

*Journal of Computer Science and Cybernetics, V.35, N.4 (2019), 319–336*

DOI 10.15625/1813-9663/35/4/14348

## A HEDGE ALGEBRAS BASED CLASSIFICATION REASONING METHOD WITH MULTI-GRANULARITY FUZZY PARTITIONING

PHAM DINH PHONG<sup>1,\*</sup>, NGUYEN DUC DU<sup>1</sup>, NGUYEN THANH THUY<sup>2</sup>,  
HOANG VAN THONG<sup>1</sup>

<sup>1</sup>*Faculty of Information Technology, University of Transport and Communications, Hanoi, Vietnam*

<sup>2</sup>*Faculty of Information Technology, University of Engineering and Technology, VNU, Hanoi, Vietnam*

\**dinhphongpham@gmail.com*



**Abstract.** During last years, lots of the fuzzy rule based classifier (FRBC) design methods have been proposed to improve the classification accuracy and the interpretability of the proposed classification models. In view of that trend, genetic design methods of linguistic terms along with their (triangular and trapezoidal) fuzzy sets based semantics for FRBCs, using hedge algebras as the mathematical formalism, have been proposed. Those hedge algebras based design methods utilize semantically quantifying mapping values of linguistic terms to generate their fuzzy sets based semantics so as to make use of the existing fuzzy sets based classification reasoning methods for data classification. If there exists a classification reasoning method which bases merely on the semantic parameters of hedge algebras, fuzzy sets based semantics of the linguistic terms in the fuzzy classification rule bases can be replaced by hedge algebras-based semantics. This paper presents a FRBC design method based on hedge algebras approach by introducing a hedge algebra based classification reasoning method with multi-granularity fuzzy partitioning for data classification so that the semantics of linguistic terms in the rule bases can be hedge algebras-based semantics. Experimental results over 17 real world datasets are compared to the existing methods based on hedge algebras and the state-of-the-art fuzzy set theory-based approaches, showing that the proposed FRBC in this paper is an effective classifier and produces good results.

**Keywords.** Classification Reasoning; Fuzzy Rule Based Classifier; Fuzziness Interval; Hedge Algebras; Multi-Granularity; Semantically Quantifying Mapping Values.

### 1. INTRODUCTION

Fuzzy rule based systems (FRBSs) have been studied and applied efficiently in many different fields such as fuzzy control, data mining, etc. Unlike classical classifiers based on the statistical and probabilistic approaches [3, 8, 27, 32] which are the “black boxes” lacking of interpretability, the advantage of the FRBC model is that end-users can use the high interpretability fuzzy rule-based knowledge extracted automatically from data as their knowledge.

In the FRBC design based on the fuzzy set theory approaches [1, 2, 6, 7, 21, 22, 23, 24, 35, 36, 38, 39, 41], the fuzzy partitions from which fuzzy rules are extracted are commonly pre-designed using fuzzy sets and then linguistic terms are intuitively assigned to

fuzzy sets. Furthermore, fuzzy partitions can be generated automatically from data by using discretization or granular computing mechanisms [37]. No matter how they are designed, the problem of the linguistic term design is not clearly studied although fuzzy rule bases are represented by linguistic terms with their fuzzy set based semantics. Many techniques have been proposed to achieve compact fuzzy rule systems with accuracy and interpretability trade-off extracted from data, such as using artificial neural network [33] or genetic algorithm [1, 2, 7, 21, 36, 38, 39, 41] by adjusting fuzzy set parameters to achieve the optimal fuzzy partitions and to select the optimal fuzzy rule based systems. However, the fuzzy set based semantics of linguistic terms are not preserved, leading to the affectedness of the interpretability of the fuzzy rule bases of classifiers.

Hedge algebras (HAs) [9, 11, 12, 14, 17, 18] provide a mathematical formalism for designing the order based semantic structure of term domains of linguistic variables that can be applied to various application domains in the real life, such as fuzzy control [10, 26, 28, 29], expert systems [12], data mining [5, 13, 15, 16, 25, 40], fuzzy database [19, 42], image processing [20], timetabling [31], etc. The crucial idea of the hedge algebra based approach is that it reflects the nature of fuzzy information by the fuzziness of information. In [13, 15], HAs are utilized to model and design the linguistic terms for FRBCs. They exploit the inherent semantic order of linguistic terms that allows generating semantic constraints between linguistic terms and their integrated fuzzy sets. More specifically, when given values of fuzziness parameters, the semantically quantifying mapping (SQM) values of linguistic terms are computed and then associated fuzzy sets of linguistic terms are automatically generated from their own semantics. So, linguistic terms along with their fuzzy sets based semantics are generated by a procedure. Based on this formalism, an efficient fuzzy rule based classifier design method is developed.

As set forth above, HAs can be utilized to design eminent FRBCs. However, we may wonder that why the semantics of linguistic terms in the fuzzy classification rule bases of FRBCs designed by the HAs based methodology are still fuzzy sets based semantics. The answer is that although linguistic terms are designed by HAs, the fuzzy set based classification reasoning methods proposed in the prior researches [21, 23, 24] are made use for data classification. If there is a classification reasoning method for data classification which bases merely on semantic parameters of hedge algebras, fuzzy sets based semantics of linguistic terms in the fuzzy classification rule bases can be replaced with hedge algebras based semantics. In response to that question, a classification reasoning method merely based on HAs for FRBC is presented in this paper. The idea is based on the Takagi-Sugeno-Hedge algebras fuzzy model proposed in [26] to improve the forecast control based on models in such a way that membership functions of individual linguistic terms in Takagi-Sugeno fuzzy model are replaced with the closeness of semantically quantifying mapping values of adjacent linguistic values. That result is enhanced to build a classification reasoning method based on HAs which enables fuzzy sets based semantics of the linguistic terms in the fuzzy rule bases to be replaced with hedge algebras based semantics. Furthermore, the design of information granules plays an important role in designing FLRBCs, i.e., it is the basis for generating interpretable FLRBCs and impacts on the classification performance. Because of the semantic inheritance, with linguistic terms that are induced from the same primary term, the shorter the term, the more generality it has and vice versa. Therefore, with the single-granularity structure, all linguistic terms just appear in a fuzzy partition leading

to the semantics of shorter terms are reduced and become more specific. Contrarily, the multi-granularity structure retains the generality of shorter linguistic terms in the rule bases because linguistic terms which have the same length form a fuzzy partition. That is why a hedge algebra based classification reasoning method with multi-granularity fuzzy partitioning for data classification is introduced in this paper. Experimental results over 17 real world datasets show the efficiency of the multi-granularity structure design in comparison with the single one as well as show the efficiency of the proposed classifier in comparison with the state-of-the-art methods based on hedge algebras and fuzzy set theory.

The rest of the paper is organized as follows: Section 2 presents fuzzy rule based classifier design based on hedge algebras and the proposed hedge algebras based classification reasoning method for the FRBCs. Section 3 presents experimental evaluation studies and discussions. Conclusions and remarks are included in Section 4.

## 2. FUZZY RULE BASED CLASSIFIER DESIGN BASED ON HEDGE ALGEBRAS

### 2.1. Hedge algebras for the semantic representation of linguistic terms

To formalize the nature structure of the linguistic variables, a mathematic structure, so-called the hedge algebra, has been introduced and examined by N. C. Ho et al. [17, 18]. Assume that  $\mathcal{X}$  is a linguistic variable and the linguistic value domain of  $\mathcal{X}$  is  $Dom(\mathcal{X})$ . A hedge algebra  $\mathcal{AX}$  of  $\mathcal{X}$  is a structure  $\mathcal{AX} = (X, G, C, H, \leq)$ , where  $X$  is a set of linguistic terms of  $\mathcal{X}$  and  $X \subseteq Dom(\mathcal{X})$ ;  $G$  is a set of two generator terms  $c^-$  and  $c^+$ , where  $c^-$  is the negative primary term,  $c^+$  is the positive primary term and  $c^- \leq c^+$ ;  $C$  is a set of term constants,  $C = \{\mathbf{0}, W, \mathbf{1}\}$ , satisfying the relation order  $\mathbf{0} \leq c^- \leq W \leq c^+ \leq \mathbf{1}$ ;  $\mathbf{0}$  and  $\mathbf{1}$  are the least and the greatest terms, respectively;  $W$  is the neutral term;  $H$  is a set of hedges of  $X$ , where  $H = H^- \cup H^+$ ,  $H^-$  and  $H^+$  are the set of negative and positive hedges, respectively;  $\leq$  is an order relation induced by the inherent semantics of terms of  $X$ .

When a hedge acts on a non-constant term, a new linguistic term is induced. Each linguistic term  $x$  in  $X$  is represented as the string representation, i.e., either  $x = c$  or  $x = h_m \dots h_1 c$ , where  $c \in \{c^-, c^+\} \cup C$  and  $h_j \in H$ ,  $j = 1, \dots, m$ . All linguistic terms generated from  $x$  by using the hedges in  $H$  can be abbreviated as  $H(x)$ . If all linguistic terms in  $X$  and all hedges in  $H$  have a linear order relation, respectively,  $\mathcal{AX}$  is the linear hedge algebras.  $\mathcal{AX}$  is built from some characteristics of the inherent semantics of linguistic terms which are expressed by the semantic order relationship “ $\leq$ ” of  $X$ .

Two primary terms  $c^-$  and  $c^+$  possess their own converse semantic tendencies. For convenience,  $c^+$  possesses the positive tendency and it has positive sign written as  $\text{sign}(c^+) = +1$ . Similarly,  $c^-$  possesses the negative tendency and it has negative sign written as  $\text{sign}(c^-) = -1$ . As the semantic order relationship, we have  $c^- \leq c^+$ . For example, “old” possesses the positive tendency, “young” possesses the negative tendency and “young”  $\leq$  “old”.

Each hedge possesses tendency to decrease or increase the semantics of two primary terms. For example, “very young”  $\leq$  “young” and “old”  $\leq$  “very old”, the hedge *very* makes the semantics of “young” and “old” increased. “young”  $\leq$  “less young” and “less old”  $\leq$  “old”, the hedge *less* makes the semantics of “young” and “old” decreased. It is said that *very* is the positive hedge and *less* is the negative hedge. We denote the  $H^- = \{h_{-q},$

$\dots, h_{-1}\}$  is a set of negative hedges where  $h_{-q} \leq \dots \leq h_{-2} \leq h_{-1}$ ,  $H^+ = \{h_1, \dots, h_p\}$  is a set of positive hedges where  $h_1 \leq h_2 \leq \dots \leq h_p$  and  $H = H^- \cup H^+$ . If  $h \in H^-$ ,  $\text{sign}(h) = -1$  and if  $h \in H^+$ ,  $\text{sign}(h) = +1$ . If both hedges  $h$  and  $k$  in  $H^-$  or  $H^+$ , we say that  $h$  and  $k$  are compatible, whereas,  $h$  and  $k$  are inverse each other.

Each hedge possesses tendency to decrease or increase the semantics of other hedge. If  $k$  makes the semantic of  $h$  increased,  $k$  is positive with respect to  $h$ , whereas, if  $k$  makes the semantic of  $h$  decreased,  $k$  is negative with respect to  $h$ . The negativity and positivity of hedges do not depend on the linguistic terms on which they act. For example,  $V$  is positive with respect to  $L$ , we have  $x \leq Lx$  then  $Lx \leq VLx$ , or  $Lx \leq x$  then  $VLx \leq Lx$ . One hedge may have a relative sign with respect to another.  $\text{sign}(k, h) = +1$  if  $k$  strengthens the effect tendency of  $h$ , whereas,  $\text{sign}(k, h) = -1$  if  $k$  weakens the effect tendency of  $h$ . Thus, the sign of term  $x$ ,  $x = h_m h_{m-1} \dots h_2 h_1 c$ , is defined by

$$\text{sign}(x) = \text{sign}(h_m, h_{m-1}) \times \dots \times \text{sign}(h_2, h_1) \times \text{sign}(h_1) \times \text{sign}(c).$$

The meaning of the sign of term is that  $\text{sign}(hx) = +1 \rightarrow x \leq hx$  and  $\text{sign}(hx) = -1 \rightarrow hx \leq x$ .

Semantic inheritance in generating linguistic terms by using hedges: When a new linguistic term  $hx$  is generated from a linguistic term  $x$  by using the hedge  $h$ , the semantic of the new linguistic term is changed but it still conveys the original semantic of  $x$ . This means that the semantic of  $hx$  is inherited from  $x$ .

As set forth above, HAs are the qualitative models. Therefore, to apply HAs to solve the real world problems, some characteristics of HAs need to be characterized by quantitative concepts based on qualitative term semantics.

On the semantic aspect,  $H(x)$ ,  $x \in X$ , is the set of linguistic terms generated from  $x$  and their semantics are changed by using the hedges in  $H$  but still convey the original semantic of  $x$ . So,  $H(x)$  reflects the fuzziness of  $x$  and the length of  $H(x)$  can be used to express the *fuzziness measure* of  $x$ , denoted by  $fm(x)$ . When  $H(x)$  is mapped to an interval in  $[0, 1]$  following the order structure of  $X$  by a mapping  $v$ , it is called the *fuzziness interval* of  $x$  and denoted by  $\mathfrak{F}(x)$ .

A function  $fm: X \rightarrow [0, 1]$  is said to be a *fuzziness measure* of  $\mathcal{AX}$  provided that it satisfies the following properties:

$$(FM1) \quad fm(c^-) + fm(c^+) = 1 \text{ and } \sum_{h \in H} fm(hu) = fm(u) \text{ for } \forall u \in X;$$

$$(FM2) \quad fm(x) = 0 \text{ for all } H(x) = x, \text{ especially, } fm(\mathbf{0}) = fm(W) = fm(\mathbf{1}) = 0;$$

$$(FM3) \quad \forall x, y \in X, \forall h \in H, \text{ the proportion } \frac{fm(hx)}{fm(x)} = \frac{fm(hy)}{fm(y)} \text{ which does not depend on any particular linguistic term on } X \text{ is called the fuzziness measure of the hedge } h, \text{ denoted by } \mu(h).$$

From (FM1) and (FM3), the fuzziness measure of linguistic term  $x = h_m \dots h_1 c$  can be computed recursively that  $fm(x) = \mu(h_m) \dots \mu(h_1) fm(c)$ , where  $\sum_{h \in H} \mu(h) = 1$  and  $c \in \{c^-, c^+\}$ .

Semantically quantifying mappings (SQMs): The semantically quantifying mapping of  $\mathcal{AX}$  is a mapping  $v: X \rightarrow [0, 1]$  which satisfies the following conditions:

(SQM1) It preserves the order based structure of  $X$ , i.e.,  $x \leq y \rightarrow v(x) \leq v(y), \forall x \in X$ ;

(SQM2) It is one-to-one mapping and  $v(x)$  is dense in  $[0, 1]$ .

Let  $fm$  be a fuzziness measure on  $X$ .  $v(x)$  is computed recursively based on  $fm$  as follows:

1.  $v(W) = \theta = fm(c^-)$ ,  $v(c^-) = \theta - \alpha fm(c^-) = \beta fm(c^-)$ ,  $v(c^+) = \theta + \alpha fm(c^+)$ ;
2.  $v(h_j x) = v(x) + \text{sign}(h_j x) \left( \sum_{i=\text{sign}(j)}^j fm(h_i x) - \omega(h_j x) fm(h_j x) \right)$ ,

where  $j \in [-q, p] = \{j: -q \leq j \leq p, j \neq 0\}$  and

$$\omega(h_j x) = \frac{1}{2} [1 + \text{sign}(h_j x) \text{sign}(h_p h_j x) (\beta - \alpha)] \in \{\alpha, \beta\}.$$

## 2.2. Fuzzy rule based classifier design based on hedge algebras

A fuzzy rule based classifier design problem  $\mathcal{P}$  is defined as: A set  $\mathbf{P} = \{(\mathbf{d}_p, C_p) \mid \mathbf{d}_p \in \mathbf{D}, C_p \in \mathbf{C}, p = 1, \dots, m\}$  of  $m$  patterns, where  $\mathbf{d}_p = [d_{p,1}, d_{p,2}, \dots, d_{p,n}]$  is the row  $p^{th}$ ,  $\mathbf{C} = \{C_s \mid s = 1, \dots, M\}$  is the set of  $M$  class labels,  $n$  is the number of features of the dataset  $\mathbf{P}$ .

The fuzzy rule based system of the FRBCs used in this paper is the set of weighted fuzzy rules in the following form [21, 23, 24]

$$\text{Rule } R_q : \text{IF } X_1 \text{ is } A_{q,1} \text{ AND } \dots \text{ AND } X_n \text{ is } A_{q,n} \text{ THEN } C_q \text{ with } CF_q, \text{ for } q=1, \dots, N, \quad (1)$$

where  $X = \{X_j, j = 1, \dots, n\}$  is the set of  $n$  linguistic variables corresponding to  $n$  features of the dataset  $\mathbf{P}$ ,  $A_{q,j}$  is the linguistic terms of the  $j^{th}$  feature  $F_j$ ,  $C_q$  is a class label and  $CF_q$  is the rule weight of  $R_q$ . The rule  $R_q$  is abbreviated as the following short form

$$\mathbf{A}_q \Rightarrow C_q \text{ with } CF_q, \text{ for } q=1, \dots, N, \quad (2)$$

where  $\mathbf{A}_q$  is the antecedent part of the  $q^{th}$ -rule.

Solving the problem  $\mathcal{P}$  is to extract from  $\mathbf{P}$  a set  $\mathbf{S}$  of fuzzy rules in the form (1) in order to achieve a compact FRBC based on  $\mathbf{S}$  comes with high classification accuracy and suitable interpretability. The general method of FRBC design with the semantics of linguistic terms based on the hedge algebras comprises two following phases [15, 16]:

1. Genetically design linguistic terms along with their fuzzy-set-based semantics for each feature of the designated dataset in such a way that only semantic parameter values are adjusted, as a result, near optimal semantic parameter values are achieved by the interaction between semantics of linguistic terms and the data.
2. An evolutionary algorithm is applied to select near optimal fuzzy classification rule based systems having a quite suitable interpretability-accuracy trade-offs from data by using a given near optimal semantic parameter values provided by the first phase for fuzzy rule based classifiers.

HAs provides a formalism basis for generating quantitative semantics of linguistic terms from their qualitative semantics. This formalism is applied to genetically design linguistic terms along with the integrated fuzzy set based semantics for fuzzy rule based classifiers. Hereafter are the summaries of two above steps:

Each feature  $j^{th}$  of the designated dataset is associated with an hedge algebra  $\mathcal{AX}_j$ , induces all linguistic terms  $X_{j,(k_j)}$  with the maximum length  $k_j$  having the order based inherent semantics of linguistic terms. Given a value of the semantic parameters  $\Pi$ , which includes fuzziness measures  $fm(c_j^-)$  and  $\mu(h_{j,i})$  of the negative primary term  $c_j^-$  and  $h_{j,i}$ , respectively, and a positive integer  $k_j$  for limiting the designed term lengths, quantifying mapping values  $v(x_{j,i}), x_{j,i} \in X_{j,k}$  for all  $k \leq k_j$  and the  $k_j$ -similarity intervals  $\mathfrak{S}_{k_j}(X_{j,i})$  of linguistic terms in  $X_{j,k_j+2}$  are computed and they constitute a unique fuzzy partition of the  $j^{th}$  attribute. After fuzzy partitions of all attributes are constructed, fuzzy rule conditions will be specified based on these partitions.

Among the  $k_j$ -similarity intervals of a given fuzzy partition, there is a unique interval  $\mathfrak{S}_{k_j}(x_{j,i(i)})$  containing  $j^{th}$ -component  $d_{p,j}$  of  $d_p$  pattern. All  $k_j$ -similarity intervals which contain  $d_{p,j}$  component define a hyper-cube  $\mathcal{H}_p$ , and fuzzy rules are only induced from this type of hyper-cube. A fuzzy rule generated from  $\mathcal{H}_p$  for the class  $C_p$  of  $d_p$  is so-called a *basic fuzzy rule* and it has the following form

$$\text{IF } X_1 \text{ is } x_{1,i(1)} \text{ AND } \dots \text{ AND } X_n \text{ is } x_{n,i(n)} \text{ THEN } C_p. \quad (R_b)$$

Only one basic fuzzy rule which has the length  $n$  can be generated from the data pattern  $d_p$ . To generate the fuzzy rule with the length  $L \leq n$ , so-called the *secondary rules*, some techniques should be used for generating fuzzy combinations, for example, generate all  $k$ -combinations ( $1 \leq k \leq L$ ) from the given set of  $n$  features of dataset  $\mathcal{P}$ .

$$\text{IF } X_{j_1} \text{ is } x_{j_1,i(j_1)} \text{ AND } \dots \text{ AND } X_{j_t} \text{ is } x_{j_t,i(j_t)} \text{ THEN } C_q, \quad (R_{snd})$$

where  $1 \leq j_1 \leq \dots \leq j_t \leq n$ . The consequence class  $C_q$  of the rule  $R_q$  is determined by the confidence measure  $c(\mathbf{A}_q \Rightarrow C_h)$  [20, 21] of  $R_q$

$$C_q = \text{argmax}(c(\mathbf{A}_q \Rightarrow C_h) \mid h = 1, \dots, M). \quad (3)$$

The confidence of a fuzzy rule is computed as

$$c(\mathbf{A}_q \Rightarrow C_h) = \sum_{d_p \in C_h} \mu_{\mathbf{A}_q}(d_p) / \sum_{p=1}^m \mu_{\mathbf{A}_q}(d_p), \quad (4)$$

where  $\mu_{\mathbf{A}_q}(d_p)$  is the compatibility grade of the pattern  $d_p$  with the antecedent of the rule  $R_q$  and commonly computed as

$$\mu_{\mathbf{A}_q}(d_p) = \prod_{j=1}^n \mu_{q,j}(d_{p,j}). \quad (5)$$

As trying to generate all possible combinations, the maximum of number fuzzy combinations is  $\sum_{i=1}^L C_n^i$ , so the maximum of the *secondary rules* is  $m \times \sum_{i=1}^L C_n^i$ .

To eliminate less important rules, a screening criterion is used to select a subset  $\mathcal{S}_0$  with  $NR_0$  fuzzy rules from the candidate rule set, called an *initial fuzzy rule set*. Candidate rules are divided into  $M$  groups, sort rules in each group by a screening criterion. Select from each group  $NB_0$  rules, so the number of initial fuzzy rules is  $NR_0 = NB_0 \times M$ . The screening criterion can be the confidence  $c$ , the support  $s$  or  $c \times s$ . The confidence is computed by the formula (4), the support is computed as following formula [20]

$$s(\mathbf{A}_q \Rightarrow C_h) = \sum_{d_p \in C_h} \mu_{\mathbf{A}_q}(d_p) / m. \quad (6)$$

To improve the accuracy of classifiers, each fuzzy rule is assigned a rule weight and it is commonly computed by the following formula [20]

$$CF_q = c(\mathbf{A}_q \Rightarrow C_q) - c_{q,2nd}, \quad (7)$$

where  $c_{q,2nd}$  is computed as

$$c_{q,2nd} = \max(c(\mathbf{A}_q \Rightarrow \text{Class } h) \mid h = 1, \dots, M; h \neq C_q). \quad (8)$$

The classification reasoning method commonly used to classify the data pattern  $d_p$  is Single Winner Rule (SWR). The winner rule  $R_w \in \mathcal{S}$  (a classification rule set) is the rule which has the maximum of the product of the compatibility grade  $\mu_{\mathbf{A}_q}(d_p)$  and the rule weight  $CF(\mathbf{A}_q \Rightarrow C_q)$ , and the classified class  $C_w$  is the consequence part of this rule.

$$\mu_{\mathbf{A}_w}(d_p) \times CF_w = \operatorname{argmax} (\mu_{\mathbf{A}_q}(d_p) \times CF_q \mid R_q \in \mathcal{S}). \quad (9)$$

This fuzzy rule generation process is called the initial fuzzy rule set generation procedure **IFRG**( $\Pi, \mathbf{P}, NR_0, L$ ) [15], where  $\Pi$  is a set of semantic parameter values and  $L$  is the maximum of rule length.

Each specific dataset needs a different set of semantic parameter values to adapt to the data distribution of it, i.e., the quality of the classifier is improved. Thus, an evolutionary algorithm is needed to find optimal semantic parameter values for a specific dataset. When having optimal semantic parameter values, they are used to extract an *initial fuzzy rule set* and an evolutionary algorithm used to find a subset of the fuzzy classification rules  $\mathcal{S}$  from  $\mathcal{S}_0$  having a suitable interpretability–accuracy trade-offs for FRBCs.

### 2.3. Hedge algebras based reasoning method for fuzzy rule based classifier

Up to now, fuzzy rule based classifier design methods, using the hedge algebra methodology [13, 15] induce fuzzy sets based semantics of linguistic terms for FRBCs because the authors would like to make use of the fuzzy set based classification reasoning method proposed in the fuzzy set based approaches [21, 23, 24]. This research aims at proposing hedge algebras based classification reasoning method with multi-granularity fuzzy partitioning for FRBCs and shows the efficiency of the proposed ones by the experiments on a considerable real world datasets.

In [26], the authors propose a Takagi-Sugeno-Hedge algebra fuzzy model to improve the forecast control based on the models by using the closeness of semantically quantifying mapping values of adjacent linguistic terms instead of the grade of the membership function of each individual linguistic term. That idea is summarized as follows:

- $v(x_i)$ ,  $v(x_0)$  and  $v(x_k)$  are the SQM values of the linguistic terms  $x_i$ ,  $x_0$  and  $x_k$  with the semantic order  $x_i \leq x_0 \leq x_k$ , respectively.
- $\eta_i$  which is the closeness of  $v(x_i)$  to  $v(x_0)$  is defined as  $\eta_i = \frac{(v(x_k) - v(x_0))}{(v(x_k) - v(x_i))}$  and  $\eta_k$  which is the closeness of  $v(x_k)$  to  $v(x_0)$  is defined as  $\eta_k = \frac{(v(x_0) - v(x_i))}{(v(x_k) - v(x_i))}$ , where  $\eta_i + \eta_k = 1$  and  $0 \leq \eta_i, \eta_k \leq 1$ .

That idea is advanced to apply to make the hedge algebra based classification reasoning methods for FRBCs in two cases as follows.

#### ***In case of single granularity structure***

In the single granularity structure design, all linguistic terms  $X_{(k_j)}$  with different term length  $k$  ( $1 \leq k \leq k_j$ ) appear at the same level  $k_j$ . Therefore, at the level  $k_j$  of the  $j^{th}$ -feature of the designated dataset, there are the SQM values of all linguistic terms  $X_{(k_j)}$  with the semantic order  $v(x_{j,i-1}) \leq v(x_{j,i}) \leq v(x_{j,i+1})$ ,  $x_{j,i} \in X_{(k_j)}$ . For a data point  $d_{p,j}$  of the data pattern  $d_p$  (has been normalized to  $[0, 1]$ ), the closeness of  $d_{p,j}$  to  $v(x_{j,i})$  is defined as:

- If  $d_{p,j}$  is between  $v(x_{j,i})$  and  $v(x_{j,i+1})$  then  $\eta_{d_{p,j}} = \frac{v(x_{j,i}) - v(x_{j,i-1})}{d_{p,j} - v(x_{j,i-1})}$ ,
- If  $d_{p,j}$  is between  $v(x_{j,i-1})$  and  $v(x_{j,i})$  then  $\eta_{d_{p,j}} = \frac{v(x_{j,i+1}) - v(x_{j,i})}{v(x_{j,i+1}) - d_{p,j}}$ .

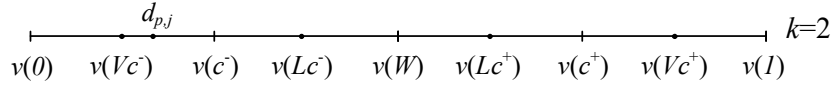


Figure 1. The position of data point  $d_{p,j}$  at the level  $k_j = 2$  in the single granularity structure

Figure 1 shows the position of data point  $d_{p,j}$  between the SQM values of the linguistic terms in case  $k_j$  is 2. In this example,  $d_{p,j}$  is between  $v(Vc^-)$  and  $v(c^-)$ , so the closeness of  $d_{p,j}$  to  $v(c^-)$  is  $\eta_{d_{p,j}} = \frac{v(Lc^-) - v(c^-)}{v(Lc^-) - d_{p,j}}$ .

#### ***In case of multi-granularity structure***

In the multi-granularity structure design, linguistic terms with the same term length  $X_k$  (including two constants 0 and 1) which have the partial order make a separate fuzzy partition. At the level  $k$  ( $0 \leq k \leq k_j$ ), there are SQM values of linguistic terms  $X_k$  with the partial semantic order, i.e.,  $v(x_{j,i-1}^k) \leq v(x_{j,i}^k) \leq v(x_{j,i+1}^k)$ ,  $x_{j,i}^k \in X_k$ .

For a data point  $d_{p,j}$  of the data pattern  $d_p$ , the closeness of  $d_{p,j}$  to  $v(x_{j,i}^k)$  is defined as:

- If  $d_{p,j}$  is between  $v(x_{j,i}^k)$  and  $v(x_{j,i+1}^k)$  then  $\eta_{d_{p,j}} = \frac{v(x_{j,i}^k) - v(x_{j,i-1}^k)}{d_{p,j} - v(x_{j,i-1}^k)}$ ;
- If  $d_{p,j}$  is between  $v(x_{j,i-1}^k)$  and  $v(x_{j,i}^k)$  then  $\eta_{d_{p,j}} = \frac{v(x_{j,i+1}^k) - v(x_{j,i}^k)}{v(x_{j,i+1}^k) - d_{p,j}}$ .



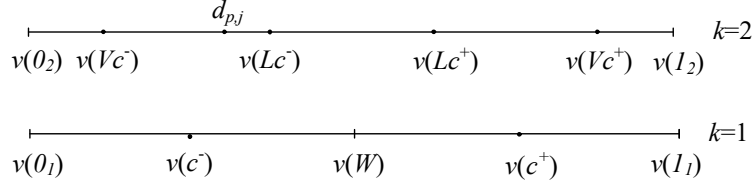


Figure 2. The position of data point  $d_{p,j}$  at the level  $k = 2$  in the multi-granularity structure

For example, Figure 2 shows the position of data point  $d_{p,j}$  between SQM values of linguistic terms in case  $k_j$  is 2. In this case,  $d_{p,j}$  is between  $v(Vc^-)$  and  $v(Lc^-)$ , so the closeness of  $d_{p,j}$  to  $v(Lc^-)$  is  $\eta_{d_{p,j}} = \frac{v(Lc^+) - v(Lc^-)}{v(Lc^+) - d_{p,j}}$ .

We can see that the generality of shorter linguistic terms are preserved with the multi-granularity structure design. The predictability can be improved by high generality classifiers, whereas, high specificity classifiers are good for the particular data. The problem of finding a suitable trade-off between the generality and the specificity of linguistic terms can be given out with the multi-granularity structure design method.

After the formula of the closeness measure of a data point to a specified SQM value of a linguistic term is defined, it is used to compute the compatibility grade of a data pattern  $d_p$  with the antecedent of the rule  $R_q$  as follows:

+ The compatibility grade  $\mu_{A_q}(d_p)$  in the formula (4), (6) and (9) is replaced with  $\eta_{A_q}(d_p)$ .

+  $\eta_{A_q}(d_p)$  is computed as

$$\eta_{A_q}(d_p) = \prod_{j=1}^n \eta_{q,j}(d_{p,j}). \quad (10)$$

+ The formula (4) becomes

$$c(A_q \Rightarrow C_h) = \sum_{d_p \in C_h} \eta_{A_q}(d_p) / \sum_{p=1}^m \eta_{A_q}(d_p). \quad (11)$$

+ The formula (6) becomes

$$s(A_q \Rightarrow C_h) = \sum_{d_p \in C_h} \eta_{A_q}(d_p) / m. \quad (12)$$

+ The formula (9) becomes

$$\eta_{A_w}(d_p) \times CF_w = \operatorname{argmax}(\eta_{A_q}(d_p) \times CF_q \mid R_q \in S). \quad (13)$$

Because the new compatibility grade  $\eta_{A_q}(d_p)$  is computed purely based on the SQM values of the linguistic terms, there is not any fuzzy sets in the proposed model. In the proposed hedge algebras based classification reasoning method, the membership function is replaced with the closeness measure of the data point to the SQM value of the linguistic term.

### 3. EXPERIMENTAL STUDY EVALUATIONS AND DISCUSSIONS

This section presents experimental results of the FRBC applying the proposed hedge algebras based classification reasoning with multi-granularity fuzzy partitioning in comparison with the state-of-the-art results of methods based on hedge algebras [13, 15] and fuzzy sets theory [2]. The real world datasets used in our experiments can be found on the KEEL-Dataset repository: <http://sci2s.ugr.es/keel/datasets.php> and shown in the Table 1. Firstly, two granularity design methods, single granularity and multi-granularities, are compared with each other in order to show the better one. Secondly, the better one is compared to the existing hedge algebras based classifiers proposed in [13, 15] and the fuzzy set theory based approaches proposed in [2]. The comparison conclusions will be made based on the test results of the Wilcoxon's signed rank tests [4]. To make a comparative study, the same cross validation method is used when comparing the methods. All experiments use the ten-folds cross-validation method in which the designated dataset is randomly divided into ten folds, nine folds for the training phase and one fold for the testing phase. Three experiments are executed for each dataset and results of the classification accuracy and the complexity of the FRBCs are averaged out, respectively.

Table 1. The datasets used to evaluate in this research

No.	Dataset Name	Number of attributes	Number of classes	Number of patterns
1	Australian	14	2	690
2	Bands	19	2	365
3	Bupa	6	2	345
4	Dermatology	34	6	358
5	Glass	9	6	214
6	Haberman	3	2	306
7	Heart	13	2	270
8	Ionosphere	34	2	351
9	Iris	4	3	150
10	Mammogr.	5	2	830
11	Pima	8	2	768
12	Saheart	9	2	462
13	Sonar	60	2	208
14	Vehicle	18	4	846
15	Wdbc	30	2	569
16	Wine	13	3	178
17	Wisconsin	9	2	683

In order to have significant comparisons, reduce the searching space in the learning processes and there is no big imbalance between  $fm(c_j^-)$  and  $fm(c_j^+)$ , and between  $\mu(L_j)$  and  $\mu(V_j)$ , constraints on semantic parameter values should be the same as ones used in the compared methods (in [13]) and they are applied as follows: The number of both negative and positive hedges is 1, the negative hedge is “Less” ( $L$ ) and the positive hedge is “Very”

$(V)$ ;  $0 \leq k_j \leq 3$ ;  $0.2 \leq \{fm(c_j^-), fm(c_j^+)\} \leq 0.8$ ;  $fm(c_j^-) + fm(c_j^+) = 1$ ;  $0.2 \leq \mu(L_j)$ ,  $\mu(V_j) \leq 0.8$ ; and  $\mu(L_j) + \mu(V_j) = 1$ .

To optimize semantic parameter values and select the best fuzzy rule set for FRBCs, the multi-objective particle swarm optimization (MOPSO) [30, 34] is utilized. The algorithm parameter values of MOPSO used in the semantic parameter value optimization process are as follows: The number of generations is 250; The number of particles of each generation is 600; Inertia coefficient is 0.4; The self-cognitive factor is 0.2; The social cognitive factor is 0.2; The number of the initial fuzzy rules is equal to the number of attributes; The maximum of rule length is 1. Most of the algorithm parameter values of MOPSO used in the fuzzy rule selection process are the same, except, the number of generations is 1000; The number of initial fuzzy rules  $|S_0| = 300 \times \text{number of classes}$ ; The maximum of rule length is 3.

### 3.1. Single granularity versus multi-granularities

In the fuzzy set theory based approaches, as there is no formal links between linguistic terms of variables and their intuitively designed fuzzy sets, one may be confused to assign linguistic terms to pre-designed fuzzy sets of the multi-granularity structures. Whereas, in the HAS-approach, linguistic terms which have the same length and partially ordered form a fuzzy partition. So, there is no interpretability loss when using multi-granularity structures. This sub-section represents the comparison results between the fuzzy rule based classifier applying the hedge algebras based classification reasoning with single granularity structure (namely HABR-SIG) and the one applying the hedge algebras based classification reasoning with multi-granularity structure (namely HABR-MUL) and shows the important role of the information granule design.

Experimental results of HABR-MUL and HABR-SIG are shown in the Table 2, noting that the column  $\#R \times \#C$  shows the complexities of extracted fuzzy rule bases of the classifiers;  $P_{te}$  is the classification accuracies of the test sets;  $\neq P_{te}$  and  $\neq R \times C$  columns show the differences of the classification accuracies and the complexities of the compared classifiers, respectively. Better values are shown in bold face.

As intuitively recognized from the Table 2, classification accuracies of testing sets of HABR-MUL are better than HABR-SIG on 13 of 17 datasets. The mean value of the classification accuracies on all experimented datasets of HABR-MUL is greater than HABR-SIG while the mean value of the complexity measures of fuzzy rule based systems between them are not much different. Therefore, to know whether the differences of experimental results between two granularity structures are significant or not, Wilcoxon's signed-rank test is applied to test the accuracies and the complexities of fuzzy rule based systems extracted from two granularity structures. It is assumed that their accuracies and complexities are statistically equivalent (null-hypothesis), respectively.

Statistical testing results of the accuracies and the complexities obtained by Wilcoxon's signed-rank tests at level  $\alpha = 0.05$  are shown in the Table 3 and Table 4, respectively. The abbreviation terms used in the statistical test result tables from now on: VS column is the list of the compared method names; E. is Exact; A. is Asymptotic.

As shown in the Table 4, since the  $p\text{-value} > 0.05$ , the *null-hypothesis* is not rejected. There is no significant difference of the complexities between the two compared methods. Therefore, there is no need to take the complexity of the FRBCs into account in this case

Table 2. The experimental results of the HABR-MUL and the HABR-SIG classifiers

Dataset	HABR-MUL		HABR-SIG		$\neq R \times C$	$\neq P_{te}$
	$\#R \times \#C$	$T_{te}$	$\#R \times \#C$	$T_{te}$		
Australian	46.38	<b>87.29</b>	53.24	86.33	-6.86	0.96
Bands	53.22	73.53	60.60	<b>73.61</b>	-7.38	-0.08
Bupa	152.76	<b>72.13</b>	203.13	71.82	-50.37	0.31
Dermatology	215.64	<b>96.55</b>	191.84	95.47	23.80	1.08
Glass	403.08	73.09	318.68	<b>73.77</b>	84.40	-0.68
Haberman	9.00	77.11	8.82	77.11	0.18	0.00
Heart	105.16	<b>83.95</b>	122.92	83.70	-17.76	0.25
Ionosphere	58.29	<b>93.18</b>	92.80	92.22	-34.51	0.96
Iris	30.35	<b>98.67</b>	28.41	97.56	1.94	1.11
Mammogr.	50.57	<b>84.35</b>	85.04	84.33	-34.47	0.02
Pima	57.65	<b>77.28</b>	52.02	76.18	5.63	1.10
Saheart	59.40	72.23	56.40	<b>72.60</b>	3.00	-0.37
Sonar	64.62	<b>79.29</b>	61.80	77.52	2.82	1.77
Vehicle	236.17	<b>68.20</b>	333.94	68.01	-97.77	0.19
Wdbc	47.35	<b>96.31</b>	47.15	95.26	0.20	1.05
Wine	34.00	<b>99.61</b>	43.20	99.44	-9.20	0.17
Wisconsin	49.85	96.99	66.71	<b>97.19</b>	-16.86	-0.20
<b>Mean</b>	<b>98.44</b>	<b>84.10</b>	<b>107.45</b>	<b>83.65</b>		

Table 3. The comparison result of the accuracy of HABR-MUL and HABR-SIG classifiers using the Wilcoxon signed rank test at level  $\alpha = 0.05$ 

VS	$R^+$	$R^-$	E. $P$ -value	A. $P$ -value	Hypothesis
HABR-MUL vs HABR-SIG	112.0	24.0	0.0214	0.020558	Rejected

Table 4. The comparison result of the complexity of HABR-MUL and HABR-SIG classifiers using the Wilcoxon signed rank test at level  $\alpha = 0.05$ 

VS	$R^+$	$R^-$	E. $P$ -value	A. $P$ -value	Hypothesis
HABR-MUL vs HABR-SIG	104.0	49.0	$\geq 0.2$	0.185016	Not rejected

of comparison. The comparison result of the classification accuracies is shown in the Table 3. Since the  $p$ -value = 0.0214 < 0.05, the *null-hypothesis* is rejected. Based on statistical testing results, we can state that the multi-granularity based classifier outperforms the single granularity based classifier. In the next sub-sections, the multi-granularity structure is the default granular design method in our experiments.

### 3.2. The proposed classifier versus the existing hedge algebras based classifiers

This sub-section presents the evaluation of the proposed classifier (HABR-MUL) in comparisons with the existing hedge algebras based classifiers. For the reading convenience, the hedge algebras based classifier with the triangular [13] and trapezoidal [15] fuzzy set based semantics of linguistic values are named as HATRI and HATRA, respectively. Their experimental results in the Table 5 show that HABR-MUL has better classification accuracies on 15 and 13 of 17 experimental datasets than HATRI and HATRA, respectively. The mean value of the classification accuracies of HABR-MUL is higher than HATRI and HATRA (84.10% in comparison with 82.82% and 83.58, respectively). The mean value of the fuzzy rule base complexities of HABR-MUL is a bit lower than both HATRI and HATRA (98.44 in comparison with 104.52 and 103.79, respectively).

Table 5. Experimental results of HABR-MUL, HATRI and HATRA classifiers

Dataset	HABR-MUL		HATRI		$\neq R \times C$	$\neq P_{te}$	HATRA		$\neq R \times C$	$\neq P_{te}$
	$\#R \times \#C$	$T_{te}$	$\#R \times \#C$	$T_{te}$			$\#R \times \#C$	$T_{te}$		
Australian	46.38	87.29	36.20	86.38	10.17	0.91	46.50	87.15	-0.12	0.14
Bands	53.22	73.53	52.20	72.80	1.02	0.73	58.20	73.46	-4.98	0.07
Bupa	152.76	72.13	187.20	68.09	-34.44	4.04	181.19	72.38	-28.44	-0.25
Dermatology	215.64	96.55	198.05	96.07	17.58	0.48	182.84	94.40	32.80	2.15
Glass	403.08	73.09	343.60	72.09	59.49	1.00	474.29	72.24	-71.20	0.85
Haberman	9.00	77.11	10.20	75.76	-1.20	1.35	10.80	77.40	-1.80	-0.29
Heart	105.16	83.95	122.72	84.44	-17.56	-0.49	123.29	84.57	-18.13	-0.62
Ionosphere	58.29	93.18	90.33	90.22	-32.04	2.96	88.03	91.56	-29.73	1.62
Iris	30.35	98.67	26.29	96.00	4.06	2.67	30.37	97.33	-0.02	1.34
Mammogr.	50.57	84.35	92.25	84.20	-41.69	0.15	73.84	84.20	-23.27	0.15
Pima	57.65	77.28	60.89	76.18	-3.24	1.10	56.12	77.01	1.53	0.27
Saheart	59.40	72.23	86.75	69.33	-27.35	2.90	59.28	70.05	0.12	2.18
Sonar	64.62	79.29	79.76	76.80	-15.14	2.49	49.31	78.61	15.31	0.68
Vehicle	236.17	68.20	242.79	67.62	-6.62	0.58	195.07	68.20	41.10	0.00
Wdbc	47.35	96.31	37.35	96.96	10.00	-0.65	25.04	96.78	22.31	-0.47
Wine	34.00	99.61	35.82	98.30	-1.82	1.31	40.39	98.49	-6.39	1.12
Wisconsin	49.85	96.99	74.36	96.74	-24.51	0.25	69.81	96.95	-19.96	0.04
Mean	98.44	84.10	104.52	82.82			103.79	83.58		

Table 6. The comparison result of the accuracy of HABR-MUL, HATRI and HATRA classifiers using the Wilcoxon signed rank test at level  $\alpha = 0.05$

VS	$R^+$	$R^-$	E. $P$ -value	A. $P$ -value	Hypothesis
HABR-MUL vs HATRI	143.0	10.0	6.562E-4	0.001516	Rejected
HABR-MUL vs HATRA	107.0	29.0	0.04432	0.041102	Rejected

To make sure the differences are significant, Wilcoxon's signed-rank test at level  $\alpha = 0.05$  is used to test the equivalent hypotheses. As shown in the Table 6, all  $p$ -values are less than  $\alpha = 0.05$ , all *null-hypotheses* are rejected. In the Table 7, all  $p$ -values are greater than  $\alpha = 0.05$ , all *null-hypotheses* are not rejected. Thus, we can state that the HABR-MUL has better classification accuracy than HATRI and HATRA while the complexities of the fuzzy rule bases are equivalent.

Table 7. The comparison result of the complexity of HABR-MUL, HATRI and HATRA classifiers using the Wilcoxon signed rank test at level  $\alpha = 0.05$

VS	$R^+$	$R^-$	E. $P$ -value	A. $P$ -value	Hypothesis
HABR-MUL vs HATRI	104.0	49.0	$\geq 0.2$	0.185016	Not rejected
HABR-MUL vs HATRA	97.0	56.0	$\geq 0.2$	0.320174	Not rejected

### 3.3. The proposed classifier versus the fuzzy set theory based classifiers

To show more about the efficiency of the proposed classifier, we run a comparison study of the proposed classifier with existing fuzzy rule based classifiers examined by M. Antonelli et al. 2014 so-called PAES-RCS in conjunction with non-evolutionary classification algorithms so-called FURIA [2].

Table 8. The experimental results of the HABR-MUL, the PAES-RCS and the FURIA classifiers

Dataset	HABR-MUL		PAES-RCS		$\neq R \times C$	$\neq P_{te}$	FURIA		$\neq R \times C$	$\neq P_{te}$
	$\#R \times \#C$	$T_{te}$	$\#R \times \#C$	$T_{te}$			$\#R \times \#C$	$T_{te}$		
Australian	46.38	87.29	329.64	85.80	-283.26	1.49	89.60	85.22	-43.22	2.07
Bands	53.22	73.53	756.00	67.56	-702.78	5.97	535.15	64.65	-481.93	8.88
Bupa	152.76	72.13	256.20	68.67	-103.44	3.46	324.12	69.02	-171.36	3.11
Dermatology	215.64	96.55	389.40	95.43	-173.76	1.12	303.88	95.24	-88.24	1.31
Glass	403.08	73.09	487.90	72.13	-84.82	0.96	474.81	72.41	-71.73	0.68
Haberman	9.00	77.11	202.41	72.65	-193.41	4.46	22.04	75.44	-13.04	1.67
Heart	105.16	83.95	300.30	83.21	-195.14	0.74	193.64	80.00	-88.48	3.95
Ionosphere	58.29	93.18	670.63	90.40	-612.34	2.78	372.68	91.75	-314.39	1.43
Iris	30.35	98.67	69.84	95.33	-39.49	3.34	31.95	94.66	-1.60	4.01
Mammogr.	50.57	84.35	132.54	83.37	-81.97	0.98	16.83	83.89	33.74	0.46
Pima	57.65	77.28	270.64	74.66	-212.99	2.62	127.50	74.62	-69.85	2.66
Saheart	59.40	72.23	525.21	70.92	-465.81	1.31	50.88	69.69	8.52	2.54
Sonar	64.62	79.29	524.60	77.00	-459.98	2.29	309.96	82.14	-245.34	-2.85
Vehicle	236.17	68.20	555.77	64.89	-319.60	3.31	2125.97	71.52	-1889.80	-3.32
Wdbc	47.35	96.31	183.70	95.14	-136.35	1.17	356.12	96.31	-308.77	0.00
Wine	34.00	99.61	170.94	93.98	-136.94	5.63	80.00	96.60	-46.00	3.01
Wisconsin	49.85	96.99	328.02	96.46	-278.17	0.53	521.10	96.35	-471.25	0.64
Mean	<b>98.44</b>	<b>84.10</b>	<b>361.98</b>	<b>81.62</b>			<b>349.19</b>	<b>82.32</b>		

Table 9. The comparison result of the accuracy of HABR-MUL, PAES-RCS and FURIA classifiers using the Wilcoxon signed rank test at level  $\alpha = 0.05$

VS	$R^+$	$R^-$	E. $P$ -value	A. $P$ -value	Hypothesis
HABR-MUL vs PAES-RCS	153.0	0.0	1.5258E-5	0.000267	Rejected
HABR-MUL vs FURIA	113.0	23.0	0.01825	0.018635	Rejected

PAES-RCS [2] is a multi-objective evolutionary approach deployed to learn concurrently the fuzzy rule bases and databases of FRBCs. It exploits the pre-specified granularity of each attribute for generating the candidate fuzzy set by applying the C4.5 algorithm [32]. Then,

Table 10. The comparison result of the complexity of the HABR-MUL, the PAES-RCS and the FURIA classifiers using the Wilcoxon signed rank test at level  $\alpha = 0.05$

VS	$R^+$	$R^-$	E. $P$ -value	A. $P$ -value	Hypothesis
HABR-MUL vs PAES-RCS	153.0	0.0	1.5258E-5	0.000267	Rejected
HABR-MUL vs FURIA	147.0	6.0	2.136E-4	0.000777	Rejected

the multi-objective evolutionary process is performed to select a set of fuzzy rules from the candidate rule set in conjunction with a set of conditions for each selected rule, named as the rule and condition selection (RCS). The membership functions of linguistic terms are concurrently learned during the RCS process.

The comparison of the classification accuracies on test sets and the complexities between the proposed classifier and the two other classifiers PAES-RCS and FURIA are shown in the Table 8. The HABR-MUL has better classification accuracies and better classifier complexities than PAES-RCS on all test datasets. HABR-MUL has better classification accuracies and better classifier complexities than FURIA on 15 of 17 test datasets. Based on mean values of the classification accuracies and the classifier complexities, the proposed classifier is much better than PAES-RCS and FURIA on both classification accuracy and complexity measures. To make sure the differences are significant, Wilcoxon's signed-rank test at level  $\alpha = 0.05$  is used to test the equivalent hypotheses. As shown in the Table 9 and the Table 10, since all  $p$ -values are less than  $\alpha = 0.05$ , all *null-hypotheses* are rejected. Thus, we can state that the proposed classifier strictly outperforms PAES-RCS and FURIA classifiers.

#### 4. CONCLUSIONS

Fuzzy rule based systems which deal with uncertainty information have been applied successfully in solving the FRBC design problem. There is the fact that although fuzzy rule bases are represented by linguistic terms associated with their fuzzy set based semantics, the problem of the linguistic term design is not clearly studied in the fuzzy set theory approaches. HAs provide a mathematical formalism of term design so that the fuzzy set based semantics of all linguistic terms are generated from qualitative semantics of terms. So far, the FRBCs design based on hedge algebra approach generate fuzzy rule bases with fuzzy sets based semantics of linguistic terms for classifiers. This paper presents a pure hedge algebra based classifier design methodology which generates fuzzy rule based classifiers with the semantics of the linguistic terms in the fuzzy rule bases are the hedge algebras based semantics. To do so, a hedge algebra based classification reasoning with multi-granularity fuzzy partitioning method is applied for data classification. The new classification reasoning method enables fuzzy sets based semantics of linguistic terms in fuzzy rule bases of classifiers to be replaced with hedge algebra based semantics. Experimental results on 17 real world datasets have shown that the multi-granularity structure is more efficient than the single granularity structure and the proposed classifier outperforms the existing ones. By research results of this paper, we can state that fuzzy rule based classifiers can be designed purely based on hedge algebras based semantics of linguistic terms.

## 5. ACKNOWLEDGMENT

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant No. 102.01-2017.06.

## REFERENCES

- [1] R. Alcal, Y. Nojima, F. Herrera, and H. Ishibuchi, "Multiobjective genetic fuzzy rule selection of single granularity-based fuzzy classification rules and its interaction with the lateral tuning of membership functions," *Soft Computing*, vol. 15, no. 12, pp. 2303–2318, 2011.
- [2] M. Antonelli, P. Ducange, and F. Marcelloni, "A fast and efficient multi-objective evolutionary learning scheme for fuzzy rule-based classifiers," *Information Sciences*, vol. 283, pp. 36–54, 2014.
- [3] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," in *Proceedings of Int Conference on Data Mining and Knowledge Discovery*, vol. 2, no. 2. Boston. Manufactured in The Netherlands: Kluwer Academic Publishers, 1998, pp. 121–167.
- [4] J. Demar, "Statistical comparisons of classifiers over multiple data sets," *The Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.
- [5] D. K. Dong, T. D. Khang, and D. K. Dung, "Fuzzy clustering with hedge algebra," in *Proceedings of the 2010 Symposium on Information and Communication Technology*, Hanoi, Vietnam, 2010, pp. 49–54.
- [6] M. Elkano, M. Galar, J. Sanz, and H. Bustince, "Chi-bd: A fuzzy rule-based classification system for big data classification problems," *Fuzzy Sets and Systems*, vol. 348, pp. 75–101, 2018.
- [7] M. Fazzolari, R. Alcal, and F. Herrera, "A multi-objective evolutionary method for learning granularities based on fuzzy discretization to improve the accuracy-complexity trade-off of fuzzy rule-based classification systems: D-mofarc algorithm," *Applied Soft Computing*, vol. 24, pp. 470–481, 2014.
- [8] A. K. Ghosh, "A probabilistic approach for semi-supervised nearest neighbor classification," *Pattern Recognition Letters*, vol. 33, no. 9, pp. 1127–1133, 2012.
- [9] N. C. Ho, "A topological completion of refined hedge algebras and a model of fuzziness of linguistic terms and hedges," *Fuzzy Sets and Systems*, vol. 158, no. 4, pp. 436–451, 2007.
- [10] N. C. Ho, V. N. Lan, and L. X. Viet, "Optimal hedge-algebras-based controller: Design and application," *Fuzzy Sets and Systems*, vol. 159, no. 8, pp. 968–989, 2008.
- [11] N. C. Ho and N. V. Long, "Fuzziness measure on complete hedges algebras and quantifying semantics of terms in linear hedge algebras," *Fuzzy Sets and Systems*, vol. 158, no. 4, pp. 452–471, 2007.
- [12] N. C. Ho, H. V. Nam, T. D. Khang, and L. H. Chau, "Hedge algebras, linguistic-valued logic and their application to fuzzy reasoning," *Internat. J. Uncertain. Fuzziness Knowledge-Based Systems*, vol. 7, no. 4, pp. 347–361, 1999.
- [13] N. C. Ho, W. Pedrycz, D. T. Long, and T. T. Son, "A genetic design of linguistic terms for fuzzy rule based classifiers," *International Journal of Approximate Reasoning*, vol. 54, no. 1, pp. 1–21, 2013.



- [14] N. C. Ho, T. T. Son, T. D. Khang, and L. X. Viet, “Fuzziness measure, quantified semantic mapping and interpolative method of approximate reasoning in medical expert systems,” *Journal of Computer Science and Cybernetics*, vol. 18, no. 3, pp. 237–252, 2002.
- [15] N. C. Ho, T. T. Son, and P. D. Phong, “Modeling of a semantics core of linguistic terms based on an extension of hedge algebra semantics and its application,” *Knowledge-Based Systems*, vol. 67, pp. 244–262, 2014.
- [16] N. C. Ho, H. V. Thong, and N. V. Long, “A discussion on interpretability of linguistic rule based systems and its application to solve regression problems,” *Knowledge-Based Systems*, vol. 88, pp. 107–133, 2015.
- [17] N. C. Ho and W. Wechler, “Hedge algebras: an algebraic approach to structures of sets of linguistic domains of linguistic truth values,” *Fuzzy Sets and Systems*, vol. 35, pp. 281–293, 1990.
- [18] —, “Extended algebra and their application to fuzzy logic,” *Fuzzy Sets and Systems*, vol. 52, no. 3, pp. 259–281, 1992.
- [19] L. N. Hung and V. M. Loc, “Primacy of fuzzy relational databases based on hedge algebras,” in *Volume 144 of the series Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*. Ho Chi Minh City, Vietnam: Springer, Cham, 2015, pp. 292–305.
- [20] N. H. Huy, N. C. Ho, and N. V. Quyen, “Multichannel image contrast enhancement based on linguistic rule-based intensifiers,” *Applied Soft Computing Journal*, vol. 76, pp. 744–762, 2019.
- [21] H. Ishibuchi and Y. Nojima, “Analysis of interpretability-accuracy tradeoff of fuzzy systems by multi-objective fuzzy genetics-based machine learning,” *International Journal of Approximate Reasoning*, vol. 44, pp. 4–31, 2007.
- [22] H. Ishibuchi, K. Nozaki, and H. Tanaka, “Distributed representation of fuzzy rules and its application to pattern classification,” *Fuzzy Sets and Systems*, vol. 52, no. 1, pp. 21–32, 1992.
- [23] H. Ishibuchi and T. Yamamoto, “Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining,” *Fuzzy Sets and Systems*, vol. 141, no. 1, pp. 59–88, 2004.
- [24] —, “Rule weight specification in fuzzy rule-based classification systems,” *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 4, pp. 428–435, 2005.
- [25] L. V. T. Lan, N. M. Han, and N. C. Hao, “An algorithm to build a fuzzy decision tree for data classification problem based on the fuzziness intervals matching,” *Journal of Computer Science and Cybernetics*, vol. 32, no. 4, pp. 367–380, 2016.
- [26] V. N. Lan, T. T. Ha, P. K. Lai, and N. T. Duy, “The application of the hedge algebras in forecast control based on the models,” in *In Proceedings of The 11st National Conference on Fundamental and Applied IT Research*, Hanoi, Vietnam, 2018, pp. 521–528.
- [27] H. Langseth and T. D. Nielsen, “Classification using hierarchical naïve bayes models,” *Machine Learning*, vol. 63, no. 2, pp. 135–159, 2006.
- [28] B. H. Le, L. T. Anh, and B. V. Binh, “Explicit formula of hedge-algebras-based fuzzy controller and applications in structural vibration control,” *Applied Soft Computing*, vol. 60, pp. 150–166, 2017.

- [29] B. H. Le, N. C. Ho, V. N. Lan, and N. C. Hung, "General design method of hedge-algebras-based fuzzy controllers and an application for structural active control," *Applied Intelligence*, vol. 43, no. 2, pp. 251–275, 2015.
- [30] M. S. Lechuga, "Multi-objective optimization using sharing in swarm optimization algorithms," 2016.
- [31] D. T. Long, "A genetic algorithm based method for timetabling problems using linguistics of hedge algebra in constraints," *Journal of Computer Science and Cybernetics*, vol. 32, no. 4, pp. 285–301, 2016.
- [32] M. M. Mazid, M. M. Mazid, and K. S. Tickle, "Improved c4.5 algorithm for rule based classification," in *Proceedings of the 9th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Databases*, University of Cambridge, UK, 2010, pp. 296–301.
- [33] D. Nauck and R. Kruse, "Nefclass: A neuro-fuzzy approach for the classification of data," in *Proceedings of the 1995 ACM symposium on Applied computing*, Nashville, TN, 1995, pp. 461–465.
- [34] P. D. Phong, N. C. Ho, and N. T. Thuy, "Multi-objective particle swarm optimization algorithm and its application to the fuzzy rule based classifier design problem with the order based semantics of linguistic terms," in *In Proceedings of The 10th IEEE RIVF International Conference on Computing and Communication Technologies (RIVF-2013)*, Hanoi, Vietnam, 2013, pp. 12–17.
- [35] M. Pota, M. Esposito, and G. D. Pietro, "Designing rule-based fuzzy systems for classification in medicine," *Knowledge-Based Systems*, vol. 124, pp. 105–132, 2017.
- [36] M. I. Rey, M. Galende, M. J. Fuente, and G. I. Sainz-Palmero, "Multi-objective based fuzzy rule based systems (frbss) for trade-off improvement in accuracy and interpretability: A rule relevance point of view," *Knowledge-Based Systems*, vol. 127, pp. 67–84, 2017.
- [37] S.-B. Roh, W. Pedrycz, and T.-C. Ahn, "A design of granular fuzzy classifier," *Expert Systems with Applications*, vol. 41, pp. 6786–6795, 2014.
- [38] F. Rudziski, "A multi-objective genetic optimization of interpretability-oriented fuzzy rule-based classifiers," *Applied Soft Computing*, vol. 38, pp. 118–133, 2016.
- [39] J. Sanz, A. Fernndez, H. Bustince, and F. Herrera, "A genetic tuning to improve the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets: Degree of ignorance and lateral position," *International Journal of Approximate Reasoning*, vol. 52, no. 6, pp. 751–766, 2011.
- [40] T. T. Son and N. T. Anh, "Partition fuzzy domain with multi-granularity representation of data based on hedge algebra approach," *Journal of Computer Science and Cybernetics*, vol. 34, no. 1, pp. 63–75, 2018.
- [41] M. Soui, I. Gasmi, S. Smiti, and K. Ghdira, "Rule-based credit risk assessment model using multi-objective evolutionary algorithms," *Expert Systems with Applications*, vol. 126, pp. 144–157, 2019.
- [42] D. V. Thang and D. V. Ban, "Query data with fuzzy information in object-oriented databases an approach the semantic neighborhood of hedge algebras," *International Journal of Computer Science and Information Security*, vol. 9, no. 5, pp. 37–42, 2011.

*Received on August 22, 2019*

*Revised on October 03, 2019*