A novel approach of rough set-based attribute reduction using fuzzy discernibility matrix

Ming Yang [1,2] , Songcan Chen[1] and Xubing Yang[1]
[1] Department of Computer Science and Engineering
Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[2] Department of Computer Science, Nanjing Normal University
122 Ninghai Road,Nanjing 210097,China
m.yang@njnu.edu.cn

Abstract

Rough set approach is one of effective attribute reduction (also called a feature selection) methods that can preserve the meaning of the attributes(features). However, most of existing algorithms mainly aim at information systems or decision tables with discrete values. Therefore, in this paper, we introduce a novel rough set-based method followed by establishing a fuzzy discernibility matrix by using distance preserving strategy for attribute reduction, and only choose fisher discriminant analysis with kernels as discriminant criteria for testing the effectiveness of selected attribute subsets with relatively higher fitness values, since the proposed method is independent of post-analysis algorithms (predictors).Experimental results show that the classifiers developed using the selected attribute subsets have better or comparable performance on all eight UCI benchmark datasets than those obtained by all attributes. Thus, our newly developed method can, in most cases, get effective attribute subsets. In addition, this method can be directly incorporated into other learning algorithms, such as PCA, SVM and etc. and can also be more easily applied to many real applications, such as Web Categorization, Image recognition and etc.
Keywords: *Rough Set, Fuzzy Discernibility Matrix, Attribute Reduction*

1. Introduction

Many approaches have been proposed for dimensionality reduction in the literature [1][2][3][4]. Generally, the existing dimensionality reduction methods can be roughly divided into the following two categories: attribute extraction(feature extraction) and attribute reduction(feature selection). Feature extraction means the original features in the measurement space are transformed into a new dimension-reduced space via some specified transformation[2][3]. Although the significant variables determined in the new space are related to the original variables, the interpretation may be lost.

Unlike feature extraction, attribute reduction(also called feature selection) has been viewed as one of the most key problems in the field of the machine learning[1][4]. 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 overall 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[4][5]. 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 an efficient data preprocessing tool, rough set approach is one of effective attribute reduction methods that

Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)
0-7695-2874-0/07 $25.00 © 2007 IEEE

can preserve the meaning of the attributes[6][7], it has been widely applied in many fields of machine learning, such as data mining, intelligent data analyzing and control algorithm acquiring [8][9][10], etc. The rough set theory-based attribute reduction methods can roughly be divided into the following categories: discernibility matrix-based methods[1][11][12][13], information entropy-based methods[14],positive region-based methods[15] and other evolutionary methods [16]. However, these models are only fit for information systems or decision tables with discrete values, that is, before using these algorithms, discretizing continuous-valued attributes(for short,also called continuous attributes) is considered firstly[17]. However, the following disadvantages still remain in the state-of-the-art discretization methods: 1) loss of some useful information;2) it is very difficult to determine the partitioning granularity for the continuous attributes, hence both the number of intervals and the positions of cuts may be adjusted repeatedly for obtaining useful and interesting attribute subsets. In addition, due to discretization operator, it is also inconvenient to incorporate rough reduction approaches into other machine learning algorithms, such as KFDA[18], PCA[19] and etc. Therefore, seeking a novel rough attribute reduction method that directly aims at continuous attributes is also our main motivation of this study.

In this study, a new notation, called fuzzy discernibility matrix, is introduced first, which is same as the classical discernibility matrix in form, but their difference is: its every entry is directly generated by distance preserving strategy between two objects with different class labels (also called decision attribute values), that is, every cell in a fuzzy discernibility matrix records those significant attributes that can effectively discern the corresponding two objects. Then, we present a novel rough attribute reduction algorithm based on fuzzy discernibility matrix by using genetic operators. Since the proposed method is independent of post-analysis algorithms (predictors), that is, how to select effective attribute subsets don't depend on some special classification algorithms. So, in this paper we only choose the well known fisher discriminant analysis with kernels as evaluating criteria for testing the classification accuracy of each selected attribute subset with higher fitness value, the corresponding experimental datasets are eight standard UCI datasets.

In addition, since this method need not discrete the continuous attributes, it can be directly efficiently incorporated into other learning algorithms, such as PCA, SVM and etc. and can also be more easily applied to many real applications, such as Web Categorization, Image recognition and etc. Hence, this studies are of real value.

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, a new concept that is called fuzzy discernibility matrix is introduced; A new rough attribute reduction algorithm based on fuzzy discernibility matrix 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 [1][8].

In rough set theory, a data set can be formally described using a decision table [8]. A decision table is denoted as $DT =< U, Q, V, f >$, where $U = \{x_1, x_2, ..., x_n\}$ is a non-empty finite set of objects or cases, called universe, Q is a non-empty finite set of attributes or attributes, $Q = C \bigcup D$, where C is the set of condition attributes and D is the set of decision attributes, $C \bigcap D = \{\}$. 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: $U \times Q \to V$ is a total function such that for every $a \in Q, x_i \in U, i = 1, 2, ..., card(U)$.Throughout this paper, $\{\}$ denotes the empty set, and $card(X)$ denotes the function that returns the cardinality of the argument set X.

Given a decision table DT, a discernibility matrix DM [7] is defined as an n× n matrix of DT with the (i,j) th entry $m_{i,j}$ given by:

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

Given a decision table DT, let its corresponding discernibility matrix be DM=$\{m_{i,j}\ \}$ by formula (1), where x_i and x_j are two objects of U. The following equivalent definition of attribute reduction was introduced in [11].

An attribute subset R of C is an attribute reduction iff $R \bigcap m_{i,j} \neq \{\}$ 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. $S \bigcap m_{i,j} = \{\}$.

Based on discernibiliy matrix, many efficient algorithms were developed for attribute reduction[11][12]. Unfortunately, this method is only fits for a decision table with discrete values. So, how to construct directly a discernibility matrix, that aims at a decision table with continuous attributes, will be discussed in Section 3.

3. Construction of a fuzzy discernibility matrix

In short, in this paper,we only focus on the decision table in which only decision attributes are discrete at-

Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)
0-7695-2874-0/07 $25.00 © 2007 IEEE

tributes,while all conditional attributes are continuous.

Let DT be a decision table, as described in Section 2, where $C = \{a_1, a_2, ..., a_m\}$ ($V_{a_j} \subset R$,$1 \leq j \leq m$), $D = \{d\}$($V_d = \{c_1, c_2, ...c_s\}$,namely it is a set of s-class), $U = \{x_1, x_2, ..., x_n\}$, $x_i = (x_{i1}, x_{i2}, ..., x_{im}, y_i)$, $x_{ij} \in V_{a_j}$, $y_i \in V_d$, $1 \leq i \leq n, 1 \leq j \leq m$. According to the classical discernility matrix definition, continuous attributes need to be discretized first before using formula (1).

As analyzed in Section 1, seeking a novel strategy, that can directly generate a discernibiliy matrix by using the original data, is necessary and also the main motivation of this study. In essence, each cell $m_{i,j}$ in formula (1) records those significant attributes that discern objects x_i and $x_j (1 \leq i, j \leq n)$ with different class labels. In order to effectively discern two objects with different class labels, in this paper we introduce a novel strategy called so-called "distance preserving", the details are as follows.

Firstly, all continuous attributes are normalized for helping prevent attributes with initially large ranges from outweighing attributes with initially smaller ranges. In this paper, the so-called Min-max normalization method is employed , which maps a value v of a to $v\,'$ in the ranges $[new\_min_a, new\_max_a]$ by computing

$$v\,' = \frac{v - min_a}{max_a - min_a}(new\_max_a - new\_min_a) + new\_min_a \quad (2)$$

here, let new_min_a and new_max_a be 0 and 1, respectively.

Secondly, the distance between each pair of objects, x_i and x_j,on a given attribute subset B is computed, the well-known Manhattan distance is used in this paper, which is defined as

$$d_B(x_i, x_j) = \sum_{a_k \in B} |x_{ik} - x_{jk}| \quad (3)$$

by formula (3), the dissimilarity between the objects on the given attribute subset is computed, i.e., the larger the distance of two objects the more the dissimilarity. Along this, if $d_C(x_i, x_j) < \delta$ where δ is a pre-set threshold, we call objects x_i and x_j are indispensable or inconsistent; dispensable or consistent otherwise. Also, for two consistent objects x_i and x_j, namely $d_C(x_i, x_j) \geq \delta$ in order to get the attributes in C that are relatively more significant than others, we need to seek those attributes that have more contribution to $d_C(x_i, x_j)$, i.e., the larger the value of $|f(x_i, a_k) - f(x_j, a_k)|$ or $|x_{ik} - x_{jk}|$,the more the significance of the attribute a_k, where $a_k \in C$. To effectively control the cardinality of relatively significant attribute subset, here we provide a concise but effective strategy as follows.

$$\min card(B) \quad (4)$$

$$\text{s.t } \frac{d_B(x_i, x_j)}{d_C(x_i, x_j)} \geq \varepsilon, B \subseteq C$$

where $1 \leq i, j \leq n$, ε is a pre-set threshold, which is also an adjustable parameter. Solving this optimization problem, those significant attributes that discern each pair of objects can be obtained. For simplicity, for every pair of objects x_i and x_j, the solution of eqn.(4) is denoted by $g_\varepsilon(x_i, x_j)$ under a controlled parameter ε.The solution to eqn.(4) can be obtained by the following concise and efficient method.

For given pre-set threshold ε, suppose that $|f(x_i, a_k) - f(x_j, a_k)| \geq |f(x_i, a_{k+1}) - f(x_j, a_{k+1})|$ holds,$1 \leq k \leq m-1$, the solution of this optimization problem, $B (B = \{a_1, a_2, ..., a_t\})$ can be obtained by using the following equivalent equations.

$$\frac{\sum_{k=1}^{t} |x_{ik} - x_{jk}|}{\sum_{k=1}^{m} |x_{ik} - x_{jk}|} \geq \varepsilon \wedge \frac{\sum_{k=1}^{t-1} |x_{ik} - x_{jk}|}{\sum_{k=1}^{m} |x_{ik} - x_{jk}|} < \varepsilon \quad (5)$$

Generally, the conditional attributes are sorted in descending order of the distances of $|f(x_i, a_k) - f(x_j, a_k)| (k = 1, 2, ..., m)$(larger distances mean more significance), hence a solution of eqn.(4), that containing t best significant attributes, is certainly the attribute subset corresponding to the t biggest distances in the set, $\{|f(x_i, a_k) - f(x_j, a_k)|\}_{k=1}^{m}$.

Following by above analysis, a new discernibility matrix called $fuzzy\ discernibility\ matrix$ is introduced, denoted by FDM for simplicity. FDM is also an $n \times n$ matrix of DT with the (i,j)th entry $c_{i,j}$ given by:

$$c_{i,j} = \begin{cases} g_\varepsilon(x_i, x_j) & \text{if } d_C(x_i, x_j) \geq \delta \text{ and} \\ & f(x_i, d) \neq f(x_j, d), 1 \leq i, j \leq n \\ \{\} & \text{otherwise} \end{cases} \quad (6)$$

where both ε and δ are two pre-set thresholds. The proposed algorithm for generating FDM is summarized as follows:

Algorithm 1.$Gen\_FDM(C, D, U, \varepsilon, \delta)$
/*construction of a fuzzy discernibility matrix*/
Input: C: conditional attributes, D: decision attribute, U:objects, ε, δ:pre-set thresholds;
Output: FDM.

Step 1. Min-max normalization for each continuous attribute;

Step 2. for (i=1 to $card(U) - 1$, $j = i + 1$ to $card(U)$)

If $f(x_i, d) \neq f(x_j, d)$ and $d_C(x_i, x_j) \geq \delta$ then $c_{i,j} \leftarrow g_\varepsilon(x_i, x_j)$;

else $c_{i,j} \leftarrow \{\}$;

Step 3. Return FDM.

By adjusting δ, Gen_FDM algorithm can effectively control the fuzzy boundary of two different classification objects for improving the performance of post-analysis algorithms. Further, using the adjustable parameter ε, some more significant attributes can be discriminated from other attributes. Intuitively, these benefit to improve the generalization of a classifier, since relevant and significant attribute subset can be obtained. Using the generated fuzzy discernibility matrix, how to obtain the useful and interesting relevant attribute subsets will can be presented in Section 4.

4. Rough attribute reduction using fuzzy discernibility matrix

There are a lot of discernibility matrix based methods available for attribute reduction[1][11][16], generally including discernibility function based methods[1][8], heuristic strategy based methods and genetic operators based methods[16]. In this paper, we will present only genetic operators based methods with heuristic strategy.

Unlike classical optimization methods, Genetic algorithms(for short, GAs) deal with a population of solutions/individuals by using selection, crossover and mutation operators. A population of solutions is repeatedly evolved over generations by optimizing a fitness function, which provides a quantitative measure of the fitness of individuals in the pool. Selection operator chooses better individuals to participate into crossover, i.e. those individuals with high fitness values. Crossover operator is responsible for creating new individuals from the old ones. Mutation also generate new individuals, but only in the vicinity of old individuals.

Generally, reducts(solutions) are represented by binary strings of length m, where m is the number of conditional attributes. In the bit representation '1' means that attribute is present and '0' means that it is not. Most of discernibility matrix based GAs for attribute reduction use the following fitness function or its variants.

$$F(v) = \frac{m-L_v}{m} + \frac{card(C_v)}{(n^2-n)*0.5} \quad (7)$$

where v is an individual,i.e., a reduct candidate, m is the number of conditional attributes, L_v is the number of 1's in v,C_v is the set of elements that v can discern in the discernibility matrix, n is the number of objects. However, the first part and second part are of a conflicting nature. Completely using the value of this fitness function as the fitness of individuals is obviously improper, because the individuals with relatively low fitness values may be more better than those with relatively high fitness values.

Hence, in order to obtain potential and more useful reducts, the following heuristic-based methods are incorporated in this standard genetic algorithms: 1) Computing the frequency of each attribute that appears in the fuzzy discernibility matrix, and the so-called 'Roulette Wheel' is used for initiating a population of individuals for enhancing their fitness. 2) Eqn.(7) is improved and eqn.(8) is obtained, where N is the number of non-empty entries in the discernibility matrix. According to eqn.(10), we can efficiently evaluate the fitness of v using the value of eqn.(10). Along this, if the classification capability of over 80% individuals are near to 1, perhaps corresponding genetic algorithm can be stopped. Thus, the value of eqn.(10) can be used as one of stop criterion of genetic algorithms; 3) The individuals, that have less attributes(few 1's) and their values corresponding to eqn.(10) are near to 1, will be used for post-analysis algorithm. This approach can overcome the disadvantages of only using a single fitness function(eqn.(8)).

$$F(v) = \frac{m-L_v}{m} + \frac{card(C_v)}{N} \quad (8)$$

$$F_1(v) = \frac{m-L_v}{m} \quad (9)$$

$$F_2(v) = \frac{C_v}{N} \quad (10)$$

As analyzed above, a novel fuzzy discernibility matrix based algorithm for attribute reduction is summarized as follows.

Algorithm 2. $RARABFDM(C, D, U, \varepsilon, \delta)$
/*Rough Attribute Reduction Algorithm Based on Fuzzy Fiscernibility Matrix*/
Input: C: conditional attributes, D: decision attribute, U:objects, ε, δ:pre-set thresholds;
Output: Solutions or reducts.

Step 1. parameter initialization :
- $\theta \leftarrow A$ pre-defined threshold; /*$\theta \in [0.8, 1]$,it usually is near to 1*/
- $P_c \leftarrow$ Crossover probability;
- $P_m \leftarrow$ Mutation probability;
- $T \leftarrow$ maximum number of iterations;
- $K \leftarrow 0$;

Step 2. $FDM \leftarrow Gen\_FDM(C, D, U, \varepsilon, \delta)$;

Step 3. Computing the frequency of each attribute that appears in FDM;

Step 4. $P \leftarrow A$ random population of size m is generated by using the principle of 'Roulette Wheel' according to the frequency of each attribute;

Step 5. The three fitness values for each individuals is calculated; $t \leftarrow All$ individuals that their fitness values $F_2 \geq \theta$ (eqn.(10));

Step 6. while($k < T$ and $\frac{card(t)}{card(p)} < 0.9$) do

- **Step 6.1.** Select individuals using 'Roulette Wheel' strategy;
- **Step 6.2.** Crossover with P_c;
- **Step 6.3.** Mutation with P_m;
- **Step 6.4.** Some uninteresting individuals are replaced and offspring individuals P are created;
- **Step 6.5.** The three fitness values for each individuals is repeatedly calculated;
- **Step 6.6.** $K \leftarrow K + 1$;

Step 7. Choose some suboptimal individuals.

In contrast with the existing algorithm, this algorithm directly creates a fuzzy discernibility matrix for a decision table with continuous attributes, hence which can overcome the disadvantages induced by discretization methods. In addition, some heuristic strategies are embedded in the standard genetic algorithms, which can effectively assure that the interesting and useful reducts could be obtained in most cases.

Table 1.Experimental results

datasets	RARABFDM		KFDA
	Reduct(individual):$v(\varepsilon)$	$F_2(v)$	accuracy(%)
CMC	-	1	92.54
	$\{2, 3, 6, 8\}$(0.86)	0.9984	88.47
	$\{1, 6, 7, 8\}$(0.86)	0.9976	88.81
	{1,2,6,7,8}(0.91)	0.9999	92.20
Diabete	-	1	71.43
	{1,2,5,8}(0.86)	0.9958	69.48
	{1,2,3,7,8}(0.86)	0.9999	79.22
	{1,2,3,5,6,8}(0.91)	1	71.43
Glass	-	1	95.46
	{1,2,4,8}(0.86)	0.9865	100
	{3,4,5,7,9}(0.86)	1	95.46
	{1,2,3,4,5,7}(0.91)	1	95.46
iris23	-	1	100
	{3,4 }(0.86)	0.9939	100
	{1,4}(0.86)	0.9966	100
	{1,2,3}(0.91)	1	100
BLD	-	1	63.77
	{1,4,6}(0.86)	0.9932	60.87
	{1,2,3,6}(0.86)	0.9994	59.4
	{1,3,4,6}(0.91)	0.9997	66.67
Ion	-	1	88.73
	{1,11,13,20,25,33,34}(0.86)	0.9996	95.78
	{1,11,13,17,21,24,25,30,34}(0.86)	0.9999	95.78
	{3,6,7,10,13,16,17,18,22,23,25,27,30}(0.86)	1	95.78
Sonar	-	1	69.77
	{9,25,33,58,59}(0.86)	0.9907	69.77
	{15,25,33,48,58,59}(0.86)	0.9979	83.72
	{7,9,11,13,26,30,37,53,60}(0.91)	1	83.72
WPBC	-	1	75
	{9,20,23,28,29}(0.86)	0.9965	75
	{1,8,9,18,26,28}(0.86)	0.9995	75
	{7,12,16,18, 22,24,29,32}(0.91)	1	75

Notes: "-" represents all conditional attributes.

5. Experiments

To test the efficiency of algorithms developed in this paper, we performed the experiments on eight 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) Contraceptive Method Coice(CMC): 1473 objects, 9-conditional attributes(9F)(after the 9th attribute is deleted), 1-decision attribute(1D), 2-classes(2C)(class 1 includes original classes 1,2 and 3; class 2 consists of others). For short, denoted as (1473,9F,1D,2C); (2) Pima Indians Diabetes(Diabete): (768, 8F,1D,2C); (3) Glass Identification(Glass): (214, 9F, 1D,2C)(the 1th attribute is removed in original dataset, class 1: float data; class 2: non float data); (4) iris23: (100,4F,1D,2C) (iris23 consists of those objects contained in second and third classes); (5) BUPA Liver Disorders(BLD): (345, 6F,1D,2C); (6) Ionosphere (Ion): (351, 34F, 1D,2C); (7) Sonar: (208,60F,1D,2C); (8) 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 80% objects as training data, and use remaining 20% objects as testing data.

For simplicity of discussion, in our experiments, the conditional attributes are named as $a_1, a_2, ..., a_m$ in order(for short, here we use the index i to denote the attribute a_i),and the related parameters as follows: let ε=0.86 or 0.91, δ=0.04*m, P_c=0.6, P_m=0.1, T=200, and pop_size=30 and θ =0.98. In our experiments, why the controlled parameter ε is set a moderate value? our intention attempts to find those relatively more significant attributes but not lost useful information.

Because the proposed method is independent of post-analysis algorithms (predictors), here we only choose the well known fisher discriminant analysis with kernels as evaluating criteria for testing the classification accuracy of each selected attribute subset. In the KFDA algorithm, we use RBF kernel as the kernel function.

The experimental results are shown in Table 1. From Table 1, we can find that some interesting and suboptimal individuals could be obtained, and corresponding classifiers have better or comparable performance on all datasets than those based on originated attributes. Further, we also find that most interesting and potential reducts can be obtained under ε=0.86, this also indicates the controlled parameter ε has weak dependency on given datasets. Further, we also find the interesting solutions usually come from those individuals v, their $F_2(v)$ is near to 1 but not equal to 1.

In summary, here we provided results on how to directly create a fuzzy discernibiliy matrix from a decision table with continuous attributes, and how to establish a novel ge-

0-7695-2874-0/07 $25.00 © 2007 IEEE

netic algorithm based on it for attribute reduction. More studies which including more tests and some novel improving strategies are under exploring currently.

6. Conclusions

In this paper, a new concept called fuzzy discernibiliy matrix is introduced first, then corresponding constructing algorithm for fuzzy discernibility matrix is developed by using the so-called ”distance preserving” strategy. Along this, we presented a novel fuzzy discernibility matrix based genetic algorithm followed by establishing new fitness functions and corresponding heuristic techniques. The experimental results showed that the algorithms developed in this paper can effectively and efficiently obtain the relevant attribute subsets. In the future, further studies on improving strategies of the proposed approach for creating fuzzy discernibility matrix are required.

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

References

[1] R.Jenshen, Q.Shen. Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough-Based Approaches. *IEEE Transactions on Knowledge and Data Engineering*,2004,16(12):1457-1471.

[2] I.T.Jolliffe. *Principal Component Analysis*. Second ed. Wiley,2002.

[3] A.K.Jain, R.P.W.Duin, and J.Mao. Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,2000,22(1):4-37.

[4] R.Kohavi and G.John. Wrapper for feature subset selection. *Artificial Intelligence*,1997.273-324.

[5] 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, 2004.

[6] Z.Pawlak. Rough sets. *International Journal of Information and Computer Science*,1982,11(5): 341-356.

[7] G.Y.Wang. *Rough Set Theory and Data Mining*. Xi’an Jiaotong University Press,Xi’an, 2001.

[8] Q.Liu. *Rough sets and rough reasoning*. Beijing: Science Press,2001.

[9] R. W. Swiniarski and A.Skowron. Rough set methods in feature selection and recognition. *Pattern Recognition Letters*,2003 (24) :833-849.

[10] 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*,2002,13(5): 1163-1177.

[11] A.Skowron and C.Rauszer. The Discernibility Matrices and Functions in Information Systems. *Intelligent Decision Support*,1992.331-362.

[12] Jue.Wang and Ju.WANG. Reduction algorithm based on discernibility matrix the ordered attributes method. *Journal of Computer Science and Technology*,2001,16(6):489-504.

[13] M.Yang. An Incremental Updating Algorithm of the Computation of a Core Based on the improved Discernibility Matrix. *Chinese Journal of Computers*,2006,29(3):407-413.

[14] G.Y.Wang, J. Zhao, J.J.An, et al. Theoretical study on attribute reduction of rough set theory: comparison of algebra and information views. *In: Proceedings of the Third IEEE International Conference on Cognitive Informatics*,2004.

[15] J.Zh.Dong, N.Zhong, and S.Ohsuga. Using rough sets with heuristics for feature selection. *N.Zhong (Eds.):RSFDGrC’99*,LNAI 1711,1999.178-187

[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] D.Janssens,T.Brijs, K.Vanhoof and G.Wets. Evaluating the performance of cost-based discretization versus entropy- and error-based discretization. *Computers and Operations Research*,2006(33):3107-3123.

[18] 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., pp. 41-48. IEEE Press, New York, 1999.

[19] K.R.Tan and S.C.Chen. Adaptively weighted sub-pattern PCA for face recognition. *Neurocomputing*,2005(64): 505-511.