Z}(\overline{S}_{k}) = \{ \mathcal{Z}(x) : x \in \overline{S}_{k} \} \);
- A fuzzy MGr-representation of \( \overline{S}_{k} \) with \( \kappa + 1 \) trapezoid fuzzy partitions of specificity levels \( \overline{S}_{k,k}, k = 0, 1, \ldots, \kappa \).

As an illustration, Fig. 4 represents an MGr-structure with three levels of trapezoid fuzzy sets produced by the procedure \( Pr_{MGr} \).

Now, we are going to examine the VI-interpretability of the constructed trapezoid fuzzy set representations of the given LFoCs of variables. By \( \overline{S}_{\kappa} \) we denote the potential infinite word-set LFoC of a variable \( \mathcal{X} \): \( \overline{S}_{\kappa} = \bigcup_{0 \leq \alpha < \infty} \overline{S}_{\alpha} \).

By Constraint 4, we first construct an order-based computational structure \( (C_{Trp}(\mathcal{X}), \preceq) \) consisting of trapezoid fuzzy sets, each of which is denoted by a triple \((a, b, c)\), where \( a, b \) and \( c \) are the parameters of a trapezoid constructed for a word of \( \overline{S}_{\kappa} \) by the above procedure \( Pr_{MGr} \). \([a, c]\) is the large base and \( b \) indicates the small based of the constructed trapezoid. This establishes an interpretation assignment \( \mathcal{Z}: Dom(\mathcal{X}) \rightarrow C_{Trp}(\mathcal{X}) \). The order relation \( \preceq \) and the generality-specificity relation, denoted by \( GST_{Trp} \), on \( C_{Trp}(\mathcal{X}) \) is defined as follows:

**Definition 3.2.** For any two trapezoids \((a, b, c)\) and \((a', b', c')\) in \( C_{Trp}(\mathcal{X}) \),

(i) \( GST_{Trp} \): The relation \( GST_{Trp} \) of the two triples is defined as follows: \( GST_{Trp}(a, b, c), (a', b', c') \) iff \([a, c] \supset [a', c']\). \( (ii) \preceq : (a, b, c) \preceq (a', b', c') \) if either at least two parameters necessarily including the core of the first fuzzy set are respectively smaller than the corresponding ones of the second one or, for any two fuzzy sets whose both cores are positioned either at the left end- or the right end-point of the universe
singleton set or the interval core of the fuzzy set of $x$.

Semantics of the words of $\mathcal{F}_k$

Constraint 6. Assumed to be of the form $\mathbf{E}$.

Arrangements of the relative fuzzy set positions of the respective MGr-representations of a given LFoC are maintained by its respective semantic order-based structures of the EnHAs.

In addition, importantly, for both cases we have that the trapezoid fuzzy sets of the words of $X$’s from their respective semantic order-based structures of the EnHAs.

To prove Constraints 3 and 4 we observe that triangle MGr-representations and the respective trapezoid MGr-representations of LFoCs are different only from the cores of their fuzzy sets: the cores of trapezoid fuzzy sets are singleton sets while the ones of triangle fuzzy sets are intervals. Therefore, it is crucial that for both cases the fuzzy set semantics of a word $x$ can be written as triples in the form $(a_x, b_x, c_x)$, where the bold letter $b_x$ is understood as the core of singleton set or the interval core of the fuzzy set of $x$.

In addition, importantly, for both cases we have that $x \leq y \Rightarrow b_x \leq b_y$. This ensures that the arrangements of the relative fuzzy set positions of the respective MGr-representations of a given LFoC are identified. Accordingly, we can repeat the corresponding arguments in the study [41] to prove that the trapezoid MGr-representations of a given LFoC meet Constraints 3 and 4 presented above.

By the procedure $Pr_{MG}$, the given LFoC is always assumed to be of the form $\mathcal{F}_k$ and, hence, it satisfies Constraint 6.

To prove that the trapezoid MGr-representation of $\mathcal{F}_k$ produced by the procedure $Pr_{MG}$ meets Constraint 5, $\mathcal{F}_k$ is augmented by adding a new word $z$. Since $\mathcal{F}_k$ is full, it implies that the length $|z| > \kappa$, which results in the fact that the trapezoid fuzzy set of $z$ constructed by $Pr_{MG}$ lies on the level greater than $\kappa$. Therefore, this cannot change the shapes of the trapezoids appearing in any level $k$ lower or equal to $\kappa$. That is, the trapezoid semantics of the words of $\mathcal{F}_k$ is maintained by its augmentation and the theorem is completely proved.

4. Genetic method to design of LFoC-interpretable LRBSs for regression problems

The objectives of the section are:

- a) Contribute to show the role of the LFoC-interpretability of fuzzy set representation of LFoCs in designing LRBSs.
- b) To show the performance of trapezoid fuzzy sets over the one of triangle fuzzy sets in the situation that rare studies to design genetic FRBSs using trapezoid fuzzy sets to solve classification/regression problems.

Noting that words that considered as mathematical objects and as subjects to generate their computational semantics, including their fuzzy set based semantics have not been examined in the framework of fuzzy sets. In this context, though multi-granularity structure of fuzzy sets representing the semantics of word-sets in the HA-approach were applied in [39,40], till the study [41] a novel approach to the low-interpretability of FRBSs is first proposed and discussed. This establishes an approach distinguished from the existing ones as it is developed based on the own semantics of linguistic words, from which the computational word semantics is generated. So, the first objective of the section is to contribute to conclude that the LFoC-interpretability of LRBSs can actually enhance the performance of the designed LRBSs, noting that, in the uncertainty environment, such a conclusion requires to make sufficient number of comparative simulation studies. Triangle fuzzy sets are used to represent the semantics of words in [41]. It seems that the shortcoming is the fact that the membership values of their surrounding cores are unreasonably changed and this feature of trapezoid fuzzy sets is clearly better. To meet the objective b), it is necessary to require that we should apply the same method to design LRBSs except the fuzzy set based semantics of words: for one experiments used triangles but for the others used trapezoids. By this reason, we used the same HA-based method to design LRBSs to solve regression problems examined in [41] with some slight modification, regardless which fuzzy set are used. Since the most part of the method is described in details in that paper, we focus mainly on the following modification points:

- A more detailed discussion on how determining suitable fuzziness parameters of the variables.
- A way to determine the linguistic rule corresponding to a pattern of the given dataset: By the presence of the specific words representing the semantics core of the words in $X_{m(4)}$ with the length less than $k$, we use the fuzziness intervals of the words in $X_{m(k)}$, which forms a binary partition of the feature universe, instead of the similarity intervals as in the studies [39-41] in generating rules from given patterns.
- Using BLX-alpha crossover instead of one-point crossover on the chromosomes coding the independent fuzziness parameters of the variables.
In the evolutionary process, when the crossover operator was not realized on the rule base (LRB) part, operator adding \( y \) rules to the current LRB must be done and, then, only new rules generated from patterns that are not covered by the current LRB are taken into account in order to increase the effectiveness of the new generated rules.

### 4.1. Descriptions of an HA-based genetic method to design LRBSs

Let be given a regression dataset \( D = \{d_i = (a_{i1}, a_{i2}, \ldots, a_{in}) \in [0, 1]^{n+1} : i = 1 \) to \( N_D \} \) with \( n \) input-features, \( X_j \) \( j = 1 \) to \( n \) and one output feature \( X_{n+1} \), whose universes are all normalized to be \([0, 1]\). First we need some notations. A linguistic rule representing a piece of knowledge from the dataset \( D \) is of the form:

\[
\text{if } r \text{ then } X_{n+1} \Rightarrow X_{x,r+1} (4)
\]

where \( x_{r+1} \in \text{Dom}(X_j) \), the word-domain of \( X_j \) \( j = 1 \) to \( n + 1 \). Denoting the antecedent of \( r \) simply as \( x_r \), the rule \( r \) can be written shortly by \( x_r \Rightarrow y_r \). Assume that the membership function of the fuzzy set assigned to each word \( w \), denoted by \( \mu_w \), and put \( \tilde{a}_i = (a_{i1}, a_{i2}, \ldots, a_{in}) \) the input vector of a given datum \( d_i = (a_{i1}, a_{i2}, \ldots, a_{in}) \in D \). Then, the product \( \mu_{x_{r+1}}(a_{i1}) \cdots \mu_{x_{r+1}}(a_{in}) \) is denoted by \( \mu_{x_r}(\tilde{a}_i) \).

The proposed method aims to design an interpretable LRBS \( S_D \), whose RB consists of the linguistic rules of the form of (4) and whose parameters are the independent fuzziness parameters of the dataset features, such that the system \( S_D \) can approximates the function \( f_0 \) defined by \( D \) in the sense that it may minimize the MSE defined as follows:

\[
\text{MSE} = \frac{1}{2N_D} \sum_{i=1}^{N_D} (\hat{y}_i - y_i)^2 , \quad \text{where, for } i = 1 \text{ to } N_D \]

\[
\hat{y}_i = \frac{\sum_{w \in \text{Dom}(X_j)} \mu_w(a) \text{Cent}(v_w(x_{r+1})))}{\sum_{w \in \text{Dom}(X_j)} \mu_w(a)}
\]

where \( \text{Cent}(v_w(x_{r+1})) \) denotes the center of \( v_w(x_{r+1}) \) which is the fuzziness interval of the word \( h_0 \). So, the center of \( v_w(x_{r+1}) \) is used instead of the gravity center of the fuzzy set \( FS(x_{r+1}) \), if \( \sum_{w \in \text{Dom}(X_j)} \mu_w(a) \neq 0 \), and, otherwise, \( \hat{b} \) is calculated by the method given by Alcalá et al. [1] which is applied in [41] and, hence, not repeated here.

The main tasks of the proposed HA-based method are described as follows:

**Task 1: Determining EnHAIs associated with the variables (or features) of a given regression dataset**

This task is specific to the HA-approach. An EnHA \( AX_{en} \) associated with a variable \( X \) is uniquely determined by specifying two primary words, \( c^- \) and \( c^+ \), a negative and a positive hedge set, \( H^- \) and \( H^+ \), an artificial hedge \( h_0 \) to generate the semantics core of the words and a table of the relative signs, \( \text{sign}(h, k) \) \( \in \{-1, +1\} \), between the hinges of \( H = H^- \cup H^+ \). As shown in [39–41], in many applications, it is sufficient to use only singleton sets of hinges, say \( H = \{L\} \) and \( H = \{V\} \). Of course, the effect of a developed method depends on which hinges are selected, but it is not the subject of the study. To quantify the word-domains of \( X \), it is necessary to specify the values of the independent fuzziness parameters, \( \text{fm}(c^-), \text{fm}(W), \mu(L), \mu(V), \mu(h_0) \). The values of the remaining fuzziness parameters are determined by the equalities, \( \text{fm}(0) = \text{fm}(c^+) + \text{fm}(W) + \text{fm}(c^-) + \text{fm}(1) = 1 \) and \( \mu(L) + \mu(V) + \mu(h_0) = 1 \).

In general, human user/expert can be more easy to specify the fuzziness parameters values the to specify the parameters of fuzzy sets, as he daily uses his words and, hence, he can immediately feel the meaning of these words and hedges, while he does not do so when dealing with the fuzzy set parameters. In genetic methods, these fuzziness parameter values can be optimized based on given optimal criteria. In such both manual and automatic cases, certain additional constraints imposed on parameter values can be specified, e.g. the following:

\[
\begin{align*}
0 & \leq \mu_{\text{min}} \leq \mu(0) \leq \mu_{\text{max}} <= 1 \\
0 & < \mu_{\text{min}} \leq \mu(c^+) \leq \mu_{\text{max}} <= 1 \\
0 & \leq \mu(L) \leq \mu(W) \leq \mu_{\text{max}} <= 1 \\
0 & \leq \mu_{\text{min}} \leq \mu(1) \leq \mu_{\text{max}} <= 1 \\
0 & < \mu_{\text{min}} \leq \mu(h_0) \leq \mu_{\text{max}} <= 1
\end{align*}
\]

\[
(5)
\]

**Task 2: Using the fuzziness intervals of the words of the set of all words of length not greater than \( k \), \( X_{en(k)} \), to generate linguistic rules from selected individual numeric patterns**

In many approaches to design genetic FRBSs to solve classification or regression problems, initial fuzzy rules are generated by considering all combinations of the fuzzy sets specified for the individual features of datasets, which results in an exponential explosion of rules under consideration. Inspired by studies of the so-called ad hoc data-oriented approach to generate initial fuzzy rules immediately from a given collection of numeric data samples, such as by Wang and Mendel [48], Ishibuchi et al. [23] or by Cordón and F. Herrera [15], Nguyen et al. [39] proposed a method to generate linguistic rules from each individual real pattern based on the
similarity intervals of the words of $X_{\text{ant}}$, which constitute a binary partition of the feature universe, for a given positive integer $k$ to limit the lengths of its words. This ensures that every numeric feature value of a pattern must fall into a similarity interval of a words of the respective feature and, then, the corresponding words are selected to form a linguistic rule.

In the case of an EnHAs, $\mathcal{AX}_{\text{en}} = (X_{\text{en}}, G, C, H_{\text{en}}, \preceq)$ – an enlargement of $\mathcal{AX} = (X, G, C, H, \preceq)$, the artificial words of the form $h_{\text{en}}$ are utilized to represent the semantics core of $x$, as they are not original words of the respective linguistic variable and, hence, they are formalized as fixed points of EnHAs. A specific feature of EnHAs is that the presence of the semantics cores of the words, the fuzziness intervals of the words of $X_{\text{en}(i)}$ constitute a binary partition and the fuzziness interval of every word of $X_{\text{en}(i)}$ consists of values can be considered as most compatible with this word in comparison with the other ones.

We consider now a given regression dataset $D = \{d_i = (x_{i1}, a_{i2}, \ldots, a_{in}, x_{i(n+1)}); i = 1$ to $N\}$ From the above binary partition generated from $X_{\text{en}(i)}$ it follows that, for every pattern $d_i$ of $D$ and for $j = 1$ to $n + 1$, if the feature value $a_{ij}$ of $d_i$ falls in a fuzziness interval of a word $x_{ij}$ of $X_{\text{en}(i)}$, then the following rule $r(d_i)$ is understood as generated from $d_i$, called the primary (linguistic) rule:

$$r(d_i) = X_1 = x_{i1} \land \ldots \land x_{in} \Rightarrow x_{i(n+1)} \quad (6)$$

It is known that the length of a rule is defined by the number of conditions of the form “$X_j = x_{ij}$” in its antecedent. So, when $n$ is large, the length of the primary rules generated from the patterns are incomprehensive or they are not interpretable in terms of complexity. Therefore, as usual, only shorter rules produced from the primary rules are considered. Firstly, for convenience, the rule (6) can be written shortly as $x_{i1} \land \ldots \land x_{in} \Rightarrow x_{i(n+1)}$. Then, the shorter rules produced from (6) by replacing some words in its antecedent with “don’t care”, called secondary rules of the primary rule (6), are taken into account, similarly as examined by Ishibuchi et al. [23]:

$$(r_{q}) x_{i1} \land \ldots \land x_{iq} \Rightarrow x_{i(q+1)} . \quad (7)$$

where $x_{iq} \in \mathcal{I}_j \cup \{\text{don’t care}\}$.

For illustration, we give the following example.

Example 4.1. Linguistic rules generated from a numeric pattern: Let us consider a pattern $d = (d_1, d_2, d_3, d_4, d_5) = (0.571427, 0.678572, 0.449666, 0.192073, 0.432208)$ of the regression dataset ELE2 with 5 variables. The given independent fuzziness parameter values of each variable are provided in this case by the proposed algorithm EnHA-Tz-0-MG (TzOM) developed in Section 5 and run on ELE2. For instance, for the variable $X_4$ they are $f_{m}(0_8) = 0.02110243$, $f_{m}(1_8) = 0.08699807, f_{m}(c_8) = 0.3016947, f_{m}(l_8) = 0.08666857, \mu(b_8) = 0.500458, \mu(b_8) = 0.1933572$. Then, all fuzziness intervals of the words of LFOC = $X_{(3)}$ and they form a binary partition, for $j = 1$ to $5$, are computed. Since for every variable they are disjoint, we have: $d_1 \in \mathcal{I}_3(W_1) = \{0.5246565, 0.6116368\}, d_2 \in \mathcal{I}_1(\text{Lc}_1) = \{0.6578662, 0.6870604\}, d_3 \in \mathcal{I}_5(\text{Lc}_5) = \{0.3987681, 0.4574956\}$, $d_4 \in \mathcal{I}_3(\text{C}_3) = \{0.175283, 0.2726489\}$ and $d_5 \in \mathcal{I}_5(\text{C}_5) = \{0.3987681, 0.4574956\}$. By the above method, $d$ generates the following primary rule:

$$r(d) : (X_1 = W) \land (X_2 = \text{Lc}) \land (X_3 = \text{Lc}) \land (X_4 = c) \Rightarrow (X_5 = \text{Lc}).$$

Assuming that $\tau_{\text{min}} = 3$, the secondary rules then are produced from $r(d)$ by replacing the words of $n - 1$ attributes arbitrarily selected from the antecedent of $r(d)$ by “don’t care”, where $l < 3$. For instance, for $l = 1$, the secondary rules are of length 1 are the following:

$$(X_1 = W) \Rightarrow (X_2 = \text{Lc}); (X_2 = \text{Lc}) \Rightarrow (X_5 = \text{Lc}); (X_3 = \text{Lc}) \Rightarrow (X_5 = \text{Lc}) \land (X_4 = c) \Rightarrow (X_5 = \text{Lc}).$$

For $l = 2$, some examples of the secondary rules are of length 2 are as follows:

$$(X_1 = W) \land (X_3 = \text{Lc}) \Rightarrow (X_4 = \text{Lc}); (X_2 = \text{Lc}) \land (X_4 = c) \Rightarrow (X_3 = \text{Lc}); (X_3 = \text{Lc}) \land (X_4 = c) \Rightarrow (X_5 = \text{Lc}).$$

It can be seen that the consideration of only such secondary rules generated from the numeric patterns will decrease considerably the searching space in comparison with the case where the rules under consideration are generated by considering all possible combinations of the words of LFOCs specified for the individual features of datasets.

Task 3: Coding individuals and generating initial population

As aforementioned, the study is aimed to show the performance of the interpretable trapezoid MGr-representation of LFOCs in designing LRBSs, therefore, different from the usual FRBSs in which fuzzy rules are ready in the form of fuzzy expressions containing linguistic labels assigned to their fuzzy sets, the RBs of LRBSs are sets of linguistic rules of the form (7), written for short by LRBs, whose elements are very similar sentences in natural languages. To handle these linguistic rules, words should be mapped
to the fuzzy sets using the interpretation mapping defined by the MGR-representations of the given LFoCs produced by the procedure Pr\(_{\text{MGR}}\) for the given independent fuzziness parameter values. In this formalism, the defuzzification values of fuzzy sets are simply the center of the corresponding fuzzy set core regions (the \(v_{\text{core}}\)-values or the interval-values) of the corresponding words. The fuzzy reasoning are the same as those examined many recent approaches in this field. So, the main difference of the LRBSs to the counterpart FRBSs is the fuzzy MGr-representations of this field. So, the main difference of the LRBSs to the same as those examined many recent approaches in this field.

Denoting the vector of the independent fuzziness parameter values of each attribute \(X_j\) by \(\pi_j = (\mu_0(X_j), \mu(L_j), \mu(L_j), f m(0), f m(c_j^-), f m(W_j), f m(1)),\) for \(j = 1, \ldots, n + 1\), the component \(C_j\) is the vector \(\pi = (\pi_1, \ldots, \pi_n, \pi_n, 1, \ldots, n + 1),\) by which the MGR-representations of given LFoCs \(\mathcal{Y}_j\)’s of the form \(X_{\text{core}}(\mathcal{Y}_j)\) can be calculated, where we assume for simplifying the presentation that the cardinalities of \(\mathcal{Y}_j, |\mathcal{Y}_j|\), for all \(j\)’s, are equal to each other.

Suppose that, for every \(X_j\), its independent fuzziness parameters are constrained by (5), (fm1) and (fm4).

Similarly also as the studies [1,2] mentioned above, the component \(C_{\text{RB}}\) is designed to encode linguistic rules with real numbers by the following way. Assume that \(\mathcal{Y}_j \cup \{\text{don’t care}\} = \{y_{jk} : k = 0, 1, \ldots, N_j\} \subseteq X_{\text{core}},\) based on which every word \(y_{jk}\) is encoded by its index \(k\) for \(k = 0, 1, \ldots, N_j\) and \(j = 1, \ldots, n + 1\), where \(y_0 = \text{don’t care}\) and, hence, the code of “don’t care” is “0”. Then, any rule \(r_j\) of the form (7) with \(x_{dj} = y_{jk_{dj}}\) can be encoded as a vector \((k_{j1}, \ldots, k_{jn+1})\).

For an LRBS \(\mathcal{P}\) of the form (7), where we assume \(M\) being constrained by \(M \in [M_{\text{min}}, M_{\text{max}}]\), the individual representing both the LRBS and the vector \(\pi\) is encoded as represented in Fig. 5.

A specific feature of the HA-approach is that, for a given fuzziness parameter vector \(\pi = (\pi_1, \ldots, \pi_n, \pi_n, 1, \ldots, n + 1)\), the set of the fuzziness intervals of the LFoCs \(\mathcal{Y}_j\)’s, \(\mathcal{Y}(\mathcal{Y}_j) = \{y_{jk} : k = 0, 1, \ldots, n + 1\}\) can be computed and, then, each numeric pattern can generates a set of linguistic rules of length not greater than a prespecified positive integer \(\tau_{\text{max}}\). Similarly as in Nguyen et al. [41], we introduce the following procedure:

**Procedure Pr\((d, \mathcal{Y}(\mathcal{Y}_j) ; j \leq n + 1), \tau_{\text{max}}\)** for generating a linguistic rule of length not greater than \(\tau_{\text{max}}\) from a given pattern \(d\) for a given vector \(\pi\) of fuzziness parameters:

- For the given pattern \(d_j = (d_{j1}, d_{j2}, \ldots, d_{j(n+1)})\) \(\in D\), produce the linguistic rule \(r(d_j)\) of the form (6);  
- Generate randomly in turn an integer \(l, 0 < l \leq \tau_{\text{max}}\) and, then, integers \(j_1, \ldots, j_l\) such that \(0 < j_1 < \ldots < j_l \leq n\) to produce from the obtained rule \(r(d_j)\) a rule of length \(l\) of the form (7), denoted by \(r_{j_1 \ldots j_l}\)  

\(r_{j_1 \ldots j_l} = \mathcal{X}_{j_1} \Rightarrow \mathcal{X}_{j_1} \ldots \mathcal{X}_{j_l} \Rightarrow \mathcal{X}_{j_1 \ldots j_l}\), where \(d_{i(kj_{j})} \in \mathcal{X}_{j_1 \ldots j_l}\). □

The procedure \(\text{Pr} (d, \mathcal{Y}(\mathcal{Y}_j) ; j \leq n + 1), \tau_{\text{max}}\) will be applied to construct an initial population \(P_0\) of \(N_0\) rules of an evolutionary algorithm or to generate any rule when it is required by performing genetic operations.

**Task 4. Defining genetic operators**

**Crossover operators:** For any parent \(p_1\) and \(p_2\), each of which has two disjoint components \(C_u\) and \(C_{\text{RB}}\) with different semantic functionalities, to find optimized solutions we define two crossover operators that are independently done on each component.

(i) BLX-α crossover operator on \(C_u\): This operator works on the component \(C_u\) of \((n + 1)\) chromosomes with 6 genes. For a piece \(C\) of any parent \(p\), denote by \(p|C|\) the restriction of a parent \(p\) on \(C\). To define this type of crossover, we assume that \(p_1|C|\) can be written as follows, where \(\sigma_j : j = 1\) to \((n + 1) = \sigma_1 \sigma_2 \ldots \sigma_{n+1}\), \(\sigma_j\)’s are strings of symbols:

\[p_1|C|_j = (\mu_0(L_j)_{p_1}, \mu(L_j)_{p_1}, f m(0)_{p_1, j}, f m(c_j^-)_{p_1, j}, f m(W_j)_{p_1, j}, f m(1)_{p_1, j})\]
where $\mu(h_0)_p \in [\mu_{\min}, \mu_{\max}]$, $\mu(L)_p \in [\mu_{\min}, \mu_{\max}]$, $f_m(0)_p \in [0, \max]$, $m(0)_p \in [c_{\min}, c_{\max}]$, $m(W) \in [W_{\min}, W_{\max}]$ and $m(1)_p \in [1, \max]$ are real codes of the fuzziness parameters of the $j$th chromosomes of $C_p$ of $p$, for $j = 1$ to $n + 1$ and $i = 1, 2$. The intervals are subsets of $[0, 1]$, the normalization of the variable universes, and given by domain users to maintain the semantics of the hedges and the primary words of the variable. The BLX-$\alpha$ crossover operator works on two $C_p$-components of $p_1$ and $p_2$ to randomly produce a $C_p$-component of a child $\sigma$, whose $j$th chromosome is denoted by $(\mu(h_0)_p, \mu(L)_p, m(0)_p, m(1)_p)$. $f_m(c_j), f_m(W), f_m(1)_p$, defined as follows:

1. $\mu(h_0)_p$ is randomly selected from the interval $[l_{\mu(h_0)_p}, u_{\mu(h_0)_p}]$ calculated from the given values $\mu_{\min}$ and $\mu_{\max}$ as follows: putting $\mu_{\min} = \min(\mu(h_0)_p_1, \mu(h_0)_p_2)$, $\mu_{\max} = \max(\mu(h_0)_p_1, \mu(h_0)_p_2)$ and $t_1 = (t_{\mu_{\min}} - t_{\mu_{\max}})^\gamma$, we have $l_{\mu(h_0)_p} = \max(\mu_{\min}, \mu_{\min} - \gamma I_1)$ and $u_{\mu(h_0)_p} = \min(\mu_{\max}, t_{\mu_{\max}} + I_1)$.

2. $\mu(L)_p$ is randomly selected from the interval $[l_{\mu(L)_p}, u_{\mu(L)_p}]$ calculated from the given values $\mu_{\min}$ and $\mu_{\max}$ as follows: putting $t_{\mu_{\min}} = \min(\mu(L)_p_1, \mu(L)_p_2)$, $t_{\mu_{\max}} = \max(\mu(L)_p_1, \mu(L)_p_2)$ and $t_2 = (t_{\mu_{\max}} - t_{\mu_{\min}})^\gamma$, we have $l_{\mu(L)_p} = \max(\mu_{\min}, t_{\mu_{\min}} - t_{\mu_{\max}})$ and $u_{\mu(L)_p} = \min(\mu_{\max}, t_{\mu_{\max}} + t_{\mu_{\min}})$.

3. $f_m(c_j)_p$ is randomly selected from the interval $[l_{f_m(c_j)_p}, u_{f_m(c_j)_p}]$ calculated from the values $c_{\min}$ and $c_{\max}$ as follows: putting $t_{c_{\min}} = \min(f_m(c_j)_p_1, f_m(c_j)_p_2)$, $t_{c_{\max}} = \max(f_m(c_j)_p_1, f_m(c_j)_p_2)$ and $t_3 = (t_{c_{\max}} - t_{c_{\min}})^\gamma$, we have $l_{f_m(c_j)_p} = \max(t_{c_{\min}}, t_{c_{\max}} - t_{c_{\min}})$ and $u_{f_m(c_j)_p} = \min(t_{c_{\max}}, t_{c_{\max}} + t_{c_{\min}})$.

4. $f_m(W)_p$ is randomly selected from the interval $[l_{f_m(W)_p}, u_{f_m(W)_p}]$ calculated from the values $W_{\min}$ and $W_{\max}$ as follows: putting $t_{W_{\min}} = \min(f_m(W)_p_1, f_m(W)_p_2)$, $t_{W_{\max}} = \max(f_m(W)_p_1, f_m(W)_p_2)$ and $t_5 = (t_{W_{\max}} - t_{W_{\min}})^\gamma$, we have $l_{f_m(W)_p} = \max(t_{W_{\min}}, t_{W_{\max}} - t_{W_{\min}})$ and $u_{f_m(W)_p} = \min(t_{W_{\max}}, t_{W_{\max}} + t_{W_{\min}})$.

5. $f_m(1)_p$ is randomly selected from the interval $[l_{f_m(1)_p}, u_{f_m(1)_p}]$ calculated from the values $1_{\min}$ and $1_{\max}$ as follows: putting $t_{1_{\min}} = \min(f_m(1)_p_1, f_m(1)_p_2)$, $t_{1_{\max}} = \max(f_m(1)_p_1, f_m(1)_p_2)$ and $t_6 = (t_{1_{\max}} - t_{1_{\min}})^\gamma$, we have $l_{f_m(1)_p} = \max(t_{1_{\min}}, t_{1_{\max}} - t_{1_{\min}})$ and $u_{f_m(1)_p} = \min(t_{1_{\max}}, t_{1_{\max}} + t_{1_{\min}})$.

(a) An ordinary one-point crossover operator on $C_{RB'}$. The single cut-point is randomly selected in the interval $[1, \rho_{\min} - 1]$, where $\rho_{\min}$ is the least number of linguistic rules appearing in each of the parent $p_1$ and $p_2$.

- **Mutation operators**: Similarly as above, given an individual $p$ encoded as above, two mutation operators will be defined on the two components $C_p$ and $C_{RB'}$ of $p$ independently and described as follows.

Mutation operator on the component $C_p$: Select randomly a $f$-gene of $C_p$, $i \in [1, (n+1)]$. If it represents a value of $\mu(h_p)$, or, respectively, $f_m(c_p)$, $f_m(W_p)$, $f_m(1)_p$, it is replaced with a real value selected randomly from $[\mu_{\min}, \mu_{\max}]$ (or, respectively, from $[l_{f_m(c_p)}, u_{f_m(c_p)}]$, $[l_{f_m(W_p)}, u_{f_m(W_p)}]$, $[l_{f_m(1)_p}, u_{f_m(1)_p}])$.

Mutation operator on $C_{RB'}$: Since the coding of $C_{RB'}$ is the same as in Nguyen et al. [41], the two mutation operators $\rho_{\min}$ and $\rho_{\max}$ on it are the same as in that paper:

+ **Mutation operator $M_{Ad}$ adding $\gamma$ rules to the current $RB$ of $M$ rules represented by $C_{RB'}$.** It is a procedure that selects randomly an integer $\gamma$ in $[1, \gamma_{\max}]$, where $\gamma_{\max}$ is prespecified, and, if $M + \gamma > M_{\max}$ then, it adds $\gamma' = \min(\gamma, M_{\max} - M)$ rules generated by $\gamma'$ patterns randomly selected in turn from the given dataset to the current $RB$, using the procedure $Pr(d, [\delta(\gamma)]; j \leq n + 1)$, $\gamma_{\max}$.

+ **Mutation operator $M_{Ch}$ changing the gene values at $\delta$ points of $C_{RB'}$: It** is a procedure that selects randomly an integer $\delta$ in $[1, \delta_{\max}]$, where $\delta_{\max}$ is prespecified, and, then, randomly select $\delta$ points of $C_{RB'}$. Change in turn the gene value, which is encoded by an integer in $[0, N_j]$, where $N_j = \lceil \delta \rceil$, at every selected point into a value selected randomly from the integers either in $[0, N_j]$, for $j = n + 1$, or in $[1, N_j]$, for $j = n + 1$, noting that “0” is the code of “don’t care”.

Note that, in evolutionary process next, if a rule becomes of length 0, due to the presence of “don’t care”.

![Figure 5](image-url): An individual of $(n + 1) + M$ chromosomes representing an $RB$ of $M$ linguistic rules.
Care” in all input variables, it will be eliminated and if several rules become identical, retain only one (sub-step 2.1.3, 2.1.4).

Task 5. Two objective evolutionary algorithm to solve regression problems

As aforementioned, the developed Multiple-objective Optimization Evolutionary Algorithm (MOEA) is aimed to design effective but interpretable LRBSs to solve regression problems. Therefore, there are two main objectives: for regression problems, the first one is, of course, the MSE criterion. The second one is the interpretability of the desired LRBSs that comprises two characteristics: the interpretability of the computational representation of the LFOCs under considerations and the complexity of their linguistic rule base measured by the total number of the rule conditions appearing in the rule base, denoted by “Comp” or sometimes by “#RFL,” understood as the product of the number of rules and the average of the rule lengths, e.g. see Alcalá et al. [2] and Antonelli et al. [8]. Since we will use strong trapezoid representation of the LFOCs produced by the procedure P_MG, it is advantageous that their full interpretability is ensured by Th. 4.1 and by six constraints on the fuzziness parameter values given in (5) to handle the compatibility of the trapezoid fuzzy sets with their words, where the upper and lower limits are specified by application user. So, the optimization objectives of the proposed MOEA are only MSE and Comp.

The proposed algorithm of the HA-approach composes two main tasks: - The first task comprises an action, which is a feature distinguished from the fuzzy set based approaches and aimed to determine the desired hedge algebras, including the LFOCs of the linguistic variables, and to produce the trapezoid representations of the LFOCs from the intermediate fuzziness parameters of their respective variables, and other actions to initialize elementary elements necessary for realization of the second task.

- The second task is to realize evolutionary steps to produce optimized solutions establishing a Pareto surface, each of which represent optimized independent fuzziness parameters and an optimized linguistic rule base.

Similarly as in the studies by Alcalá et al. [2] and Antonelli et al. [8], we apply the (2 + 2)M-PAES proposed by Cococcioni et al. [13], which differs from the original classical (2 + 2)PAES examined by Knowles and Corne [30] in using not only mutation, but also crossover to generate new candidate solutions stored in the archive.

Since the (2 + 2)M-PAES is familiar in the literature, the algorithm, named by EnHA-based PAES with trapezoid MG- representations of LFOCs (written briefly by EnHA-0-MG), whose MG-structure started with the 0-level as represented in Fig. 4, can briefly be described as follows:

Algorithm EnHA-0-MG

1. Initialization block: Construction of LFOCs, their trapezoid MG- representations and produce in turn two initial individuals 1 and 2 to be stored in the archive.

Step 1.1. To produce LFOCs and select the independent fuzziness parameters: For every variable X_j, j = 1,..., n + 1,
- Generate its LFOC \( \mathcal{F}_j = \mathcal{F}_j(\mathcal{X}) \) from the primary words \( c^-_j \) and \( c^+_j \) and the constants \( 0_j, W_j \) and \( 1_j \) by using hedges \( h_0, L_j \) and \( V_j \);
- Randomly select the components values of \( \pi_j = (\mu(h_0), \mu(L_j), f_m(0), f_m(c^-_j), f_m(W_j), f_m(1_j)) \) from their aforementioned corresponding prespecified interval-domains so that they satisfy (5), (fm1) and (fm4). In such a way, a \( C_{RB} \)- part is randomly generated;

Step 1.2. Compute semantic characteristics of the variables and construct MG- representations of the produced LFOCs:

For every \( \mathcal{X}_j, j = 1, ..., n + 1 \), do

Procedure Pr_MG(\( \pi_j, \mathcal{F}_j \))

Step 1.3. Generate linguistic rule base from individual numeric patterns: To generate the \( C_{RB} \) corresponding \( C_{RB} \)- part,
- Randomly select an integer \( M_k \) from the prespecified interval \([M_{min}, M_{max}]\);
- Applying the procedure Pr(d, \( \mathfrak{S}(\mathcal{F}_j): j \leq n + 1 \), \( \tau_{max} \)) to generate \( M_f \) linguistic rules from \( M_f \) patterns d’s that are randomly selected from a given dataset D.

Step 1.4. Evaluate the generated linguistic rule bases: Two aspects in this step need be computed for evaluation:
- Compute the MSE and the complexity Comp of the LRB represented by the part \( C_{RB} \) of the current individual.
- Compute the subset of the dataset D’ that cannot be covered by the LRB, i.e. the patterns of D’ that cannot fired by any rules of the LRB.

Step 1.5. Initialize archive: Execute Steps 1.1 to 1.4 two times to create two individuals 1 and 2 and add them to the archive which is viewed as an initial Pareto surface \( \mathfrak{B} \).

2. Evolutionary Block to generate optimized solutions stored in the archive representing a Pareto surface (repeated MaxGen times, where MaxGen denotes the prespecified number of generations)
**Step 2.1** Randomly select in turn two individuals from the archive interpreted as a current parent \( p_1 \) and \( p_2 \) and generate two offsprings from the parent by a procedure using genetic operators as follows:

2.1.1. Generate two intermediate offsprings \( \Phi, k = 1, 2 \), by applying in turn crossover operators working independently on their parts \( C_\mu \) and \( C_{RB} \):

- Apply BLX-\( \alpha \) crossover to \( C_\mu \) with a probability \( P_{C_\mu} \);
- Apply one-point crossover to \( C_{RB} \) with a probability \( P_{C_{RB}} \).

2.1.2. Generate two offsprings from the two current parents by applying the following mutation operators:

1) Apply a mutation operator on \( C_\mu \) with a probability \( P_{mC_\mu} \), which is executed by randomly selecting a gene in \( C_\mu \) and replaces its value with another one randomly selected from its defined interval-domain. Then, execute the procedure \( Pr(M_{Ad}(\pi, \Psi_{\mu})) \) to construct a semi-fully interpretable trapezoid MGr-representation of the LFoC \( \Psi_{\mu} \):

2) If the crossover on \( C_{RB} \) is not done, execute the following mutation procedure working on \( C_{RB} \), otherwise do it with probability \( P_{mC_{RB}} \):

When this procedure is applied, the two following mutation operators \( M_{Ad} \) and \( M_{Ch} \) are applied as follows:

- The operator \( M_{Ad} \) is executed as follows with a prespecified probability \( P_{Ad} \):
  - In the case the crossover has been applied to \( C_{RB} \) or the set \( D' = \emptyset \), where \( D' \) consists of the patterns that are not covered any current rules appearing in \( C_{RB} \), execute the \( M_{Ad} \) to add \( \gamma \) rules produced by applying the procedure \( Pr(d, \{S(\bar{D}; j \leq n + 1); \tau_{max}\}) \) where these \( d's \) are randomly selected from the dataset \( D \).
  - In the case the crossover has not been applied to \( C_{RB} \) and \( D' \neq \emptyset \), execute the \( M_{Ad} \) to add \( \gamma \) rules similarly as above, except that \( d's \) are randomly selected from the dataset \( D' \).

- When the operator \( M_{Ad} \) is not applied, the mutation operator \( M_{Ch} \) is always applied to change the gene values at \( \delta \) points representing rule conditions randomly selected in \( C_{RB} \).

2.1.3. Avoiding rule repetition: If there are several identical rules appearing in \( C_{RB} \), retain only one of them.

2.1.4. Eliminating the rules of length of zero appearing in \( C_{RB} \).

**Step 2.2.** Compute the objectives and, if possible, \( D' \) of the rule bases defined by the \( C_{RB} \)-part of the two obtained offsprings and store them into the archive so that the current solutions in the archive form a Pareto surface \( \mathcal{R} \).

---

**5. Simulation experiments and an analysis of the performance of the proposed method**

Following the purpose of the study, the section aims to show the benefits of the proposed trapezoid MGr-representation of LFoCs examined in Sections 3 and 4 in designing LRBSs by means of computer simulation. To meet this aim, we apply the same algorithm described just above, but using distinct MG-structures for representing LFoCs to analyze the meaning of certain semantic features of the words of LFoCs.

**5.1. Computer simulation design and simulation results**

Firstly, the MGr-representations considered in this section should be constructed to be different from each other only in two facts:

(i) Their 1-level fuzzy partition to represent the word-set \( \{0_j, c_j^-, W_j, c_j^+, \mathbf{1}_{1j}\} \) may be decomposed into two levels: a 0-level to represent \( \{0_j, W_j, \mathbf{1}_{1j}\} \) and a new 1-level to represent \( \{0_j, c_j^-, c_j^+, \mathbf{1}_{1j}\} \). The MGr-representations whose 1-level is not decomposed into the two levels as above is called MGr-representations started with the 1-level and denoted by 1-MGr-representations. In the opposite case, they are called MGr-representations started with the 0-level and denoted by 0-MGr-representations;

(ii) Their fuzzy sets are either all triangles or all trapezoids.

Secondly, the algorithms to design the desired genetic LRBSs should be developed in such a way that their simulation results can be sufficient to conclude that the following factors of the word semantics of the computational representation of LFoCs can actually enhance the performance of the designed LRBSs:

- The maintenance of the generality-specificity relation or, in practice of the recent study in this field, the increase of the generality of the word semantics in designing MGr-representations of LFoCs.
- The use of trapezoids instead of triangles in representing LFoCs.
- The use of interpretable trapezoid MGr-representations of LFoCs in designing of genetic LRBSs.

To solve these questions, it requires that the algorithms developed for the simulation should be as similar as possible so that no factors other than those used in representing the LFoCs may influence on the performance of their outputs. Thus, for simulation studies, we develop and examine the following four algorithms to produce genetic LRBSs or FRBSs to solve the same regression problems:
A1) OTzM: the EnHA-0-MG presented in the end of the above section using trapezoid 0-MGr-representations of LFoCs.
A2) 1TzM: a modification of EnHA-0-MG using trapezoid 1-MGr-representations.
A3) 0TgM: it is an EnHA-0-MG using triangles instead of trapezoids examined by Nguyen et al., [41].
A4) PSKB: the algorithm (2+2)M-PAES developed and examined by Alcalá et al. [2].

The algorithms A1) – A3) are identical ones except their MGr-representations of LFoCs. The algorithm PAESKB is taken into account, because the idea and general techniques underlying the development of the algorithms A1) – A3) are similar to it. That is, the main factor that may enhance the performance of their output FRBSs is only just the applying MOEAs to learn concurrently the fuzzy sets assigned to their current word-labels (DBs) and their rule bases (RBs) represented in the fuzzy set formalism. Techniques like clustering, decrease in the number of the dataset feature under consideration, or those for dealing with imbalanced datasets and so on, to improve the performance of the algorithms are still not applied.

In simulation, the performance of the first three algorithms A1 – A3 is examined based on their simulation experiment executed on twelve regression datasets given at

http://sci2s.ugr.es/keel/datasets.php

and described in Tab. 1-A, which indicates that the table is given in the Appendix. Since the algorithm PSKB examined by Alcalá et al. [2] is executed only on the first nine datasets of the table, the comparative study of the performance of the algorithms in question is performed only on these nine datasets. Similarly as the work by Alcalá et al. [2], the simulation is run by 5-fold cross-validation method, repeated 6 times, and, hence, it results in 30 runs for each dataset. For comparison, instead of considering all the solutions in the archive that form a Pareto surface of each run, similarly as in that work, to simplify the presentation we focus on the comparison of the experimental results at three specific points: the centroids of the FIRST (with the lowest MSE), the MEDIAN and the LAST (with the highest MSE) solutions of each Pareto surface.

The obtained simulation results of the run of all algorithms A1) – A4), using the parameter values given in Tab. 2.A are exposed in Tab. 3.A – 5.A, where the simulation results in the training phase are removed to reduce the tables.

The data in the columns headed by the notation PSKB are simulation results examined by Alcalá et al. [2] of the first nine datasets in the tables. The column with header of the form “Diff$$\alpha$$n$, for n = 1, …, 4, means that its data values are the difference of the respective data values in the column “Tz0M” and the corresponding data values in the column with header “A$$\alpha$$” indicating one of the algorithms A1 – A3.

5.2. Comparative studies to demonstrate the efficiency of interpretable MGr-representations of LFoCs

Theoretically, as discussed in Section 3, the interpretable MGr-representations of LFoCs represent or maintain much semantic information of the words of the variables in the sense that the interpretation mappings can convey formally discovered essential semantic characteristics of the words of the respective variables, e.g. the order-based semantics of words induced by their inherent semantics and, hence, the semantic characteristics of words formulated in terms of the semantics order relations of the respective variables and the generality-specificity relation of words. In particular, the mentioned semantic characteristics represented by these representations are invariant with respect to any extensions of the respective LFoCs. On the theoretical point of view, this is very essential for any investigation of the interpretations of a computational representation of LFoCs.

Practically, we show that they are not only essential theoretical requirements, but they are also important factors in designing LRBSs that actually enhance the designed LRBSs in terms of statistical tests. Similarly as in the paper by Alcalá et al. [2] and Nguyen et al. [41], the Wilcoxon signed-rank test, a non-parametric statistical hypothesis test, is applied using $$\alpha = 0.5$$, i.e. 95% confidence, and its results are arranged in Tab.’s 6 and 7.

These simulation experiments allow us to conclude some important facts about the meaning of the interpretable MGr-representations of LFoCs.

Firstly, it is observed that the two algorithms Tz0M and Tz1M are identical, except the fact that the trapezoid MGr-representations of the LFoCs applied in the algorithms are different from each other in only some features related to the 0- and 1-levels: by the decomposition of the 1-level partition of Tz1M representing the semantics of the words in $$\{0_0, c, W, c^-, 1\}$$, as shown in Fig. 6, into two level partitions of Tz0M, the one is the 0-level partition representing the words in $$\{0_0, W, 1\}$$ and the other is the new 1-level partition representing the words in $$\{0_0, c, W, c^-, 1\}$$, see Fig. 4. In this way, the fuzzy-set-based semantics of the words in $$\{0_0, c, W, c^-, 1\}$$ applied in Tz0M are of more generality than the one of the corresponding words in $$\{0_0, c, W, c^-, 1\}$$ applied in Tz1M and the fuzzy-set-based semantics of the remaining words are the same in the both algorithms. At the same time, it is
observed from the first row of the three points under consideration, FIRST, MEDIAN and LAST, exposed in Tab. 6A and 7A that while the complexity of the rule bases (RBs) of the LRBSs produced by the both algorithms are almost the same, the LRBSs produced by Tz0M at the FIRST point outperform the corresponding ones produced by Tz1M in terms of their accuracy with the confidence 95%. Since the accuracy of LRBSs attracts more user attention than the their complexity and the results of the LAST point are in general rather critical (see Tab. 5A), as this solution the accuracy of the LRBS at this point is the worst, whilst its complexity attains the maximum in comparison with other solutions in the archive. Therefore, though no noticeable conclusions can be deduced from the last remaining points, we can still state the following:

Cls1) Generality-specificity relation of words in representing LF0Cs is actually a considerable factor that can enhance the performance of the designed genetic LRBSs.

Secondly, from the statistic test results given in Tab. 6A and 7A, we can state that while the complexity of the RBs of the LRBSs produced by the both algorithms, Tz0M and Tg0M, are more or less equivalent, the Tz0M outperforms the Tg0M in terms of the accuracy at the FIRST and the MEDIAN points with the confidence 95%. Since the two algorithms Tz0M and Tg0M are identical, except the Tg0M applying triangle instead of trapezoid fuzzy sets in representing LF0Cs, though the null hypothesis is rejected at the LAST point, we can still draw the conclusion that:

Cls2) Applying trapezoids instead of triangles in representing LF0Cs is actually an important factor to enhance the performance of the designed genetic LRBSs or, in other words, the LRBSs produced by the Tz0M statistically outperform those produced by the Tg0M.

Thirdly, we are going to a more general question about the meaning of trapezoid MGr-representations of LRBSs, i.e. whether the proposed method to design LRBSs outperform non-HA-based methods. Noting that there are not many studies applying FRBSs to sole regression problems up to now as pointed out by Alcalá et al. [2] and by the reasons mentioned at the beginning of Section 5.1, the method proposed and examined in that paper is selected to compare with our proposed method. Similar as above, we observe that while the complexity of the RBs of the LRBSs produced by the both algorithms, Tz0M and PSKB, are equivalent, the Tz0M outperforms the PSKB in terms of the accuracy at all three examined points with the confidence 95%. So, we can strongly conclude that:

Cls3) Interpretable trapezoid MGr-representations of LF0Cs are actually an essential factor to enhance the performance of the designed genetic LRBSs in solving regression problems.

6. Conclusions

In this study, the interpretability of FRBSs is examined following the approach proposed first in [41] based on the standpoint stating that in simulating human capabilities in handing linguistic words with their own semantics, the inherent semantics of words must play a fundamental and essential role in formally designing fuzzy sets assigned to the words in question. As a consequence, such an approach requires firstly that the word-domains of variables must be formalized to become math-structures based on the inherent order-based semantics of words, noting that human words used in daily life do convey real world semantics. In accordance to our knowledge, it is difficult to study the actual interpretability of the designed FRBSs if one has no formal way to relates the fuzzy sets representing linguistic words with the own semantics of human words. Therefore, computational/fuzzy set representations of LF0Cs of variables are meaningful in representing the real world semantics conveyed by the LF0C words, only when their computational objects or fuzzy sets are designed in a strict formal

---

Figure 6. An trapezoidal MG-RP started with the 1-level of the LF0C Xε(2) of the 1st-feature of ELE-2 generated from the optimized independent fuzziness parameters.
relation with the words of the given LFoCs and with the semantic structures of just the LFoCs.

The proposed approach is developed based on the theory of hedge algebras which can be considered as immediate math-models of the word-domains of variables as word-domains are ordered by the inherent word semantics and their axioms are just selected key properties of words and hedges formulated in terms of the order relation of the respective word-domains. The order-based structures of word-domains and their quantification methodologies developed in an axiomatic way establish a formalized basis to construct

The study gives an additional explicit analysis of semantic basis of the approach proposed in [41] and proposed two additional constraints related to the semantic structure of LFoCs that seem to be very crucial and practical for constructing appropriate topological structures of arranging fuzzy sets representing the words of LFoCs to maintain the word semantics. It is shown that multi-granularity structures of fuzzy sets may provide suitable topological structures to arrange the designed fuzzy sets to represent the semantics of LFoCs to meet all proposed interpretability constraints and that the trapezoid fuzzy set semantics of words are statistically better than their triangle fuzzy set semantics. Note that tuning the trapezoid and triangle semantics of words can be performed by only tuning the few common fuzziness parameters of their respective whole variables instead of tuning individual fuzzy set parameters.

In general, it is obvious that this approach to the interpretability of FRBSs is distinguished from the existing approaches in literature and it offers many advantages, for instance:

- Providing a formalized foundation to immediately handle the words of variables and their computational semantics.
- Allowing to study semantics-based interpretability of FRBSs immediately based on words.
- There is no limit of the number of the words and their associated fuzzy sets in designing FRBSs.

References


Appendix

<table>
<thead>
<tr>
<th>No</th>
<th>Dataset name</th>
<th>#Pat</th>
<th>#Att</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Electrical Length 1 (ELE1)</td>
<td>495</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Electrical Maintenance 2 (ELE2)</td>
<td>1056</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>Weather Ankara (WA)</td>
<td>1609</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>Weather Izmir (WI)</td>
<td>1461</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>Treasury (TR)</td>
<td>1049</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>Abalone (AB)</td>
<td>4177</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>Mortgage (MTG)</td>
<td>1049</td>
<td>15</td>
</tr>
<tr>
<td>8</td>
<td>Computer Activity (CA)</td>
<td>8192</td>
<td>21</td>
</tr>
<tr>
<td>9</td>
<td>Pole Telecommunication (PT)</td>
<td>15000</td>
<td>26</td>
</tr>
<tr>
<td>10</td>
<td>Baseball (BB)</td>
<td>337</td>
<td>16</td>
</tr>
<tr>
<td>11</td>
<td>ForestFires (FF)</td>
<td>517</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>Stock (ST)</td>
<td>950</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 1. Regression problem datasets used in the study

<table>
<thead>
<tr>
<th>Parameter values with the notations:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{c_{\mu}}(P_{m_{C\mu}})$: crossover (mutation) probability on $C_{\mu}$</td>
</tr>
<tr>
<td>$P_{c_{BB}}(P_{m_{CBB}})$ crossover (mutation) probability on $C_{BB}$</td>
</tr>
<tr>
<td>$P_{add}$: mutation probability for adding rules</td>
</tr>
<tr>
<td>1) Fuzziness parameter constraints:</td>
</tr>
<tr>
<td>$\mu_{min} = fm_{max} = 0.3$, $\mu_{max} = fm_{max} = 0.7$, $fm_{0_{min}} = fm_{W_{max}} = fm_{1_{max}} = 0$; $fm_{0_{max}} = fm_{W_{max}} = fm_{1_{max}} = 0.1$, $\mu h_{0_{min}} = 0$, $\mu h_{0_{max}} = 0.2$</td>
</tr>
<tr>
<td>2) Probabilities of genetic operations:</td>
</tr>
<tr>
<td>$P_{C_{\mu}} = 0.3$, $P_{c_{BB}} = 0.5$, $P_{m_{C\mu}} = 0.1$, $P_{m_{CBB}} = 0.3$, $P_{add} = 0.75$, $\gamma_{max} = 5$, $\delta_{max} = 5$, $k_{max} = 3$, $\tau_{max} = 5$, $M_{min} = 5$, $M_{max} = 30$, $archSize = 64$, $MaxGen = 300,000$, $\alpha = 0.5$ for statistic tests</td>
</tr>
<tr>
<td>Dataset</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>ELE1</td>
</tr>
<tr>
<td>ELE2</td>
</tr>
<tr>
<td>WA</td>
</tr>
<tr>
<td>WI</td>
</tr>
<tr>
<td>TR</td>
</tr>
<tr>
<td>AB</td>
</tr>
<tr>
<td>MTG</td>
</tr>
<tr>
<td>CA</td>
</tr>
<tr>
<td>PT</td>
</tr>
<tr>
<td>BB</td>
</tr>
<tr>
<td>FF</td>
</tr>
<tr>
<td>SF</td>
</tr>
<tr>
<td>Mean MED</td>
</tr>
</tbody>
</table>

**Table 3.** Average results achieved by EnHA-Tz-0-MG (Tz0M), EnHA-Tz-1-MG (Tz1M), HA-Tg-1-MG (Tg1M) and PAESKB (KB) in correspondence to the FIRST point
### Table 4. Average results achieved by EnHA-Tz-0-MG (Tz0M), EnHA-Tz-1-MG (Tz1M), HA-Tg-1-MG (Tg1M) and PAES$_{KB}$ (KB) in correspondence to the MEDIAN point

<table>
<thead>
<tr>
<th>Dat asct</th>
<th>#R</th>
<th>Comp</th>
<th>MSE$_{kb}$</th>
<th>Tz0M</th>
<th>Tz1M</th>
<th>Diff$_{Tz}$</th>
<th>PS$_{kb}$</th>
<th>Tz0M</th>
<th>Tz1M</th>
<th>Diff$_{Tz}$</th>
<th>PS$_{kb}$</th>
<th>Tz0M</th>
<th>Tz1M</th>
<th>Diff$_{Tz}$</th>
<th>PS$_{kb}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELE1</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>212837.58</td>
<td>216069.41</td>
<td>-0.015</td>
<td>221608.03</td>
<td>-0.040</td>
<td>234179.00</td>
<td>-0.091</td>
</tr>
<tr>
<td>ELE2</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>86051.73</td>
<td>121108.61</td>
<td>-0.289</td>
<td>100666.24</td>
<td>-0.145</td>
<td>136880.00</td>
<td>-0.371</td>
</tr>
<tr>
<td>WA</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>3.68</td>
<td>3.82</td>
<td>-0.036</td>
<td>5.24</td>
<td>-0.297</td>
<td>24.88</td>
<td>-0.852</td>
</tr>
<tr>
<td>WI</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>3.15</td>
<td>2.91</td>
<td>0.077</td>
<td>3.98</td>
<td>-0.209</td>
<td>8.77</td>
<td>-0.641</td>
</tr>
<tr>
<td>TR</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>0.14</td>
<td>0.19</td>
<td>-0.262</td>
<td>0.14</td>
<td>-0.008</td>
<td>0.42</td>
<td>-0.673</td>
</tr>
<tr>
<td>AB</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>3.20</td>
<td>3.17</td>
<td>0.009</td>
<td>3.10</td>
<td>0.030</td>
<td>3.13</td>
<td>0.021</td>
</tr>
<tr>
<td>MTG</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>0.08</td>
<td>0.11</td>
<td>-0.219</td>
<td>0.13</td>
<td>-0.348</td>
<td>0.27</td>
<td>-0.692</td>
</tr>
<tr>
<td>CA</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>17.29</td>
<td>23.02</td>
<td>-0.249</td>
<td>20.45</td>
<td>-0.155</td>
<td>33.44</td>
<td>-0.483</td>
</tr>
<tr>
<td>PT</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>391.69</td>
<td>376.68</td>
<td>0.038</td>
<td>342.32</td>
<td>0.126</td>
<td>651.00</td>
<td>-0.398</td>
</tr>
<tr>
<td>BB</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>319930.66</td>
<td>341688.00</td>
<td>-0.064</td>
<td>430565.19</td>
<td>-0.257</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FF</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>2109.80</td>
<td>2157.80</td>
<td>-0.022</td>
<td>2129.39</td>
<td>-0.009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST</td>
<td>5.03</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>6.70</td>
<td>6.83</td>
<td>-0.018</td>
<td>5.47</td>
<td>0.184</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>-0.088</td>
<td>-0.094</td>
<td>0.464</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 5. Average results achieved by EnHA-Tz-0-MG (Tz0M), EnHA-Tz-1-MG (Tz1M), HA-Tg-1-MG (Tg1M) and PAES$_{KB}$ (KB) in correspondence to the LAST point

| EnHA-Tz-0-MG | VS | R$^+$ | R$^-$ | Exact P-value | Confid.-interval | Exact-confid. Hypoth. (H$_0$) | EnHA-Tz-1-MG | VS | R$^+$ | R$^-$ | Exact P-value | Confid.-interval | Exact-confid. Hypoth. (H$_0$) |
|--------------|----|------|------|--------------|-----------------|-----------------|--------------|----|------|------|--------------|-----------------|-----------------|--------------|----|------|------|--------------|-----------------|-----------------|
| The FIRST point | EnHA-Tz-1-MG | 51.0 | 27.0 | $\geq 0.2$ | [-0.9400, 03.200] | 0.95752 | Not rejected | HA-Tg-0-MG | 28.0 | 50.0 | $\geq 0.2$ | [-11.600, 19.285] | 0.95752 | Not rejected | PAES$_{KB}$ (9 datasets) | 34.0 | 11.0 | $\geq 0.2$ | [-23.200, 13.365] | 0.96094 | Not rejected |
| The MEDIAN point | EnHA-Tz-1-MG | 52.0 | 26.0 | $\geq 0.2$ | [-03.200, 01.080] | 0.95752 | Not rejected | HA-Tg-0-MG | 38.0 | 40.0 | $\geq 0.2$ | [-04.915, 06.680] | 0.95752 | Not rejected | PAES$_{KB}$ (9 datasets) | 30.0 | 15.0 | $\geq 0.2$ | [-08.235, 06.080] | 0.96094 | Not rejected |
| The LAST point | EnHA-Tz-1-MG | 35.5 | 30.5 | $\geq 0.2$ | [-00.035, 00.035] | 0.95752 | Not rejected | HA-Tg-0-MG | 38.0 | 40.0 | $\geq 0.2$ | [-00.015, 00.035] | 0.95752 | Not rejected | PAES$_{KB}$ (9 datasets) | 18.0 | 27.0 | $\geq 0.2$ | [0.0000, 00.035] | 0.96094 | Not rejected |

### Table 6. Comparison of fuzzy rule base complexity using the Wilcoxon test at level $\alpha = 0.05$ for three representative points

<p>| EnHA-Tz-0-MG | VS | R$^+$ | R$^-$ | Exact P-value | Confid.-interval | Exact-confid. Hypoth. (H$_0$) |
|--------------|----|------|------|--------------|-----------------|-----------------|--------------|----|------|------|--------------|-----------------|-----------------|--------------|----|------|------|--------------|-----------------|-----------------|
| The FIRST point | EnHA-Tz-1-MG | 60.0 | 6.0 | 0.013672 | [0.0085, 0.1585] | 0.95752 | Rejected |</p>
<table>
<thead>
<tr>
<th>Point Type</th>
<th>Method</th>
<th>MSE</th>
<th>CI</th>
<th>p-value</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>The MEDIAN point</td>
<td>HA-Tg-0-MG</td>
<td>67.5</td>
<td>10.5</td>
<td>0.023930</td>
<td>[0.0025, 0.0655]</td>
</tr>
<tr>
<td></td>
<td>PAES_{kb} (9 datasets)</td>
<td>43.0</td>
<td>2.0</td>
<td>0.011718</td>
<td>[0.1070, 0.6980]</td>
</tr>
<tr>
<td></td>
<td>EnHA-Tz-1-MG</td>
<td>61.5</td>
<td>16.5</td>
<td>0.084710</td>
<td>[-0.0015, 0.0970]</td>
</tr>
<tr>
<td></td>
<td>HA-Tg-0-MG</td>
<td>57.0</td>
<td>9.0</td>
<td>0.032220</td>
<td>[0.0020, 0.1060]</td>
</tr>
<tr>
<td></td>
<td>PAES_{kb} (9 datasets)</td>
<td>43.0</td>
<td>2.0</td>
<td>0.011718</td>
<td>[0.1070, 0.6980]</td>
</tr>
<tr>
<td>The LAST point</td>
<td>EnHA-Tz-1-MG</td>
<td>63.0</td>
<td>15.0</td>
<td>0.063960</td>
<td>[-0.0060, 0.1630]</td>
</tr>
<tr>
<td></td>
<td>HA-Tg-0-MG</td>
<td>53.0</td>
<td>13.0</td>
<td>0.083000</td>
<td>[-0.0145, 0.2125]</td>
</tr>
<tr>
<td></td>
<td>PAES_{kb} (9 datasets)</td>
<td>44.0</td>
<td>1.0</td>
<td>0.007812</td>
<td>[0.2310, 0.6825]</td>
</tr>
</tbody>
</table>

Table 7. Comparison MSE_{kb} using the Wilcoxon test at level α = 0.05 for three representative points.