ICIC Express Letters
Volume **10**, Number **6**, June **2016**

 ISSN 1881-803X
pp. **1461–1468**

A NEW DATA CLASSIFIER BASED ON FUZZY RULES

Yuliang Cai[1], Xiaodong Liu[1] and Xianchang Wang[2]

[1]The Research Center of Information and Control
Dalian University of Technology
No. 2, Linggong Road, Ganjingzi District, Dalian 116024, P. R. China
ylcaisun@hotmail.com; xdliuros@dlut.edu.cn

[2]School of Sciences
Dalian Ocean University
No. 52, Heishijiao Street, Shahekou District, Dalian 116023, P. R. China
wxcixll@sohu.com

Received September 2015; accepted December 2015

Abstract. *This study proposes a two-stage classification algorithm which extracts fuzzy rules directly from samples. First, a forward greedy fast attribute reduction algorithm based on the neighborhood rough set is introduced to select r attributes of importance from the original attribute set. This algorithm uses the threshold vector instead of a single threshold and then exploits properties of the positive region and neighborhood relation matrices generated by the threshold vector to calculate the significance of attributes. After selecting attributes, a classification algorithm is proposed to generate fuzzy rules from the selected attribute subset. When determining the membership functions, we stipulate that the total number of the fuzzy functions defined for each feature is equal to that of the classes. Finally, a series of comparative experiments is conducted with six traditional classification approaches on six UCI data sets. The experimental results show that our method exhibits the highest average classification accuracy. Moreover, our approach uses less rules on average among all methods.*

Keywords: Data classifier, Attribute reduction, Neighborhood rough set, Threshold vector, Fuzzy function

1. **Introduction.** With the development of information technology, amounts of redundant data exist in practical applications. As we all know, attribute reduction can solve the problem of data redundancy. Among existing feature selection algorithms, most of them [1-3] usually set the same neighborhood size for all properties, which will definitely affect the results of attribute reduction. To overcome this drawback, we provide a forward greedy fast attribute reduction algorithm based on the neighborhood rough set (NRS_FS_FAST) in this study. This algorithm uses properties of the positive region and neighborhood relationship matrices to calculate the importance of attributes.

Data classifiers based on fuzzy rules have been widely discussed and applied in pattern recognition and machine learning. The core problem in the design of classifiers is the construction of rule-base. There are many methods proposed to generate the rule-base. For instance, Gou et al. put forward clustering approaches [4]. Neural network methods and genetic algorithms were proposed in [5,6] respectively. And some data mining techniques also appeared in [7], etc. However, the semantics and understanding of these methods are poor. So a method to generate fuzzy rules from the selected attribute subset is proposed. This new method consists of four steps. *Step* 1: Formation of Fuzzy Numbers; *Step* 2: Generate Fuzzy Rules from the Given Data Pairs; *Step* 3: Simplify the Rule-Base; *Step* 4: Determine a Mapping Based on the Combined Fuzzy Rule-Base.

The rest of this paper is organized as follows. Section 2 describes the neighborhood rough set model. In Section 3, a forward greedy fast attribute reduction algorithm is

introduced. Section 4 presents how to generate fuzzy rules. In Section 5, several numerical experiments are carried out to verify the superiority and interpretability of the proposed algorithm. Section 6 concludes this paper.

2. **The Introduction of Neighborhood Rough Set Model.** This section merely describes a few concepts of the neighborhood rough set model. More details can be found in [8].

Definition 2.1. *Given a set of samples $U = \{x_1, x_2, \ldots, x_n\}$. A is a feature set to describe U, and D is a decision attribute set. If A can generate a group of neighborhood relations on U, we call $NDT = \langle U, A, D\rangle$ a neighborhood decision system.*

Definition 2.2. *Given a neighborhood decision system $NDT = \langle U, A, D\rangle$, $\forall B \subseteq A$, the dependency degree of the decision attribute set D with regard to the condition attribute set B is defined as:*

$$\gamma_B(D) = \frac{|POS_B(D)|}{|U|} \tag{1}$$

where $POS_B(D)$ is the positive region of the decision attribute set D.

Definition 2.3. *Given a neighborhood decision system $NDT = \langle U, A, D\rangle$, $\forall a \in A - B$, the significance of attribute a to the attribute set B is defined as:*

$$SIG(a, B, D) = \gamma_{B\cup a}(D) - \gamma_B(D) \tag{2}$$

3. **A Forward Greedy Fast Attribute Reduction Algorithm.**

Definition 3.1. *Given a neighborhood decision system $NDT = \langle U, A, D\rangle$, $SD = \{SD_1, SD_2, \ldots, SD_m\}$ is a set of standard deviation of each attribute. m is the number of attributes. The relationship matrix of neighborhood relation N_i of the attribute i on U is defined as:*

$$M_{(N_i)} = (r_{p,q})_{n\times n} \tag{3}$$

where

$$r_{p,q} = \begin{cases} 1, & \Delta(x_p, x_q) \le \delta_i \\ 0, & others \end{cases}, \quad 1 \le p \le n,\ 1 \le q \le n \tag{4}$$

n is the number of samples. $\delta_i = SD_i/L$ is regarded as the threshold of neighborhood size, and L is a given parameter.

Setting the same neighborhood size for all properties can affect the accuracy of the feature selection. This study provides a neighborhood setting method based on the distribution of the data itself. This method exploits the threshold vector instead of a single threshold (Each dimension of vector corresponds to a property, and the values of vector depend on the ratio of the standard deviation of each attribute data and parameter L.) and then makes use of properties of the positive region and neighborhood relationship matrices generated by the threshold vector to calculate the importance of attributes. Table 1 displays the process of NRS_FS_FAST algorithm.

4. **Generating Fuzzy Rules from Numerical Data.** Fuzzy rules with linguistic interpretation are corresponding with "IF-THEN" fuzzy rules. In this paper, the process of generating fuzzy rules consists of the following four steps. First of all, let us look at the way of forming fuzzy numbers.

Step 1 – Formation of Fuzzy Numbers

The data can be represented in the form $X = [x_{ij}]_{n\times m}$, where n is the number of samples, and m is the number of attributes. Let f_j $(j = 1, 2, \ldots, m)$ denote the j-th feature of X, $x_i = [x_{i1}, x_{i2}, \ldots, x_{im}]$ $(i = 1, 2, \ldots, n)$ represent the i-th pattern, and $x_{i,j}$ be the j-th feature value of the i-th pattern x_i. $C = \{1, 2, \ldots, c\}$ is a set of class labels, where c is the number of classes. X contains c equivalence classes X_l $(l = 1, 2, \ldots, c)$.

TABLE 1. NRS_FS_FAST algorithm

Algorithm: NRS_FS_FAST

Input: Neighborhood decision system $NDT=\langle U, A, D\rangle$, parameter L, and r
L: a parameter of controlling the neighborhood size.
r: the number of attributes in *Red*.
Output: *Red*: the attribute subset with r attributes.

1. Get neighborhood relation matrix: $NA = \mathbf{GetNeighborRelation}(Sample, L)$
2. Initialize the $red = \varnothing$, $pos = \varnothing$.
   red: the attribute reduction set. *pos*: the positive region of attribute reduction.
3. Select Attributes: $Red = \mathbf{Select\ Attributes}(Sample, L, NA, r)$
   for each attribute $a_i \in A - red$
   (1) $SIG(a_i, red, D) = \gamma_{red\cup a_i}(D) - \gamma_{red}(D)$.
   (2) $N_{red\cup a_i}(D) = \bigcup_{i=1}^{N} N_{red\cup a_i} X_i$. $X_1, X_2, \ldots, X_N$ are equivalence classes.
   end
4. Find out the attribute a_k with the largest importance degree and its positive region $pos = N_{red\cup a_k}(D)$. $SIG(a_k, red, D) = \max_i\{SIG(a_i, red, D)\}$.
5. **if** $SIG(a_k, red, D) > 0$, and the number of attributes in *red* is less than r
   (1) $red = red \cup a_k$.
   (2) delete a_k from A, and delete the *pos*-th rows and the *pos*-th columns of *NA*.
   (3) go to 3.
   else $Red = red$. **end**

We assume that the number of fuzzy membership functions defined for each feature is equal to c (the number of classes). Let $f_{k,j}$ denote the k-th ($k = 1, 2, \ldots, c$) fuzzy function of the j-th feature f_j. There are many kinds of membership functions, such as Gaussian function, triangular function and trapezoidal function. This paper adopts the combination of triangular function and trapezoidal function. The core of fuzzy membership functions lies in the values of parameters $ms_{k,j}$ ($k = 1, 2, \ldots, c$), and these values are taken as the mean value of all patterns falling in the k-th equivalence class X_k in this paper. The values of $\left\{ms'_{k,j}\right\}$ ($k = 1, 2, \ldots, c$) can be obtained by Formula (5), and the values of $ms_{k,j}$ ($k = 1, 2, \ldots, c$) are the same values after sorting $ms'_{k,j}$ in an ascending order.

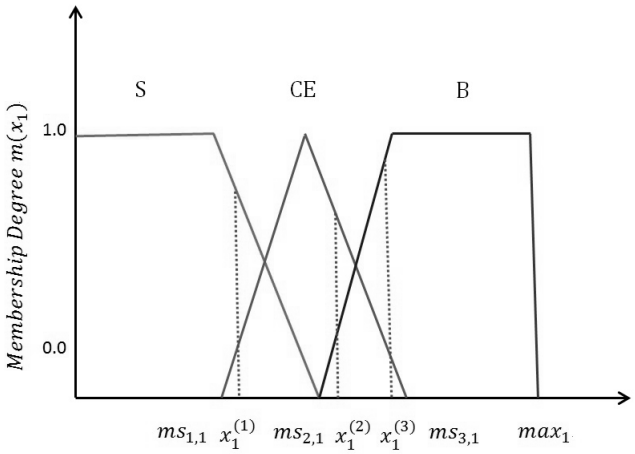
$$ms'_{k,j} = \frac{\sum x_i \in X_k x_{ij}}{|X_k|} \tag{5}$$

Step 2 – Generate Fuzzy Rules from the Given Data Pairs

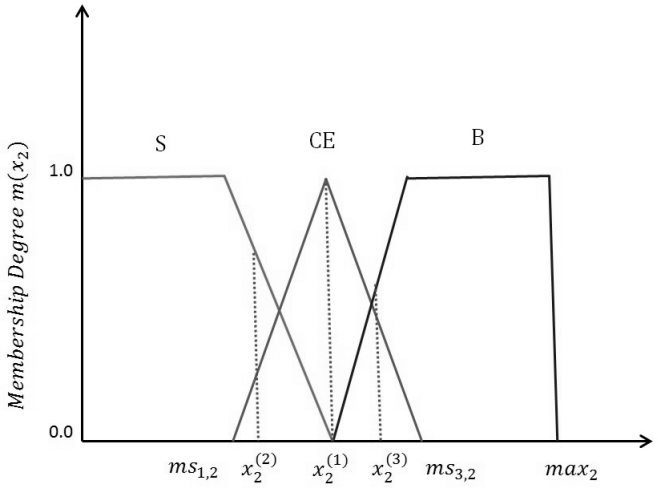
This paper chooses a simple case with three-input, one-output $\left\{\left(x_1^{(1)}, x_2^{(1)}, x_3^{(1)}; y^{(1)}\right), \left(x_1^{(2)}, x_2^{(2)}, x_3^{(2)}; y^{(2)}\right), \ldots\right\}$ to illuminate the basic view of our new approach. The specific steps of generating rules are as follows:

(1) Determine membership degrees to the given condition attributes $\left(x_1^{(i)}, x_2^{(i)}, x_3^{(i)}\right)$ and the decision attribute $y^{(i)}$ in different fuzzy regions. For instance, $x_1^{(1)}$ in Figure 1(a) has degree 0.8 in S, degree 0.2 in CE, and degree 0 in B. Similarly, $x_2^{(1)}$ in Figure 1(b) has degree 1 in CE, and degree 0 in other regions.

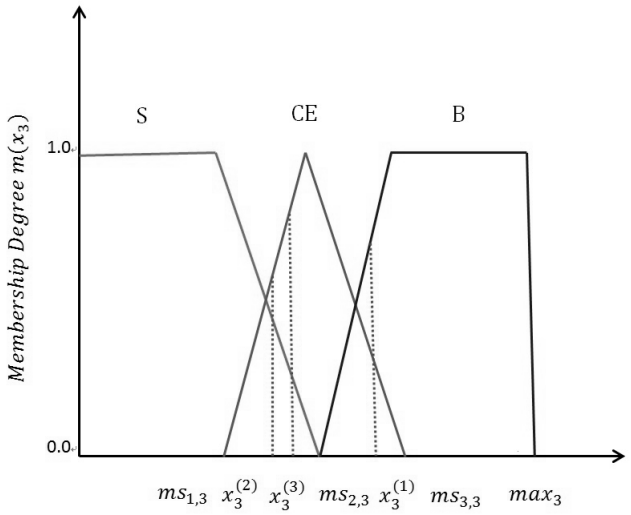
(2) Assign the given condition attributes $\left(x_1^{(i)}, x_2^{(i)}, x_3^{(i)}\right)$ to the region with maximum degree. For example, $x_1^{(1)}$ in Figure 1(a) is allocated to S area, and $x_2^{(1)}$ in Figure 1(b) is assigned to CE area.



(a) Attribute x_1



(b) Attribute x_2



(c) Attribute x_3

FIGURE 1. The triangular and trapezoidal functions formed by fuzzy numbers

(3) Obtain rules from the data pairs. Supposing that the class label of $y^{(1)}$ is 1, we can get a rule: $\left[x_1^{(1)}(0.8 \text{ in } S, \max), x_2^{(1)}(1 \text{ in } CE, \max), x_3^{(1)}(0.6 \text{ in } B, \max);\ y^{(1)} \text{ is } Class\ 1\right] \Rightarrow$ *Rule* 1. IF x_1 is S, x_2 is CE and x_3 is B; THEN y is *Class* 1 $\Leftrightarrow$ ([S, CE, B; *Class* 1]).

So, a rule-base extracted from all samples can be obtained by this way.

Step 3 – *Simplify the Rule-Base*

This study provides two processes in simplifying the rule-base. 1) Assign a degree to each rule of the rule-base, and remove the repeating and conflicting rules. 2) Use pruning algorithm to simplify the rule-base obtained by the first procedure.

First, we define a confidence degree to each rule by counting the number of data pairs with the same rules. For instance, the confidence degree of rule i is defined as:

$$D(rule\ i) = m/n \tag{6}$$

where n is the number of samples, and m is the number of samples covered by rule i.

We assign degrees to all rules by Formula (6), and accept merely the rule that has the biggest confidence degree. In this way not only the conflict problem is resolved, but also the number of rules is greatly reduced.

Then, we use the following pruning algorithm to simplify the rule-base again.

1) Delete each rule from the rule-base respectively, classify the training samples, and calculate the classification accuracy using the remaining rules. 2) Compare all accuracies and delete the rule whose remaining rules can get the maximum accuracy. 3) Repeat 1 and 2 until that removing any rule cannot improve the training accuracy.

Step 4 – *Determine a Mapping Based on the Combined Fuzzy Rule-Base*

After simplifying the rule-base, we need to determine the class labels of the new samples. First of all, generating fuzzy rules for the new samples by *Step* 2 is necessary. Taking a new sample $\left(x_1^{(3)}, x_2^{(3)}, x_3^{(3)}; y^{(3)}\right)$ for example as shown in Figure 1, we can obtain the result: $\left[x_1^{(3)}(0.9 \text{ in } B, \max), x_2^{(3)}(0.6 \text{ in } B, \max), x_3^{(3)}(0.7 \text{ in } CE, \max)\right] \Rightarrow (B, B, CE)$.

Next, we search the final rule-base to find fuzzy rules whose "IF" parts are same with (B, B, CE). However, there is a possibility that the final rule-base does not have the "IF" part (B, B, CE). If so, we only consider the combination of two attributes $\left(x_1^{(3)}, x_2^{(3)}, **\right)$, $\left(x_1^{(3)}, **, x_3^{(3)}\right)$ and $\left(**, x_2^{(3)}, x_3^{(3)}\right)$. The corresponding "IF" parts are $(B, B, **)$, $(B, **, CE)$ and $(**, B, CE)$ respectively. This study defines a decision degree to each rule. For example, the decision degree of rule r is defined as:

$$P(rule\ r) = \prod m\left(x_j^{(i)}\right) \times D\ (rule\ r) \tag{7}$$

where $m\left(x_j^{(i)}\right)$ denotes the maximal membership degree of the j-th attribute of pattern x_i in the corresponding fuzzy sets.

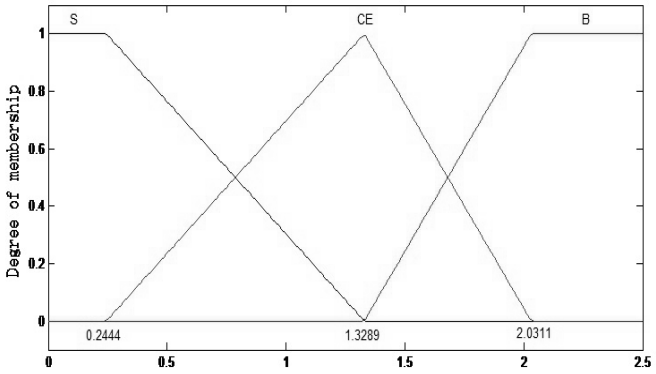
For instance, rule i([B, B, S; *Class* 3]) in the final rule-base accord with $(B, B, **)$. The decision degree of rule i is $P(rule\ i) = m\left(x_1^{(3)}\right) m\left(x_2^{(3)}\right) \times D(rule\ i) = 0.9 \times 0.6 \times D(rule\ i)$.

We assign degrees to relevant rules by Formula (7), and select the rule with the maximum decision degree to determine the class label $y^{(3)}$. However, there is still a small possibility that the final rule-base might not have any "IF" part corresponding with the combination of two attributes either. In this case, we only consider the combination of one attribute $\left(x_1^{(3)}, **, **\right)$, $\left(**, x_2^{(3)}, **\right)$, $\left(**, **, x_3^{(3)}\right)$. The corresponding "IF" parts are $(B, **, **)$, $(**, B, **)$, and $(**, **, CE)$ respectively. Then, we adopt the method mentioned above to seek the class label. The worst thing is that we could not find any "IF" part in the final rule-base corresponding with the combination of one attribute either. On this occasion, we suppose that this sample is an outlier that can be ignored.

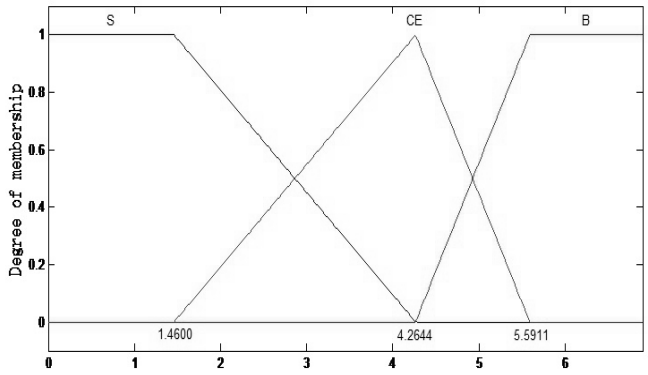
In this way, we get a mapping: $f\colon \{x_1, x_2, x_3\} \Rightarrow y$.

5. **Simulation Experiments.** In this experiment, a ten-fold cross validation is carried out. And a series of comparative experiments is conducted with six traditional classification approaches on six UCI data sets. The values of the parameters for all the algorithms are set to their defaults in the toolkit. This study uses the Euclidean distance as the distance function, and the parameters are set to be $L = 2$, $r = 3$ separately.

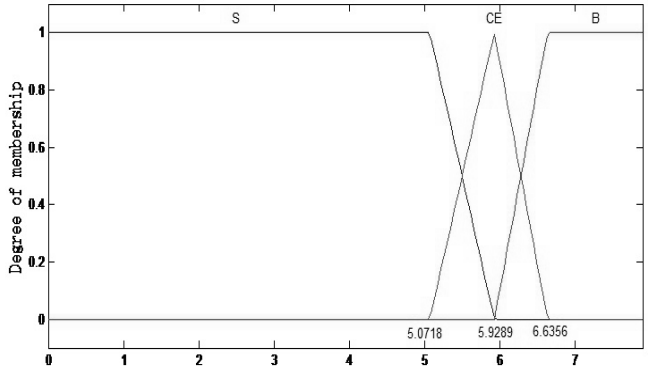
We employ the Iris data to illustrate the performance of our classifier. This data set has four features: sepal length and width, petal length and width.



(a) Petal width



(b) Petal length



(c) Sepal length

FIGURE 2. The membership functions of Iris data set

First, the NRS_FS_FAST algorithm in Table 1 which is proposed in Section 3 is used to select an attribute subset with 3 attributes. They are petal width, petal length and sepal length respectively. Next, a rule-base is extracted from the training samples described by the new attribute subset. The membership functions are shown in Figure 2. And the fuzzy numbers of each function can be computed by Formula (5). Then, we remove the repeating and conflicting rules of rule-base by the method in Step 3 of Section 4. After first simplification, the average number of rules is 11.4, and the average classification accuracy on the testing set is 89.33%. We simplify the rule-base through the pruning algorithm again. The average number of rules decreases to 5.3, and the average classification accuracy on the testing set increases to 98%.

The classification accuracies of six classification methods on six data sets are reported in Table 2 (The last line of it is the average of the each method on all data sets). It shows that the proposed algorithm can produce the highest average classification accuracy.

TABLE 2. The classification accuracy of different classification methods, and the best scores are indicated in boldface. (PM means Proposed Method)

	C4.5	JRip	Dtable	PART	OneR	NNge	PM
Wine	**0.9601**	0.9487	0.8869	0.9542	0.7460	0.9487	0.9441
Haberman	0.7061	0.7152	0.7289	0.7124	**0.7388**	0.6536	0.7221
Fertility	0.8716	0.8616	0.8807	0.8716	0.8807	0.8625	**0.8807**
Iris	0.9533	0.9533	0.9467	0.9467	0.9267	0.9600	**0.98**
Transfusion	0.7794	**0.7847**	0.7501	0.7647	0.7620	0.7059	0.7834
Liver	0.6862	0.6745	0.6028	0.6691	0.5768	0.6458	**0.6873**
Average	0.8261	0.8230	0.7994	0.8198	0.7718	0.7961	**0.8329**

We also compare the number of rules returned by the proposed algorithm with six traditional sorting techniques. By analyzing the results presented in Table 3, we can get the following results:

(1) The last line of Table 3 shows that the average number of rules of our algorithm is 4.4667, which is less than the numbers obtained by C4.5, Dtable, OneR and NNge algorithms. So our algorithm has fewer rules among all methods.

(2) The rule-base constructed by the proposed algorithm is simple in structure. It can describe each category by a small number of rules, and it has strong semantics and understandability.

TABLE 3. The number of rules of different classification methods, and the best scores are indicated in boldface. (PM means Proposed Method)

	C4.5	JRip	Dtable	PART	OneR	NNge	PM
Wine	5.1	**3.9**	19.3	4.0	4.7	9.6	9.2
Haberman	3.1	6.7	**1.5**	3.2	4.3	74.2	3.0
Fertility	2.9	1.8	**1.0**	5.1	5.6	17.4	2.0
Iris	4.8	3.9	3.5	3.5	**3.0**	9.8	5.3
Transfusion	6.7	3.0	**1.3**	4.3	3.0	161.6	3.0
Liver	10.0	3.3	**2.6**	4.5	6.5	100.6	4.3
Average	5.4333	**3.7667**	4.8667	4.10	4.5167	62.2	4.4667

6. **Conclusion.** This paper describes the neighborhood rough set model at first. Then, a forward greedy fast attribute reduction algorithm is introduced. After selecting attributes, we extract a fuzzy rule-base from the new attribute subset, and then cut down the conflicting and repeating rules through two reduction processes. The experimental results indicate that our method exhibits the highest average classification accuracy with

fewer rules. On the basis of existing research results of this paper, there are still many ways to further study. Firstly, whether different value of L has significant impact on importance of attribute or not is not clear. Secondly, we can also use other membership functions in Section 4. Thirdly, other distance functions can be utilized.

Acknowledgment. This work is supported by the Natural Science Foundation of China under Grant 61175041. And the authors would like to thank the anonymous referees for their comments and suggestions.

REFERENCES

[1] D. Wang, H. Zhang and R. Liu, t-Test feature selection approach based on term frequency for text categorization, *Pattern Recognition Letters*, vol.45, pp.1-10, 2014.

[2] Y. Lin, J. Li, P. Lin and G. Lin, Feature selection via neighborhood multi-granulation fusion, *Knowledge-Based Systems*, vol.67, pp.162-168, 2014.

[3] D. Feng, F. Chen and W. Xu, Supervised feature subset selection with ordinal optimization, *Knowledge-Based Systems*, vol.56, pp.123-140, 2014.

[4] J. Gou, F. Hou and W. Chen, Improving Wang-Mendel method performance in fuzzy rules generation using the fuzzy C-means clustering algorithm, *Neurocomputing*, vol.151, no.3, pp.1293-1304, 2015.

[5] J. C. Figueroa-Garcia and C. M. Ochoa-Rey, Rule generation of fuzzy logic systems using a self-organized fuzzy neural network, *Neurocomputing*, vol.151, no.3, pp.955-962, 2015.

[6] D. Meng and Z. Pei, Extracting linguistic rules from data sets using fuzzy logic and genetic algorithms, *Neurocomputing*, vol.78, no.1, pp.48-54, 2012.

[7] P. Kromer and S. Owaisc, Towards new directions of data mining by evolutionary fuzzy rules and symbolic regression, *Computers & Mathematics with Applications*, vol.66, no.2, pp.190-200, 2013.

[8] Q. He, Z. xie and Q. Hu, Neighborhood based sample and feature selection for SVM classification learning, *Neurocomputing*, vol.74, pp.1585-1594, 2011.