Vietnam Journal of Science and Technology ***57*** *(5) (2019) 631-644*

doi:10.15625/2525-2518/57/5/13811

A HEDGE ALGEBRAS BASED REASONING METHOD FOR FUZZY RULE BASED CLASSIFIER

Pham Dinh Phong*, Nguyen Duc Du*, Hoang Van Thong

Faculty of Information Technology, University of Transport and Communications, No. 3, Cau Giay street, Dong Da district, Ha Noi

*Email: *dinhphongpham@gmail.com*, *nducdu@gmail.com*

Received: 8 May 2019; Accepted for publication: 6 July 2019

Abstract. The fuzzy rule based classifier (FRBC) design methods have intensively been being studied during recent years. The ones designed by utilizing hedge algebras as a formalism to generate the optimal linguistic values along with their (triangular and trapezoidal) fuzzy sets based semantics for the FRBCs have been proposed. Those design methods generate the fuzzy sets based semantics because the classification reasoning method still bases on the fuzzy set theory. One question arisen is whether there is a pure hedge algebras classification reasoning method so that the fuzzy sets based semantics of the linguistic values in the fuzzy rule bases can be replaced with the hedge algebras based semantics. This paper answers that question by presenting a fuzzy rule based classifier design method based on hedge algebras with a pure hedge algebras classification reasoning method. The experimental results over 17 real world datasets are compared to the existing methods based on hedge algebras and fuzzy sets theory showing that the proposed method is effective and produces good results.

Keywords: fuzzy rule based classifier, hedge algebras, fuzziness measure, fuzziness intervals, semantically quantifying mapping value.

Classification numbers: 4.7.3, 4.7.4, 4.10.2.

1. INTRODUCTION

The fuzzy rule based classifiers (FRBCs) have been studied intensively in the data mining field and has achieved a lot of successful results [1-13]. The advantage of this classification model is that the end-users can use the high interpretability fuzzy rule based knowledge extracted automatically from data in the form of if-then sentences as their knowledge.

The FRBC design method based on the fuzzy set theory approach [1-13] exploits the pre-specified fuzzy partitions constructed by the fuzzy sets. To improve the classification accuracy and the interpretability of the fuzzy rule bases, a genetic fuzzy system is developed to adjust the fuzzy set parameters to achieve the optimal fuzzy partitions. Because there is not any formal mechanism to link the real world semantic of the linguistic values and their designed fuzzy sets, the received fuzzy sets after the learning processes do not reflect the inherent semantics of the linguistic values. Therefore, the interpretability of the fuzzy rule based systems of the classifiers is affected.

Hedge algebras (HAs) [14-18] were introduced by Ho N. C. *et al*. in the early 1990s and then HAs have been applied to many different fields such as data mining [19-25], fuzzy control [26-28], image processing [29], timetabling [30], *etc*. When applied to design the FRBCs, HAs take advantage of the algebraic approach which allows to design automatically the linguistic values integrated with their fuzzy sets from data [19, 20] for the FRBCs. To do so, the inherent semantic order of the linguistic values is exploited to generate the formal linkage between the terms and their integrated fuzzy sets in the form of triangle or/and trapezoid. This formalism helps to construct the effective fuzzy rule based classifiers introduced in [19, 20].

One question which has been arisen is that why the fuzzy sets are generated for the FRBCs designed by HAs based methodology. The reason is that the knowledge bases for the classifiers are designed by HAs, but the classification reasoning method is still based on the fuzzy set theory. Is there a pure hedge algebras classification reasoning method for the FRBCs? The research results of this paper will answer the question. In [27], a Takagi-Sugeno-Hedge algebras fuzzy model was proposed to improve the forecast control based on the models in such a way that the membership functions of the individual linguistic values in Takagi-Sugeno fuzzy model are replaced with the closeness of the semantically quantifying mapping values of the adjacent linguistic values. That idea can be enhanced to build a classification reasoning method based on HAs for the FRBC design problem. This paper presents a FRBC design method based on hedge algebras with a pure hedge algebras classification reasoning method which enables the fuzzy sets based semantics of the linguistic values in the fuzzy rule bases to be replaced with the hedge algebras based semantics. The experimental results over 17 real world datasets are compared to the existing methods based on hedge algebras and fuzzy sets theory showing that the proposed method is effective and produces good results.

The rest of this paper is organized as follows: Section 2 presents some basic concepts of hedge algebras, the fuzzy rule base classifier design method based on hedge algebras approach and the proposed pure hedge algebras classifier. Section 3 presents the experimental results and discussion. The conclusion remarks are on Section 4.

2. FUZZY RULE BASED CLASSIFIER DESIGN BASED ON HEDGE ALGEBRAS

2.1. Some basic concepts of hedge algebras

Assume that $\mathcal{X}$ is a linguistic variable and $Dom(\mathcal{X})$ is the linguistic value domain of $\mathcal{X}$. 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 (abbreviated as term) of X and $X \subseteq Dom(\mathcal{X})$.
- ✓ G is a set of two generator terms c^+ and c^-. c^- is the negative primary term, c^+ is the positive primary term and $c^- \leq c^+$.
- ✓ C is a set of term constants, $C = \{\boldsymbol{0}, W, \boldsymbol{1}\}$, satisfying the relation order $\boldsymbol{0} \leq c^- \leq W \leq c^+ \leq \boldsymbol{1}$. $\boldsymbol{0}$ and $\boldsymbol{1}$ are the least and greatest terms respectively, W is the neutral term.
- ✓ H is a set of hedges of $\mathcal{X}$.
- ✓ $\leq$ is an order relation induced by the inherent semantics of terms of $\mathcal{X}$.

When a hedge acts on a non-constant term, a new term is induced. For example, *Age* is a linguistic variable. Two generators G = {"*young*", "*old*"}, $C = \{\boldsymbol{0}, W, \boldsymbol{1}\}$ where W = {"*middle*"}, $\boldsymbol{0}$ = "*absolutely young*", $\boldsymbol{1}$ = "*absolutely old*", H = {*Less, Very*}. $X_{(2)}$ is the set of terms of variable *Age* generated from "*young*" and "*old*" using the hedges *less* and *very*, $X_{(2)}$ =

{"*absolutely young*", "*young*", "*middle*", "*old*", "*absolutely old*"} $\cup$ {"*less young*", "*very young*", "*less old*", "*very old*"}. Note that $X_{(k)}$ denotes the set of terms which have the term lengths less than and equal to k. Each term x in X can be 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 the terms generated from x by using the hedges in H can be abbreviated as $H(x)$.

Each hedge possesses tendency to decrease or increase the semantics of other hedge. If k makes the sematic of h increased, k is positive with respect to h, whereas, if k makes the sematic 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. One hedge may have a relative sign with respect to another. $Sign(k, h) = +1$ if k strengthens the effect tendency of h, whereas, $Sign(k, h) = -1$ 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:

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

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

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

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

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

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

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

From (FM1) and (FM3), the fuzziness measure of 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 $v: X \longrightarrow [0, 1]$ satisfying the following conditions:

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

(SQM2): It is one-to-one mapping and $\mathbb{f}(x)$ 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) $v(h_j x) = v(x) + sign(h_j x)(\sum_{i=sign(j)}^{j} fm(h_i x) - \omega\,(h_j x) fm(h_j x))$ where

$j \in [-q\text{^}p] = \{j: -q \leq j \leq p\ \&\ j \neq 0\}$ and

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

2.2. Fuzzy rule base classifier design based on hedge algebras

The fuzzy rule based knowledge of the FRBCs used in this paper is a set of weighted fuzzy rules in the form as following [5-7]:

$$\text{Rule } R_q\text{: IF } \mathcal{X}_1 \text{ is } A_{q,1} \text{ AND ... AND } \mathcal{X}_n \text{ is } A_{q,n} \text{ THEN } C_q \text{ with } CF_q, \text{ for } q=1, ..., N \quad (1)$$

where $\mathfrak{X} = \{\mathcal{X}_j, j = 1, .., n\}$ is a set of n linguistic variables corresponding to n features of the dataset $\boldsymbol{D}$, $A_{q,j}$ is the linguistic term of the j^{th} feature F_j, C_q is a class label, there are M class labels of each dataset, and CF_q is the weight of rule R_q. The rule R_q can be abbreviated as the short form hereafter:

$$A_q \Rightarrow C_q \text{ with } CF_q, \text{ for } q=1, \ldots, N \quad (2)$$

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

A FRBC design problem $\mathcal{P}$ is defined as: a set $\boldsymbol{P} = \{(\boldsymbol{d}_p, C_p) \mid \boldsymbol{d}_p \in \boldsymbol{D}, C_p \in \boldsymbol{C}, p = 1, \ldots, m;\}$ of m data patterns, where $\boldsymbol{d}_p = [d_{p,1}, d_{p,2}, ..., d_{p,n}]$ is the row p^{th} of $\boldsymbol{D}$, $\boldsymbol{C} = \{C_s \mid s = 1, \ldots, M\}$ is the set of M class labels. Solving the problem $\mathcal{P}$ is to extract automatically from $\boldsymbol{P}$ a set $\boldsymbol{S}$ of fuzzy rules in the form (1) in such a way as to achieve a FRBC based on $\boldsymbol{S}$ which comes with high classification accuracy, interpretability and comprehensibility.

As the previous researches, the FRBC design method based on hedge algebras comprises two following phases [19, 20]:

(1) A hybrid model between hedge algebras and an evolutionary multi-objective optimization algorithm is developed to design automatically the optimal linguistic terms along with their fuzzy-set-based semantics for each dataset feature which are the consequence of the interacting between the semantics of the linguistic terms and the data.

(2) Based on the optimal linguistic terms received from the first phase, extract the optimal fuzzy rule set for the FRBCs from the dataset in such a way as to achieve their suitable interpretability–accuracy tradeoff.

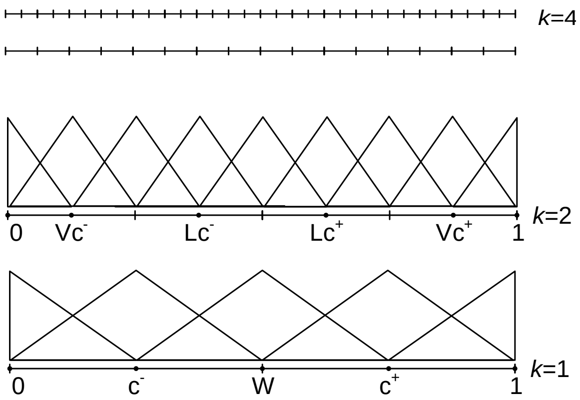

Figure 1. The fuzzy sets of the linguistic terms with $k_j = 2$.

Two phases mentioned above are summarized as follows:

The j^{th} feature of the designated dataset is associated with a hedge algebras AX_j. With the given values of the semantic parameters $\boldsymbol{\Pi}$, including $fm_j(c^-)$, $\mu(h_{j,i})$ and k_j which are the fuzziness measure of the primary term c^-, the fuzziness measure of the hedges and a positive integer to limit the linguistic term lengths of j^{th} feature respectively, the fuzziness intervals $\mathfrak{I}_k(x_{j,i})$, $x_{j,i} \in X_{j,k}$ for all $k \le k_j$ and the SQM values $v(x_{j,i})$ are computed. Based on the generated values $\mathfrak{I}_k(x_{j,i})$ and $v(x_{j,i})$, the fuzzy-set-based semantics of the terms $X_{j,(kj)}$ are computationally constructed. All the constructed fuzzy sets of the linguistic terms $X_{j,(kj)}$ which is the union of the subsets $X_{j,k}$, k = 1 to k_j, and the k_j-similarity intervals $\mathfrak{S}_{k_j}(X_{j,i})$ of the linguistic terms in $X_{j,kj+2}$ constitute a fuzzy partition of the feature reference space. For example, Figure 1 denotes the designed fuzzy sets of the linguistic terms and the k_j-similarity intervals with $k_j = 2$.

After the fuzzy partitions of all features of the dataset $\boldsymbol{P}$ are constructed, the fuzzy rules are extracted from that dataset. In a specific fuzzy partition at the level k_j, there is a unique k_j-similarity interval $\mathfrak{S}_{k_j}(X_{j,i(i)})$ compatible with the linguistic term $x_{j,i(j)}$ containing j^{th}-component $d_{p,j}$ of the data pattern d_p. All k_j-similarity intervals which contain $d_{p,j}$ component forms a hyper-cube $\boldsymbol{\mathcal{H}}_p$. The fuzzy rules are only be induced from $\boldsymbol{\mathcal{H}}_p$. So, a fuzzy rule which is so-called a *basic fuzzy rule* for the class C_p of $(d_p, C_p) \in \boldsymbol{P}$ is generated from $\boldsymbol{\mathcal{H}}_p$ in the following form:

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

Only one basic fuzzy rule with the length n are generated from a data pattern. Some techniques should be applied to generate the fuzzy rules with the length $L \leq n$, so-called the *secondary rules*. The worst case is to generate all possible combinations.

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

where $1 \leq j_1 \leq \ldots \leq j_t \leq n$. The consequence class C_q of the rule R_q is determined by the maximum of the confidence measure $c(\boldsymbol{A}_q \Rightarrow C_h)$ of R_q:

$$C_q = argmax\{c(\boldsymbol{A}_q \Rightarrow C_h) | h = 1, \ldots, M\} \tag{3}$$

The confidence measure is computed as:

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

where $\mu_{\boldsymbol{A}_q}(d_p)$ is the burning of the data pattern d_p for R_q and commonly computed as:

$$\mu_{A_q}(d_p) = \prod_{j=1}^{n} \mu_{q,j}(d_{p,j}). \tag{5}$$

In the worst case, the maximum of the number fuzzy combinations is $\sum_i^L C_n^i$, so the maximum of the *secondary rules* is $m \times \sum_i^L C_n^i$.

The inconsistent secondary fuzzy rules which have the identical antecedents and different consequence classes are eliminated by the confident measure to receive a set of the so-called *candidate fuzzy rules*. The candidate fuzzy rules may be screened by a screening criterion to select a subset S_0 with NR_0 fuzzy rules, so-called the *initial fuzzy rule set*. The above process is so-called the initial fuzzy rule set generation procedure IFRG($\boldsymbol{\Pi}$, $\boldsymbol{P}$, NR_0, L) [19], where $\boldsymbol{\Pi}$ is a set of the semantic parameter values and L is the maximal rule length.

During the classification reasoning, each rule is assigned a rule weight which is commonly computed as [6]:

$$CF(\boldsymbol{A_q} \Rightarrow C_q) = c_q - c_{q,2nd} \tag{6}$$

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

$$c_{q,2nd} = max\{c(\boldsymbol{A}_q \Rightarrow C_h) | h = 1, \ldots, M;\ C_h \neq C_q\} \tag{7}$$

The classification reasoning method *Single Winner Rule* (SWR) is commonly used to classify the data pattern d_p. The winner rule $R_w \in \boldsymbol{S}$ is the rule having the maximum of the product of the compatibility or the burning $\mu_{\boldsymbol{A}_q}(d_p)$ and the rule weight $CF(\boldsymbol{A_q} \Rightarrow C_q)$ and the classified class C_w is the consequence part of this rule.

$$\mu_{\boldsymbol{A}_w}(d_p) \times CF_w = \text{argmax}\left(\mu_{A_q}(d_p) \times CF_q \middle| R_q \in \boldsymbol{S}\right). \tag{8}$$

A different given values of the semantic parameters will generate a different fuzzy partition of the feature reference space leading to a different classification performance of a specific dataset. Therefore, to get the high classification performance, a multi-objective evolutionary algorithm is applied to find the optimal semantic parameter values for generating S_0. The objectives of the applied evolutionary algorithm are the classification accuracy of the training set and the average length of the antecedent of fuzzy rule based system.

After the training process, we have a set of best semantic parameters $\boldsymbol{Л}_{opt}$ and one of the them is randomly taken, denoted as $\boldsymbol{Л}_{opt,i*}$, to generate the initial fuzzy rule set $\boldsymbol{S}_0(\boldsymbol{Л}_{opt,i*})$ which includes NR_0 fuzzy rules by using the procedure IFRG($\boldsymbol{Л}_{opt,i*}$, $\boldsymbol{P}$, NR_0, λ) mentioned above. The second phase now is to select a subset of the fuzzy rules $\boldsymbol{S}$ from $\boldsymbol{S}_0$ by applying a multi-objective evolutionary algorithm to satisfy three objectives: the classification accuracy of the training set, the number of rules of fuzzy rules in $\boldsymbol{S}$ and the average length of the antecedent of $\boldsymbol{S}$.

2.3. The proposed pure Hedge Algebras classifier

Up to now, the FRBC design methods based on HAs methodology [19, 20] try to induce the fuzzy sets based semantics of the linguistic values for the FRBCs because the authors would like to make use of the fuzzy-set-based classification reasoning method proposed in the prior researches [5-7]. This research aims to propose a hedge algebras based classification reasoning method for the FRBCs and shows the efficiency of the proposed one by the experiments on a considerable real world dataset.

In [27], the authors propose a Takagi-Sugeno-Hedge algebras fuzzy model to improve the forecast control based on the models by using the closeness of the semantically quantifying mapping values of the adjacent linguistic values instead of the membership function of each individual linguistic value. The idea is summarized as follows:

+ $v(x_i)$, $v(x_0)$ and $v(x_k)$ are the SQM values of the linguistic values x_i, x_0 and x_k with the semantic order $x_i \le x_0 \le x_k$, respectively.

+ η_i which is the closeness of $v(x_i)$ to $v(x_0)$ is defined as: $\eta_i = (v(x_k) - v(x_0)) / (v(x_k) - v(x_i))$ and η_k which is the closeness of $v(x_2)$ to $v(x_0)$ is defined as: $\eta_k = (v(x_0) - v(x_i)) / (v(x_k) - v(x_i))$, where $\eta_i + \eta_k = 1$ and $0 \le \eta_i, \eta_k \le 1$.

That idea is advanced to apply to make a new classification reasoning method for FRBCs as follows:

+ At the k_j level of the j^{th}-feature, there are the SQM values of all linguistic values $X_{(k_j)}$ with the semantic order $v(x_{j,i-1}) \le v(x_{j,i}) \le v(x_{j,i+1})$.

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

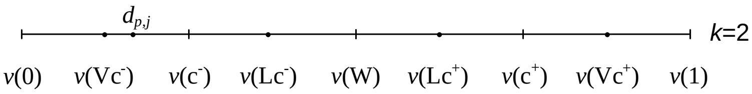

Figure 2. The SQM values of the linguistic terms with k_j = 2.

For example, Figure 2 shows the SQM values of the linguistic terms in case of k_j = 2. In this case, $\eta_{d_{p,j}} = \frac{v(Lc^-)-v(c^-)}{v(Lc^-)-d_{p,j}}$.

+ $\mu_{\boldsymbol{A}_q}(d_p)$, the burning of the data pattern d_p for the rule R_q in the formula (4) and (8), is replaced with $\eta_{\boldsymbol{A}_q}(d_p)$ which is computed as:

$$\eta_{\boldsymbol{A}_q}(d_p) = \prod_{j=1}^{n} \eta_{q,j}(d_{p,j}). \tag{9}$$

We can see that 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 measure of the closeness of the data point to the SQM value of the linguistic value.

3. EXPERIMENTAL RESULTS AND DISCUSSION

This section represents the experimental results of the pure hedge algebras classifier applying the proposed hedge algebras based classification reasoning method mentioned above. The real world datasets used in our experiments shown in the Table 1 can be found on the KEEL-Dataset repository: http://sci2s.ugr.es/keel/datasets.php.

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

The proposed pure hedge algebras classifier is compared to state-of-the-art hedge algebras based classifiers [19, 20] and some fuzzy set theory based classifiers [2, 3]. The comparison conclusions are given out based on the test results of the Wilcoxon's signed rank tests [31]. To make a comparative study, the same cross validation method is used when comparing the methods. The ten-fold cross-validation method which the designated dataset is randomly divided into ten folds, nine folds for the training phase and one fold for the testing phase, is used in all experiments. Three experiments are executed for each dataset and the results of the classification accuracy and the complexity of the classifiers are averaged out accordingly.

In order to make the comparative values, reduce the searching space in the learning processes and make sure that there is no big imbalance between $fm(c_j^-)$ and $fm(c_j^+)$, and between $\mu(L_j)$ and $\mu(V_j)$, the constraints on the semantic parameter values should be the same as the 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); $0 \le k_j \le 3$; $0.2 \le \{fm(c_j^-), fm(c_j^+)\} \le 0.8$; $fm(c_j^-) + fm(c_j^+) = 1$; $0.2 \le \{\mu(L_j), \mu(V_j)\} \le 0.8$; and $\mu(L_j) + \mu(V_j) = 1$.

The Multi-objective Particle Swarm Optimization (MOPSO) [32, 33] is used to optimize the semantic parameter values and the fuzzy rule set for FRBCs. In the optimization process of the semantic parameter values, the following parameter values of MOPSO are used: 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. In the fuzzy rule selection process, most of the algorithm parameter values are the same values of the semantic parameter optimization process, except, the number of generations is 1000; the number of initial fuzzy rules $|\boldsymbol{S_0}| = 300 \times$ *number of classes*; the maximum of rule length is 3.

3.1. The pure hedge algebras versus the existing hedge algebras based classifiers

For greater convenience, the proposed pure hedge algebras classifier is abbreviated as PHAC, the hedge algebras based classifier with the triangular [19] and trapezoidal [20] fuzzy set based semantics of linguistic values are named as HATRI and HATRA, respectively. To eliminate the possible influences of the heuristic factors on the performance of the compared classifiers, the same MOPSO algorithm with the algorithm parameters set forth above is applied to design all three classifiers.

The experimental results of the PHAC, HATRI and HATRA classifiers are shown in the Table 2, where the column #R×#C shows the complexity of the classifiers, P_{te} shows the accuracy in the testing phase, $\neq R \times C$ and $\neq P_{te}$ show the differences of the complexity and the accuracy of the comparison classifiers, respectively. By the intuitive recognition, the PHAC has better classification accuracy on 12 of 17 test datasets and the mean value of the classification accuracies is higher than the HATRI (83.65 % in comparison with 82.82 %). The mean value of the fuzzy rule base complexities of the PHAC is a bit higher than the HATRI. The PHAC has better classification accuracy on 9 of 17 test datasets and the mean value of the classification accuracies is a bit higher than the HATRA (83.65 % in comparison with 83.58 %). The mean value of the fuzzy rule base complexities of the PHAC is also a bit higher than the HATRA.

Wilcoxon's signed-rank test at level $\alpha = 0.05$ is applied to check the different significances of the classification accuracy and the complexity between the three compared classifiers. We assume that all three compared classifiers are statistically equivalent (null-hypothesis). The test result on the classification accuracy is shown in the Table 3 and the test result on the complexity is shown in the Table 4, where the VS column is the list of the classifiers which we want to compare with. The abbreviation column labels used in the Table 3 and 4: E. is Exact; A. is Asymptotic; Inte. is Interval and Conf. is Confidence. In the Table 3, since the *E. p-value* of the “PHAC vs HATRI” is less than α = 0.05, the null-hypothesis is rejected. So, the PHAC has better classification accuracy than the HATRI. The *E. p-value* of the “PHAC vs HATRA” is greater than α = 0.05, the null-hypothesis is not rejected. Furthermore, all null-hypotheses in the Table 4 are not rejected. Thus, we can statistically state that the PHAC outperforms the HATRI and the PHAC is equivalent to the HATRA.

Table 2. The experimental results of the PHAC, HATRI and HATRA classifiers.

Dataset	**PHAC**		**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	53.24	86.33	36.20	86.38	17.04	-0.05	46.50	87.15	6.74	-0.82
Bands	60.60	73.61	52.20	72.80	8.40	**0.81**	58.20	73.46	2.40	**0.15**
Bupa	203.13	71.82	187.20	68.09	15.93	**3.73**	181.19	72.38	21.94	-0.56
Dermatology	191.84	95.47	198.05	96.07	**-6.21**	-0.60	182.84	94.40	9.00	**1.07**
Glass	318.68	73.77	343.60	72.09	**-24.92**	**1.68**	474.29	72.24	**-155.61**	**1.53**
Haberman	8.82	77.11	10.20	75.76	**-1.38**	**1.35**	10.80	77.40	**-1.98**	-0.29
Heart	122.92	83.70	122.72	84.44	0.20	-0.74	123.29	84.57	**-0.37**	-0.87
Ionosphere	92.80	92.22	90.33	90.22	2.47	**2.00**	88.03	91.56	4.77	**0.66**
Iris	28.41	97.56	26.29	96.00	2.11	**1.56**	30.37	97.33	**-1.96**	**0.23**
Mammogr.	85.04	84.33	92.25	84.20	**-7.21**	**0.13**	73.84	84.20	11.20	**0.13**
Pima	52.02	76.18	60.89	76.18	**-8.87**	0.00	56.12	77.01	**-4.10**	-0.83
Saheart	56.40	72.60	86.75	69.33	**-30.35**	**3.27**	59.28	70.05	**-2.88**	**2.55**
Sonar	61.80	77.52	79.76	76.80	**-17.96**	**0.72**	49.31	78.61	12.49	-1.09
Vehicle	333.94	68.01	242.79	67.62	91.15	**0.39**	195.07	68.20	138.87	-0.19
Wdbc	47.15	95.26	37.35	96.96	9.80	-1.70	25.04	96.78	22.11	-1.52
Wine	43.20	99.44	35.82	98.30	7.38	**1.14**	40.39	98.49	2.81	**0.95**
Wisconsin	66.71	97.19	74.36	96.74	**-7.65**	**0.45**	69.81	96.95	**-3.10**	**0.24**
Mean	**107.45**	**83.65**	**104.52**	**82.82**			**103.79**	**83.58**		

Table 3. The comparison result of the accuracy of the PHAC, the HATRI and the HATRA classifiers using the Wilcoxon signed rank test at level α = 0.05.

VS	**R$^+$**	**R$^-$**	**E. *P*-value**	**A. *P*-value**	**Hypothesis**
PHAC vs HATRI	110.0	26.0	1.5258E-5	0.000267	Rejected
PHAC vs HATRA	78.0	75.0	$\geq$ 0.2	0.924572	Not rejected

Table 4. The comparison result of the complexity of the PHAC, the HATRI and the HATRA classifiers using the Wilcoxon signed rank test at level α = 0.05.

VS	R^+	R^-	E. *P*-value	A. *P*-value	Hypothesis
PHAC vs HATRI	98.0	55.0	≥ 0.2	0.297672	Not rejected
PHAC vs HATRA	44.0	109.0	≥ 0.2	1	Not rejected

3.2. The pure hedge algebras versus the fuzzy set theory based classifiers

To prove the proposed pure hedge algebras classifier outperforms the classifiers designed by the fuzzy set theory approach, its experimental results are compared to those of R. Alcalá presented in [2] and M. Antonelli presented in [3].

In [2], R. Alcalá proposed several genetic design methods of the FRBCs in such a way that the fuzzy rules are extracted from the predesigned multi-granularities (multiple partitions), then a mechanism for selecting a single granularity from the multi-granularities for each attribute is applied. The best method which a multi-objective genetic algorithm is used to tune the membership functions is the Product-1-ALL TUN.

Table 5. The experimental results of the PHAC, PAES-RCS and Product-1-ALL TUN classifiers.

Dataset	PHAC		PAES-RCS		$\neq R\times C$	$\neq P_{te}$	Product-1-ALL TUN		$\neq R\times C$	$\neq P_{te}$
	$\#R\times\#C$	T_{te}	$\#R\times\#C$	T_{te}			$\#R\times\#C$	T_{te}		
Australian	53.24	86.33	329.64	85.80	**-276.40**	**0.53**	62.43	85.65	**-9.19**	**0.68**
Bands	60.60	73.61	756.00	67.56	**-695.40**	**6.05**	104.09	65.80	**-43.49**	**7.81**
Bupa	203.13	71.82	256.20	68.67	**-53.07**	**3.15**	210.91	67.19	**-7.78**	**4.63**
Dermatology	191.84	95.47	389.40	95.43	**-197.56**	**0.04**	185.28	94.48	6.56	**0.99**
Glass	318.68	73.77	487.90	72.13	**-169.22**	**1.64**	534.88	71.28	**-216.20**	**2.49**
Haberman	8.82	77.11	202.41	72.65	**-193.59**	**4.46**	21.13	71.88	**-12.31**	**5.23**
Heart	122.92	83.70	300.30	83.21	**-177.38**	**0.49**	164.61	82.84	**-41.69**	**0.86**
Ionosphere	92.80	92.22	670.63	90.40	**-577.83**	**1.82**	86.75	90.79	6.05	**1.43**
Iris	28.41	97.56	69.84	95.33	**-41.43**	**2.23**	18.54	97.33	9.87	**0.23**
Mammogr.	85.04	84.33	132.54	83.37	**-47.50**	**0.96**	106.74	80.49	**-21.70**	**3.84**
Pima	52.02	76.18	270.64	74.66	**-218.62**	**1.52**	57.20	77.05	**-5.18**	-0.87
Saheart	56.40	72.60	525.21	70.92	**-468.81**	**1.68**	110.84	70.13	**-54.44**	**2.47**
Sonar	61.80	77.52	524.60	77.00	**-462.80**	**0.52**	47.59	78.90	14.21	-1.38
Vehicle	333.94	68.01	555.77	64.89	**-221.83**	**3.12**	382.12	66.16	**-48.18**	**1.85**
Wdbc	47.15	95.26	183.70	95.14	**-136.55**	**0.12**	44.27	94.90	2.88	**0.36**
Wine	43.20	99.44	170.94	93.98	**-127.74**	**5.46**	58.99	93.03	**-15.79**	**6.41**
Wisconsin	66.71	97.19	328.02	96.46	**-261.31**	**0.73**	69.11	96.35	**-2.40**	**0.84**
Mean	**107.45**	**83.65**	**361.98**	**81.62**			**133.26**	**81.43**		

In [3], M. Antonelli proposed a genetic design method of the FRBC namely PAES-RCS which a multi-objective evolutionary method is apply to simultaneously train the rule bases and the parameters of membership functions. The candidate rule set is generated by the C4.5 algorithm from the fuzzy partitions pre-designed for data attributes. Then, a multi-objective evolutionary process is implemented to select a set of fuzzy rules from the candidate fuzzy rule set along with the selection of a set of rules conditions for each rule. The parameters of membership functions correspond to the linguistic values are trained simultaneously in the rules and condition selection (RCS) process.

It is easy to see on the Table 5 that most of the accuracy differences between the PHAC and the Product-1-ALL TUN, and the accuracy differences between the PHAC and the PAES-RCS on 17 test datasets are positive. Review on the complexity of the classifiers, the PHAC has better complexity than the Product-1-ALL TUN on 12 of 17 test datasets and the PHAC has better complexity than the PAES-RCS on all datasets.

The comparison of the classifier accuracies and classifier complexities using Wilcoxon's signed-rank test at level $\alpha = 0.05$ are shown in the Table 6 and the Table 7, respectively. Since all *E. p-values* are less than 0.05, we can state that the PHAC outperforms the Product-1-ALL TUN and the PAES-RCS on both accuracy and complexity measures.

Table 6. The comparison result of the accuracy of the PHAC, the PAES-RCS and the Product-1-ALL TUN classifiers using the Wilcoxon signed rank test at level $\alpha = 0.05$.

VS	R^+	R^-	E. *P*-value	A. *P*-value	Hypothesis
PHAC vs PAES-RCS	153.0	0.0	1.5258E-5	0.000267	Rejected
PHAC vs Product-1-ALL TUN	139.0	14.0	0.0016784	0.002861	Rejected

Table 7. The comparison result of the complexity of the PHAC, the PAES-RCS and the Product-1-ALL TUN classifiers using the Wilcoxon signed rank test at level $\alpha = 0.05$.

VS	R^+	R^-	E. *P*-value	A. *P*-value	Hypothesis
PHAC vs PAES-RCS	153.0	0.0	1.5258E-5	0.000267	Rejected
PHAC vs Product-1-ALL TUN	124.0	29.0	0.02322	0.023073	Rejected

4. CONCLUSIONS

Fuzzy rule based systems which deal with the fuzzy information have played an important role in designing FRBCs. Hedge algebras can be regarded as an algebraic model of the semantic-order-based structure of the linguistic value domains of the linguistic variables so that hedge algebras can be used to solve the FRBC design problem with the order based semantics of linguistic values. However, the existing FRBCs designed by hedge algebras methodology generate the classifiers which still have the fuzzy rule bases with the fuzzy sets based semantics of linguistic values. This paper presents a fuzzy rule based classifier design methodology with the pure hedge algebras based semantics of linguistic values. More specifically, the fuzzy set based classification reasoning method is replaced with a hedge algebras based one in the proposed classification system model. The new classification reasoning method enables the fuzzy sets based semantics of the linguistic values in the fuzzy rule bases to be replaced with the hedge algebras based semantics. The experimental results on 17 real world datasets have shown the efficiency of the proposed classifier. By this research, we can conclude that the fuzzy rule based classifiers can be designed purely based on hedge algebras based semantics of linguistic values.

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

REFERENCES

1. Alcalá-Fdez J., Alcalá R., and Herrera F. - A Fuzzy Association Rule-Based Classification Model for High-Dimensional Problems With Genetic Rule Selection and Lateral Tuning, IEEE Transactions on Fuzzy System **19** (5) (2011) 857-872.
2. Alcalá R., Nojima Y., Herrera F., Ishibuchi H. - Multi-objective genetic fuzzy rule selection of single granularity-based fuzzy classification rules and its interaction with the lateral tuning of membership functions, Journal of Soft Computing **15** (12) (2011) 2303–2318.
3. Antonelli M., Ducange P., Marcelloni F. - A fast and efficient multi-objective evolutionary learning scheme for fuzzy rule-based classifiers, Information Sciences **283** (2014) 36–54.
4. Fazzolari M., Alcalá R., Herrera F. - 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 **24** (2014) 470–481.
5. Ishibuchi H., Yamamoto T. - Fuzzy Rule Selection by Multi-Objective Genetic Local Search Algorithms and Rule Evaluation Measures in Data Mining, Fuzzy Sets and Systems **141** (1) (2004) 59-88.
6. Ishibuchi H., Yamamoto T. - Rule weight specification in fuzzy rule-based classification systems, IEEE Transactions on Fuzzy Systems **13** (4) (2005) 428–435.
7. Ishibuchi H., Nojima Y. - Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning, International Journal of Approximate Reasoning **44** (2007) 4–31.
8. Prusty M. R., Jayanthi T., Chakraborty J. - Seetha H., Velusamy K. - Performance analysis of fuzzy rule based classification system for transient identification in nuclear power plant, Annals of Nuclear Energy **76** (2015) 63–74.
9. Rudzinski F. - A multi-objective genetic optimization of interpretability-oriented fuzzy rule-based classifiers, Applied Soft Computing, **38** (2016) 118–133.
10. Pota M., Esposito M., Pietro G. D. - Designing rule-based fuzzy systems for classification in medicine, Knowledge-Based Systems **124** (2017) 105–132.
11. Rey M. I., Galende M., Fuente M. J. - Sainz-Palmero G. I. - 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 **127** (2017) 67–84.
12. Elkanoa M., Galara M., Sanza J., Bustince H. - CHI-BD: A fuzzy rule-based classification system for Big Data classification problems, Fuzzy Sets and Systems **348** (2018) 75–101.

13. Soui M., Gasmi I., Smiti S., Ghédira K. - Rule-based credit risk assessment model using multi-objective evolutionary algorithms, Expert Systems With Applications **126** (2019) 144–157.

14. Ho N. C., Wechler W. - Hedge algebras: an algebraic approach to structures of sets of linguistic domains of linguistic truth variables, Fuzzy Sets and Systems **35** (3) (1990) 281-293.

15. Ho N. C., Wechler W. - Extended hedge algebras and their application to fuzzy logic, Fuzzy Sets and Systems **52** (1992) 259–281.

16. Ho N. C, Nam H. V., Khang D. T., Le H.C. - Hedge Algebras, Linguistic-valued logic and their application to fuzzy reasoning, Internat. J.Uncertain. Fuzziness Knowledge-Based Systems **7** (4) (1999) 347–361.

17. Ho N. C., Long N. V. - Fuzziness measure on complete hedges algebras and quantifying semantics of terms in linear hedge algebras, Fuzzy Sets and Systems **158** (2007) 452-471.

18. Ho N. C. - A topological completion of refined hedge algebras and a model of fuzziness of linguistic terms and hedges, Fuzzy Sets and Systems **158** (2007) 436–451.

19. Ho N. C., Pedrycz W., Long D. T., Son T. T. - A genetic design of linguistic terms for fuzzy rule based classifiers, International Journal of Approximate Reasoning **54** (1) (2013) 1-21.

20. Ho N. C., Son T. T., Phong P. D. - Modeling of a semantics core of linguistic terms based on an extension of hedge algebra semantics and its application, Knowledge-Based Systems **67** (2014) 244–262.

21. Ho N. C., Thong H. V., Long N. V. - A discussion on interpretability of linguistic rule based systems and its application to solve regression problems, Knowledge-Based Systems **88** (2015) 107–133.

22. Ho N. C., Dieu N. C., Lan V. N. - The application of hedge algebras in fuzzy time series forecasting, Vietnam Journal of Science and Technology **54** (2) (2016) 161-177.

23. Lan L. V. T., Han N. M., Hao N. C. - An algorithm to build a fuzzy decision tree for data classification problem based on the fuzziness intervals matching, Journal of Computer Science and Cybernetics **32** (4) (2016) 367-380.

24. Son T. T., Anh N. T. - Partition fuzzy domain with multi-granularity representation of data based on hedge algebra approach, Journal of Computer Science and Cybernetics **34** (1) (2018) 63–75.

25. Tung H., Thuan N. D., Loc V. M. - The partitioning method based on hedge algebras for fuzzy time series forecasting, Vietnam Journal of Science and Technology **54** (5) (2016) 571-583.

26. Ho N. C., Lan V. N., Trung T. T., Le B. H. - Hedge-algebras-based fuzzy controller: application to active control of a fifteen-story building against earthquake, Vietnam Journal of Science and Technology **49** (2) (2011) 13-30.

27. Lan V. N., Ha T. T., Lai L. K., Duy N. T. - The application of the hedge algebras in forecast control based on the models, In Proceedings of The 11st National Conference on Fundamental and Applied IT Research, Hanoi, Vietnam (2018) 521-528.

28. Le B. H., Anh L. T., Binh B. V. - Explicit formula of hedge-algebras-based fuzzy controller and applications in structural vibration control, Applied Soft Computing **60** (2017) 150–166.
29. Huy N. H., Ho N. C., Quyen N. V. - Multichannel image contrast enhancement based on linguistic rule-based intensificators, Applied Soft Computing Journal **76** (2019) 744–762.
30. Long D. T. - A genetic algorithm based method for timetabling problems using linguistics of hedge algebra in constraints, Journal of Computer Science and Cybernetics **32** (4) (2016) pp. 285—301.
31. Demˇsar J. - Statistical Comparisons of Classifiers over Multiple Data Sets, Journal of Machine Learning Research **7** (2006) 1–30.
32. Phong P. D., Ho N. C., Thuy N. T. - 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 Proceedings of The 10th IEEE RIVF International Conference on Computing and Communication Technologies (RIVF-2013), Hanoi, Vietnam (2013) 12 – 17.
33. Maximino S. L. - Multi-Objective Optimization using Sharing in Swarm Optimization Algorithms, Doctor thesis, School of Computer Science, The University of Birmingham (2006).