

# Evaluation of the Quality of Attribute Reductions Obtained Using Rough Set by Margin Criteria

Ming Yang Yuxiang Qiu Genlin Ji

Department of Computer Science, Nanjing Normal University, Nanjing 210097, P.R. China

## Abstract

As a *filter model*, rough set-based methods are one of effective attribute reduction (also called feature selection) that preserve the meaning of the features. In rough set theory, researchers mainly focus on extension of the classical rough set model (also called Pawlak model for short) and development of efficient attribute reduction algorithms. However, very little work has been done for aiming on the evaluation of the quality of attribute reduction, except that employing the cardinality of the given attribute subset  $P$  and so-called approximation quality of  $P$  or other equivalent criteria induced by Pawlak model. Although this discrimination strategy is simple and effective in most cases, it is very difficult to guarantee the selected attribute reduction(s) from lots of attribute reductions are the best or Top  $n$ , especially for the case containing many attribute reductions with same cardinality and approximation quality. Therefore, in this paper, we incorporate margin criteria into the proposed evaluation mechanism for guaranteeing the effectiveness of the selected attribute subsets, since margin, originally designed for binary classification problem using support vector machine, can actually determine the generalization ability. Also, an improved discernibility function-based algorithm is proposed. To further test the effectiveness of the proposed method, the algorithm of this paper is experimented using *UCI* benchmark datasets. Preliminary experimental results show that the attribute reductions with larger margin have better or comparable performance than those with relatively small margin for all reducts with same cardinality. Thus, our newly developed method can, in most cases, get more effective attribute subsets.

**Keywords:** Rough set, Attribute reduction, Margin

## 1. Introduction

Attribute reduction (also called feature selection) has been viewed as one of the most key problems in the field of the machine learning [1-3]. It is defined as a process of selecting relevant attributes (features) out of the larger set of candidate attributes. The relevant attributes are defined as attribute subset that has the same classification capability with the entire attributes. Since attribute reduction reduces the dimensionality of the data, it enables the learning algorithms to operate more effectively and rapidly. Moreover, in some cases, classification performance can be improved. In other instances, the required classifier is more compact and can be easily interpreted. Hence, the studies for attribute reduction are of great value.

Generally, supervised attribute reduction methods can be categorized into two classes: the *filter model* and the *wrapper model* [1,2]. In the *wrapper model* the attribute reduction methods tries to directly optimize the performance of a specific predictor (post-analysis algorithm). Along this, the predictor generalization performance (e.g. by cross validation) needs to be estimated for the selected attribute subset in each step. So, high computational cost is its main disadvantage.

In contrast, in the *filter model* the reduction is done as a data preprocessing, without attempting to optimize the performance of any specific predictor directly. Usually, the quality of an attribute subset is evaluated by employing the so-called fitness function, the higher the fitness value the more effective the attribute subset for two attribute subsets with the same cardinality.

As a *filter model*, rough set approach is one of effective attribute reduction methods that can preserve the meaning of the attributes [3]-[5], it has been widely applied in many fields of machine learning, such as pattern recognition [6], web mining [7] and other application fields [8]. In rough set theory, both extension of the proposed models and

development of more efficient attribute reduction algorithms are two main research issues. Some novel extended models [9]-[11] and many rough set-based attribute reduction algorithms[12]-[16] have been introduced to date. Generally, the rough set-based attribute reduction methods can roughly be divided into the following categories: *discernibility matrix-based* methods[12]-[14], *positive region-based* methods[15], *heuristic strategy-based* methods[8] and other *evolutionary* methods [16]. It is well known that the main objective of rough set-based attribute reduction methods is to provide those attribute subsets with the best or top  $n$  quality for post-analysis algorithms. To our best knowledge, very little work has been done for evaluating the quality of attribute subsets obtained by the proposed attribute reduction algorithms, except that employing the cardinality of the given attribute subset  $P$  and so-called approximation quality of  $P$  or other equivalent criteria induced by Pawlak model. However, in real applications, by this classical evaluating strategy, it is very difficult to guarantee the selected attribute reduction(s) from lots of attribute reductions are the best or Top  $n$ , especially for the case that many attribute reductions have same cardinality and approximation quality. Therefore, selecting the more effective attribute subsets from lots of attribute reductions remains very important and also is our motivation of this study.

In essence, the so-called "the more smaller the cardinality of an attribute reduction obtained by the training set, the more better its quality" means the attribute reduction with small cardinality should have a low generalization error. But this strategy lacks a theoretical foundation, hence it need to be compensated by other strategy. From a theoretical point of view, margin, originally designed for binary classification problems in the *Support Vector Machine*(SVM)[17], is a measure of the generalization performance of a learning machine[18]-[19], it is the so-called "the more larger the margin of the classification, the more lower the generalization error of the learning machine". Therefore, combining maximum margin criteria and the classical rough set-based valuation mechanism has the following advantages: providing an effective measure for ranking of the attribute subsets; enhancing the performance of the conventional rough set model and other *filter models*.

In this study, the margin of a object with respect to a given attribute subset is extended and redefined, and then we present an improved attribute reduction algorithm by embedding a maximum

margin criteria in classical discernibility function-based attribute reduction framework. Further, the quality of selected attribute subsets are tested on standard *UCI* benchmark datasets by employing both k-nearest neighbor (K-NN)[20]-[21] classifier and well known fisher discriminant analysis with kernels(KFDA) classifier[22].

The rest of this paper is organized as follows: In Section 2, some basic concepts on rough set theory are briefly introduced; In Section 3, the proposed margin is extended and redefined; A new discernibility function-based attribute reduction algorithm is developed in Section 4; Some experimental comparisons are presented in Section 5; Finally, Section 6 gives our conclusions and several issues for future work.

## 2. Preliminaries

This section introduces some essential definitions from rough set theory that are used for attribute reduction, more details and formal definitions about the rough set theory can be found in [3][8][14].

**Definition 1.** (decision table[14]-[15]) In rough set theory, a data set can be formally described using a decision table or an information system. A decision table is denoted as  $DT = \langle U, Q, V, f \rangle$ , where  $U = \{x_1, x_2, \dots, x_n\}$  is a non-empty finite set of objects or cases, called *universe*,  $Q$  is a non-empty finite set of attributes,  $Q = C \cup D$ , where  $C$  is the condition attributes and  $D$  is the decision attributes,  $C \cap D = \{\}$ . In this paper,  $D = \{d\}$  is a singleton set, where  $d$  is the class attribute that denotes classes of objects.  $V = \bigcup_{a \in Q} V_a$ , and  $V_a$  is the domain of the attribute  $a$ .  $f : U \times Q \rightarrow V$  is a total function such that  $f(x_i, a) \in V_a$  for every  $a$ . Throughout this paper,  $\{\}$  denotes the empty set, and  $card(X)$  or  $|X|$  denotes the function that returns the cardinality of the argument set  $X$ .

With any  $P \subseteq Q$  there is an associated equivalence relation  $IND(P)$ :

$$IND(P) = \{(x, y) \in U^2 | \forall a \in P, a(x) = a(y)\}.$$

The partition of  $U$ , generated by  $IND(P)$  is denoted as  $U/IND(P)$ . If  $(x, y) \in IND(P)$ , then  $x$  and  $y$  are indiscernible by attributes from  $P$ . The equivalence classes of the  $P$ -indiscernibility relation are denoted  $[x]_P$ .

**Definition 2.** (lower-approximation, positive region and approximation quality[14]-[15]) Given a *decision information table*  $S = \langle U, Q, V, f \rangle$ , for any subset  $X \subseteq U$  and indiscernibility relation  $IND(B)$ , the  $B$  lower-approximation of  $X$

is defined as:  $B_-(X) = \{x|[x]_B \subseteq X\}$ . For  $P \subseteq Q, B \subseteq Q$ , the  $P$  positive region of  $B$  is defined as  $Pos_P(B) = \bigcup_{X \in U/IND(B)} P_-(X)$ ; the  $P$ -approximation quality of  $B$  is defined as:

$$\gamma_P(B) = |Pos_P(B)|/|U|, \text{ namely}$$

$$\gamma_P(B) = \left| \bigcup_{X \in U/IND(B)} P_-(X) \right|/|U|.$$

**Definition 3.** (attribute reduction[4]-[6][22])

Given a decision table  $S = \langle U, C \cup D, V, f \rangle$ . An attribute reduction  $R$  is defined as a subset of the conditional attributes  $C$  such that  $\gamma_R(D) = \gamma_C(D)$  and  $\gamma_S(D) \neq \gamma_C(D)$  holds for any  $S \subset R$ . For short, an attribute reduction and the set of attribute reduction is denoted by *reduct* and *reducts*, respectively.

**Definition 4.** (discernibility matrix, discernibility function[3][8]) Given a decision table  $DT = \langle U, C \cup D, V, f \rangle$ , a discernibility matrix  $DM$  is defined as an  $n \times n$  matrix of  $DT$  with the  $(i, j)$  th entry  $m_{i,j}$  given by:

$$m_{i,j} = \{a \in C : f(x_i, a) \neq f(x_j, a)\} \text{ for } f(x_i, D) \neq f(x_j, D) \text{ and } \{\} \text{ otherwise;} \quad (1)$$

A discernibility function  $f_{DM}$ , induced by formula (1), is a Boolean function of  $m$  boolean variables  $a_1^*, a_2^*, \dots, a_m^*$  (corresponding to the attributes  $a_1, a_2, \dots, a_m$ ) defined as below:

$$f_{DM}(a_1^*, a_2^*, \dots, a_m^*) = \bigwedge \{ \bigvee m_{i,j}^* | 1 \leq j \leq i \leq card(U), m_{i,j} \neq \{\} \}; \text{ where } m_{i,j}^* = \{a^* | a \in m_{i,j}\}. \quad (2)$$

An attribute subset  $R$  of  $C$  is an *attribute reduction* iff  $R \cap m_{i,j} \neq \{\}$  holds for each  $m_{i,j} \in DM(m_{i,j} \neq \{\})$ , and for every  $S \subset R$ ,  $\exists m_{i,j} \in DM(m_{i,j} \neq \{\})$  s.t.  $S \cap m_{i,j} = \{\}$ . Moreover, by solving (2), all reducts of a decision table can be determined. In addition, any two reducts  $R, S$ , the quality of reducts is generally evaluated as follows: if  $card(R)$  is less than  $card(S)$ , then  $R$  is better than  $S$ ; if  $card(R)$  is equal to  $card(S)$ , then the quality of  $R$  is same as that of  $S$ . As mentioned in Section 1, this evaluation strategy lacks a theoretical foundation. So, employing the improved margin criteria as evaluation of the quality of reducts will be discussed in Section 3.

### 3. Principle for evaluating the quality of attribute subsets by margin criteria

Except for developing the efficient attribute reduction algorithms, how to seek an efficient criteria for evaluating the quality of the given attribute subsets is also an important task, since the objective of attribute reduction is to provide those effective attribute subsets for predictors or post-analysis algorithms. So, seeking more effective criteria, for evaluating the quality of given attribute subsets, is our motivation of this study.

As mentioned in Section 1, conventional measure criteria for the quality of the given reducts lacks a theoretical foundation. Especially, it is difficulty to determine which one is more better for the following two cases:  $card(A)$  is equal to  $card(B)$ ;  $card(A) < card(B)$  but  $card(A)$  is near to  $card(B)$ , where both  $A$  and  $B$  are reducts. Therefore, to enhance the discrimination capability of the proposed evaluation strategy, the following margin-based evaluation function is incorporated into the existing evaluation criteria.

As pointed out in [2,19], margins play an important role in modern machine learning research. They measure the classifier confidence when making its decision. Generally, margins are divided into two categories: sample-margin, hypothesis-margin, where sample-margin measures the distance between object(instance) and the decision boundary induced by the classifiers such as *Support Vector Machines*[17], while *hypothesis-margin* is the distance between the hypothesis and the closest hypothesis that assigns alternative decision attribute value to the given object. In this paper, we employ *hypothesis-margin* for improving the evaluation ability of rough set method, since rough set model is mainly a data preprocessing tool. So, for continuous attributes case, we introduce the following definitions of *margins* of an object and all objects.

**Definition 5.** (margin of  $x(A)$  ( $x \in U, A \subseteq C$ )) Given a decision table  $DT = \langle U, C \cup D, V, f \rangle$ , let  $x \in U$  and  $A \subseteq C$ , the margin of  $x$  with respect to  $A$  is

$$M_U(x, A) = \frac{1}{2}(\|x(A) - nearmiss(x(A))\| - \|x(A) - nearhit(x(A))\|) \quad (3)$$

where  $x(A)$  is a project of  $x$  on  $A$ , i.e.,  $x(A) = (f(x, a_1), f(x, a_2), \dots, f(x, a_s))$ ,  $A = \{a_1, a_2, \dots, a_s\}$ ;  $nearhit(x(A))$  and  $nearmiss(x(A))$  denote the

nearest point to  $x(A)$  in  $U$  with the same and different decision attribute value(label for short), i.e.,  $\forall y \in U, y(A) \neq \text{nearhit}(x(A))$  and  $x(A) \neq y(A), f(x(A), D) = f(\text{nearhit}(x(A)), D)$  and  $f(x(A), D) = f(y(A), D), \|x(A) - \text{nearhit}(x(A))\| \leq \|x(A) - y(A)\|$ ;  $\forall y \in U, x(A) \neq y(A), y(A) \neq \text{nearmiss}(x(A)), f(x(A), D) \neq f(\text{nearmiss}(x(A)), D)$  and  $f(x(A), D) \neq f(y(A), D), \|x(A) - \text{nearmiss}(x(A))\| \leq \|x(A) - y(A)\|$ . In this paper, we use the  $L_1$ -norm. Of course, in real applications, other metric strategies could be obtained by metric learning.

**Definition 6.** (margin of  $U(A), A \subseteq C$ ) Given a decision table  $DT = \langle U, C \cup D, V, f \rangle$ , let  $A \subseteq C$ , the margin of  $U$  with respect to  $A$  (for short, the margin of  $A$ ) is

$$E_U(A) = \sum_{x \in U} M_{U-\{x\}}(x, A) \quad (4)$$

Intuitively,  $E_U(A) > 0$ , i.e., in most cases, the distance between a given object and the nearest neighbor with the same label is less than that between the object and the nearest neighbor with the different label. By maximum margin criteria, the more larger the margin of an attribute subset, the more powerful its generalization ability. So, in this paper, we use margin of  $U$  as a compensated evaluation function, i.e., the margin-based evaluation function is embedded into classical evaluation criteria in rough set theory. Concretely, the hybrid evaluation criteria of the quality of the reducts is as follows.

- 1) if  $\text{card}(A) \leq \text{card}(B)$  and  $E_U(A) \geq E_U(B)$ , then the quality of  $A$  is better than that of  $B$ ;
- 2) if  $\text{card}(A) \leq \text{card}(B), \frac{\text{card}(B) - \text{card}(A)}{\text{card}(A)} < \varepsilon$  and  $E_U(B) \gg E_U(A)$ , where  $\varepsilon > 0$  is a pre-set threshold, then the quality of  $B$  is better than that of  $A$ . where both  $A$  and  $B$  are two given reducts.

In this paper, we mainly discuss the decision table in which all condition attributes are continuous attributes. Of course, **Definition 5** also can be easily extended for the case of discrete attribute-values, that is, formula (3) is replaced using the following formula (5).

$$M_U(x, A) = \frac{1}{2} [\text{dissim}(x(A) - \text{simmiss}(x(A))) - \text{dissim}(x(A), \text{simhit}(x(A)))] \quad (5)$$

where  $\text{dissim}(x, y)$  denotes the dissimilarity

between objects  $x$  and  $y$ ;  $\text{simhit}(x(A))$  and  $\text{simmiss}(x(A))$  denote the most similar point to  $x(A)$  in  $U$  with the same and different decision attribute value(label for short). Other notations are similar to definition 5.

Obviously, in most cases, the degree of the dissimilarity between a given discrete object and the most similar object with the same label should be less than that between the object and the most similar object with the different label. However, how to measure the degree of dissimilarity of two discrete objects is a very important but full of challenge. Here we provide a simple strategy to measure the similarity and dissimilarity between any two discrete objects with respect to an attribute subset  $A$  as the following formulas (6) and (7). Developing a more advanced and effective measure method is ongoing, more details will be reported in future work.

$$\text{sim}(x(A), y(A)) = \frac{\text{card}(\{a \in A | f(x, a) = f(y, a)\})}{\text{card}(A)} \quad (6)$$

$$\text{dissim}(x(A), y(A)) = 1 - \text{sim}(x(A), y(A)) \quad (7)$$

Based on formulas (6),(7),(5)and(4),we can also obtain the margin of  $U$ (all objects) for a discrete decision table.

Following by above analysis, we will introduce a rough set-based attribute reduction algorithm with margin strategy in Section 4.

## 4. Incorporating margin into rough set-based model

In this section, to illustrate how to incorporate the margin criteria in the existing rough set-based attribute reduction methods, here employing the classical discernibility matrix-based models as representative, the following **Algorithm 1** is the novel discernibility function-based attribute reduction algorithm with margin criteria. Similarly, the margin criteria could be easily embedded in other rough set-based attribute reduction models such as positive region-based and evolutionary models, etc.

**Algorithm 1.** *IMAR*//Incorporating margin criteria in an attribute reduction algorithm

**Input:** let  $DT = \langle U, C \cup D, V, f \rangle$  be a decision table,  $\varepsilon(\varepsilon > 0)$  is a pre-set threshold;

**Output:** Ranking of some selected reducts.

**Step 1.** Min-max normalization for each continuous-valued attribute, generate a new decision table  $DT1 = \langle U1, C \cup D, V, f \rangle$ ;

**Step 2.** discretize the data set  $U$ , get a new decision  $DT2 = \langle U2, C \cup D, V, f \rangle$ ;

**Step 3.**

for( $i=1$  to  $card(U2)-1$ ,  $j=i+1$  to  $card(U2)$ )

If  $f(x_i, d) \neq f(x_j, d)$  then  $m_{i,j} = \{a \in C : f(x_i, a) \neq f(x_j, a)\}$ ;

else  $m_{i,j} = \{\}$ ;

**Step 4.** construct a discernibility function  $f_{DM}$  by (2);

**Step 5.** obtain the reducts by solving the discernibility function  $f_{DM}$ ;

**Step 6.** filter the reducts;

**Step 7.** evaluate the quality of every reduct in the reducts using margin strategy for  $DT1$ ;

**Step 8.** return reducts;

In **Algorithm 1**, all continuous-valued attributes are normalized first. Here, the so-called Min-max normalization method is employed, which maps a value  $\nu$  of  $a$  to  $\nu'$  in the ranges  $[new\_min_a, new\_max_a]$  by computing

$$\nu' = \frac{\nu - \min_a}{\max_a - \min_a} (new\_max_a - new\_min_a) + new\_min_a \quad (8)$$

here, let  $new\_min_a$  and  $new\_max_a$  be 0 and 1, respectively.

Then, all continuous-valued attributes are discretized into intervals of discrete categorical values [5,8]. Next, based on formulas (1) and (2), a discernibility matrix and the corresponding discernibility function, induced by it, are generated respectively. Further, by solving the discernibility function  $f_{DM}$ , all reducts are obtained.

For the case having lots of reducts, by Step 6, most of those reducts with larger cardinality, relative to the attribute reductions with small cardinality, are filtered. Following this strategy, only those short and effective reducts are reserved. Finally, using the maximum margin criteria, these reducts are ranked, hence those more effective and relatively independent reducts could be obtained.

Especially, by setting a reasonable value for the adjustable parameter  $\varepsilon$ , some relatively long but having larger margin reducts will acquire higher

score. Hence the new criteria can effectively overcome the disadvantages of the classical evaluation method, which is also a compensation but not instead of it. Moreover, according to the number of reducts and the cardinality of reducts,  $\varepsilon$  is easily determined via user requirement.

In contrast with the existing algorithm, this algorithm incorporates margin criteria in the discernibility function-based attribute reduction model, hence which can ensure that those more effective reducts can be chosen.

## 5. Experimentanets

To test the effectiveness of algorithm developed in this paper, we performed the experiments on three publicly available datasets from *UCI* database (These datasets can be downloaded at <http://www.ics.uci.edu>).

A brief description of the *UCI* datasets is given at first: (1) Glass Identification(Glass): 214 objects, 9-conditional attributes(9F), 1-decision attribute (1D), 2-classes(2C) ( the 1th attribute is removed in original dataset, class 1: float data; class 2: non float data); For short, denoted as (214,9F,1D,2C); (2) Iris23: (100,4F,1D,2C) (iris23 consists of those objects contained in second and third classes); (3) Wisconsin Prognostic Breast Cancer(WPBC):(194,32F,1D,2C) ( 4 objects with missing values are deleted and ID attribute is removed from original WPBC). In our experiments, we randomly choose 50% objects as training data, and use remaining 50% objects as testing data.

datasets	reduct	margin	K-NN		KFDA	
			Accuracy	K=3	Accuracy	K=5
Glass	<b>{a<sub>3</sub>, a<sub>4</sub>, a<sub>6</sub>}</b>	<b>8.914239</b>	48%	<b>55%</b>	<b>57%</b>	<b>88.89%</b>
	{a <sub>3</sub> , a <sub>5</sub> , a <sub>7</sub> }	8.575893	<b>52%</b>	52%	56%	87.04%
	{a <sub>2</sub> , a <sub>3</sub> , a <sub>4</sub> , a <sub>7</sub> }	9.405252	36%	47%	48%	<b>87.04%</b>
	<b>{a<sub>3</sub>, a<sub>4</sub>, a<sub>6</sub>, a<sub>7</sub>}</b>	<b>9.831238</b>	<b>47%</b>	<b>48%</b>	<b>48%</b>	86.11%
	<b>{a<sub>1</sub>, a<sub>2</sub>, a<sub>4</sub>, a<sub>5</sub>, a<sub>6</sub>}</b>	<b>8.866564</b>	<b>24%</b>	<b>39%</b>	<b>43%</b>	<b>88.89%</b>
Iris23	{a <sub>1</sub> , a <sub>4</sub> }	8.760097	20%	20%	26%	84.26%
	{a <sub>1</sub> , a <sub>4</sub> }	3.885714	56%	36%	32%	100.0%
	<b>{a<sub>3</sub>, a<sub>4</sub>}</b>	<b>4.627778</b>	<b>72%</b>	<b>58%</b>	<b>40%</b>	<b>100.0%</b>
	<b>{a<sub>4</sub>, a<sub>27</sub>}</b>	<b>4.824138</b>	<b>18%</b>	10%	05%	<b>77.32%</b>
	{a <sub>1</sub> , a <sub>10</sub> }	4.298874	12%	06%	<b>08%</b>	76.29%
WPBC	{a <sub>2</sub> , a <sub>14</sub> }	3.585075	12%	<b>11%</b>	08%	77.32%
	<b>{a<sub>5</sub>, a<sub>16</sub>, a<sub>19</sub>}</b>	<b>5.191718</b>	08%	10%	<b>17%</b>	<b>77.32%</b>
	{a <sub>1</sub> , a <sub>15</sub> , a <sub>29</sub> }	4.200866	20%	08%	06%	76.29%
	{a <sub>14</sub> , a <sub>15</sub> , a <sub>29</sub> }	4.233819	<b>27%</b>	<b>12%</b>	15%	76.29%

Table 1. Experimental results.

Notes: The cells with boldface represent those attribute subsets with larger margin or with high accuracy, and corresponding margins and accuracies, respectively.

For simplicity of discussion, in our experiments, the conditional attributes are named as  $a_1, a_2, \dots, a_m$  in order, and let  $\varepsilon=0.1$ . Why the con-

trolled parameter  $\varepsilon$  is set a very small value? our main intention of this study attempts to evaluate those reducts with same cardinality. In addition, we use *RBF* kernel as the kernel function of the *KFDA* algorithm(*KFDA* for short).

By Steps1-5 of **Algorithm 1**, the number of the reducts obtained by Glass, Iris23 and WPBC are 6,2 and 87, respectively. All reducts generated by the former two datasets are used for evaluating their quality, while 6 reducts are chosen from 87 reducts obtained by the latter dataset, where 3 reducts are selected from 38 reducts with the cardinality 2, and another 3 reducts are chosen from the remains with the cardinality 3.

The experimental results are shown in Table 1. From Table 1, for those reducts with same cardinality, we can find that the more larger the margin, the more better the generalization ability in most cases, it is consistent with "margin theory", hence the effectiveness of the new evaluation function is further verified. Since the three datasets of this paper are non linearly separable, it is nature that the performance and stability of K-NN classifier are inferior to those of *KFDA* classifier. Further, we find that the *KFDA* classifier has more better generalization ability for those reducts with small cardinality and large margin for all three datasets, while the K-NN also has more better generalization ability for those reducts with small cardinality and large margin for the former two datasets, this is also consistent with the new criteria of this paper. For the dataset WPBC, the performance of K-NN classifier has a litter fluctuation when the value of K is increased gradually, this may be related to its non linear separability, but it has not influence for obtaining those more effective reducts.

In summary, here we provided results on how to measure the margin of a dataset with respect to an attribute subset, and how to incorporate margin criteria in the classical rough set-based algorithms for evaluating the effectiveness of the obtained reducts. More studies which including more tests and some novel improving strategies are under exploring currently.

## 6. Conclusions

To more effectively evaluate the quality of reducts obtained by rough set model, in this paper, we incorporate margin criteria in rough set-based attribute reduction model. For simplicity, using discernibility matrix-based attribute reduction as a representative, a new attribute reduction algorithm with margin criteria is introduced. By the support

of so-called "margin theory", the newly hybrid evaluation function(combining margin criteria and the classical evaluation function in rough set theory) provides a theoretical foundation for measuring the quality of reducts. The preliminary experimental results show that the attribute reductions with larger margin have better or comparable performance than those with relatively small margin for all reducts with same cardinality, this is consistent with "margin theory". Therefore, our newly developed method can, in most cases, get more effective attribute subsets.

However, in this paper, we mainly discuss the case of continuous attributes. Moreover, the proposed margin model need to be further improved by employing different metric criteria.

So, in the future, further studies on improving strategies of the proposed margin model for establishing an effective margin-based evaluation function are required. In addition, building more effective margin-based evaluation function, that aims at the discrete attributes, is also one of our future work.

## Acknowledgments

This paper is partially supported by National Natural Science Foundation of P.R.China and Jiangsu(Grant Nos.70371015 and BK2005135),and National Science Research Foundation of Jiangsu Province(Grant No.05KJB520066),respectively.

## References

- [1] R.Kohavi and G.John. Wrapper for feature subset selection. *Artificial Intelligence*,97:273-324,1997.
- [2] Gilad-B.Ran, N.Amir, and T.Naftali. Margin Based Feature Selection-Theory and Algorithms. *In Proc. of the 21st Int. Conf. on Machine Learning*, Banff, Canada, pp.337-334,2004.
- [3] R.Jensen,Q.Shen.Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough-Based Approaches, *IEEE Transactions on Knowledge and Data Engineering*,16(12):1457-1471,2004.
- [4] Z.Pawlak. Rough sets. *International Journal of Information and Computer Science*, 11(5):341-356,1982.
- [5] Liu Qing. *Rough sets and rough reasoning*. Beijing: Science Press,2001(in Chinese)

- [6] R. W. Swiniarski, A. Skowron. Rough set methods in feature selection and recognition. *Pattern Recognition Letters*, 24:833-849,2003.
- [7] S. K. Pal, V. Talwar, and P. Mitra. Web Mining in Soft Computing Framework: Relevance, State of the Art and Future Directions. *IEEE Transactions on Neural Networks*, 13(5):1163-1177,2002.
- [8] G.Y.Wang. *Rough Set Theory and Data Mining*. Xi'an Jiaotong University Press, Xi'an, 2001.
- [9] Ju-Sheng Mi, Wei-Zhi Wu, Wen-Xiu Zhang. Approaches to knowledge reduction based on variable precision rough set model. *Information Science*, 159:255-272,2004.
- [10] R. Slowinski, D. Vanderpooten. A generalized definition of rough approximations based on similarity. *IEEE Transaction on Knowledge and Data Engineering*, 12(2):331-336,2000.
- [11] William Zhu. Topological approaches to covering rough sets. *Information Science*, 177:1499-1508,2007.
- [12] Wang Jue, WANG Ju. Reduction algorithm based on discernibility matrix the ordered attributes method. *Journal of Computer Science and Technology*, 16(6):489-504,2001.
- [13] M. Yang. An Incremental Updating Algorithm of the Computation of a Core Based on the improved Discernibility Matrix. *Chinese Journal of Computers*, 29(3):407-413,2006.
- [14] M. Yang, P. Yang and Gen-lin Ji. Updating of attribute reduction for case of deleting. *2007 International Conference on Fuzzy Information and Engineering, Advances in Soft Computing(ASC) 40*, Springer-Verlag Berlin Heidelberg, pp.884-893,2007.
- [15] F. Hu, G.Y. Wang, H. Huang, Y. Wu. Incremental attribute reduction based on elementary sets. In: *Proceedings of 10th international conference on Rough sets, Fuzzy sets, Data mining, and Granular computing, Lecture Notes in computer science*, Regina, Canada, pp.185-193,2005.
- [16] M. Banerjee, S. Mitra and A. An. Feature Selection Using Rough Sets. *Studies in Computational Intelligence(SCI)*, Springer-Verlag Berlin Heidelberg, 16:3-20,2006.
- [17] V.P. Vapnik. *The Nature of Statistical Theory*. Springer-Verlag, Berlin, 1995.
- [18] Yi-Jun Sun, S. Todorovic and Jian Li. Unifying multi-class AdaBoost algorithms with binary base learners under the margin framework. *Pattern Recognition Letters*, 28:631-643,2007.
- [19] H. John Caulfield and Kaveh Heidary. Exploring margin setting for good generalization in multiple class discrimination. *Pattern Recognition*, 38:1225-1238,2005.
- [20] A.K. Jain, R.P.W. Duin, and J. Mao. Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4-37,2000.
- [21] C. Domeniconi, J. Peng, D. Gunopulos. Locally adaptive metric nearest-neighbor classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1281-1285,2002.
- [22] S. Mika, G. Ratsch, J. Weston, B. Scholkopf and K.-R. Muller, "Fisher discriminant analysis with kernels", in *Neural Networks for Signal Processing IX*, Y.-H. Hu, J. Larsen, E. Wilson, and S. Douglas, Eds., IEEE Press, New York, pp.41-48,1999.