An Approach for Fuzzy-Rough Sets Attribute Reduction via Mutual Information

Feifei Xu
Department of Computer Science and Technology, Tongji University, Shanghai 201804
xufeifei1983@hotmail.com

Duoqian Miao
Department of Compute Science and Technology, Tongji University, Shanghai 201804
miaoduoqian@163.com

Lai Wei
Department of Computer Science and Technology, Tongji University, Shanghai 201804
weily105@hotmail.com

Abstract

This paper presented a novel approach, based on an integrated use of fuzzy and rough set theories, to greatly reduce data redundancy. The information framework in rough set theory is introduced in fuzzy-rough set. Mutual information-based algorithm for attribute reduction in fuzzy–rough set is introduced and illustrated with a simple example. Experimental result shows that it is an effective technique.

1. Introduction

The classical rough set theory[1] was proposed by Z. Pawlak in 1982. It provides a kind of formal framework for the automated transformation of data into knowledge. Its main idea is to generate classification rules of concepts induced by partitioning via knowledge reduction, while preserving the ability of classification. With many advantages, rough sets has been applied widely in various application fields, such as machine learning, pattern recognition, information retrieval, data mining and so on. But it has its own limitations, such as sensitivity to noises, and lack of the disposal methods for fuzzy data sets. As a result, many scholars take interest in researching generalized rough sets so as to enlarge the application domain. Then, rough-fuzzy set theory[2] and fuzzy rough set theory[3] were brought forward.

Unfortunately, there is few effective ways to do knowledge reduction in fuzzy rough set theory. An approach based on the extension of rough set in algebraic frame was proposed in [8]. But it is hard to understand and the proceeding is also difficult to operate. In this paper, a new method based on the information framework of fuzzy rough set theory was proposed. The mathematic background of this approach is understandable. An example also shows that the algorithm is effective.

The rest of the article is organized as follows: Section 2 discusses the primary concepts of rough set theory in information frame, in particular focusing on dimensionality reduction based on mutual information. Section 3 introduces the algorithm for fuzzy–rough attribute reduction. To help the understanding of this procedure, a demonstrative case is given to show the key stages involving the use of the former mentioned. Finally, a summery will be delivered, with the future research work.

2. Background

The theory of rough sets offers rigorous mathematical techniques on data analysis, optimization and recognition. However, it is not easy for us to understand the essence of knowledge under the algebra frame. Therefore, information interpretation is introduced into the theory of rough sets. First, we'd like to review some correlative concepts in information representation.

2.1 Rough set attribute reduction

Rough sets attribute reduction has been employed to remove redundant conditional attributes from discrete-valued data sets, while retaining their information content. A successful way of this is the rough set attribute reduction (**MIBARK**) method[4]. Here we illustrate some relative concepts.

Definition 1[6] (Probability Distribution of Knowledge) Let U be a universe, P,Q denote a family of equivalence relations on the universe. According to the papers [6], P,Q may be considered as random variables on the $\sigma-$algebra that is composed of the subsets of the universe U. Let X,Y be two kind of partitions of the universe induced respectively by P,Q, where

$X=U/IND(P)=\{X_1,X_2,\cdots,X_n\}$,

$Y=U/IND(Q)=\{Y_1,Y_2,\cdots,Y_m\}$,

then probability distributing of X,Y are defined respectively by:

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

$$[X;p]=\begin{bmatrix} X_1 & X_2 & \cdots & X_n \\ p(X_1) & p(X_2) & \cdots & p(X_n) \end{bmatrix},$$

$$[Y;p]=\begin{bmatrix} Y_1 & Y_2 & \cdots & Y_m \\ p(Y_1) & p(Y_2) & \cdots & p(Y_m) \end{bmatrix},$$

where $p(X_i)=\frac{card(X_i)}{card(U)}, i=1,2,...,n$; $p(Y_j)=\frac{card(Y_j)}{card(U)}, j=1,2,...,m$

The" $card(.)$ "denotes the cardinality of a set.

Having defined probability distribution of knowledge, we can give easily the definitions of information entropy, conditional entropy and mutual information according to information theory.

Definition 2[4] (Information Entropy $H(P)$ of Knowledge P) The information entropy $H(P)$ of the knowledge P is defined by:

$$H(P)=-\sum_{i=1}^{n} p(X_i)\log p(X_i) \qquad (1)$$

The entropy is a nonnegative function , $H(P)\geq 0$. It may be interpreted as a measure of the uncertainty about knowledge P . Information entropy reaches the maximum value $\log|U|$, when the knowledge P becomes finest. The minimum value 0 is obtained, when the distribution of the knowledge P focuses on a particular value x_0 , i.e. $p(x_0)=1$ and $p(x)=0, x\neq x_0$. Entropy depends on the probabilities.

Definition 3[4] (Conditional Entropy $H(Q|P)$ of the knowledge Q by the knowledge P) The Conditional Entropy $H(Q|P)$ of the knowledge Q by the knowledge P is expressed by:

$$H(Q|P)=-\sum_{i=1}^{n} p(X_i)\sum_{j=1}^{m} p(Y_j|X_i)\log p(Y_j|X_i) \qquad (2)$$

Conditional entropy is nonnegative and non-symmetric, namely, $H(Q|P)\geq 0$ and in general $H(P|Q)\neq H(Q|P)$. It measures the additional amount of information provided by Q if P is already known.

Definition 4[4] (Mutual Information $I(P;Q)$ of the knowledge P and Q) Mutual information can be defined by using entropy and conditional entropy as follows:

$$I(P;Q)=H(Q)-H(Q|P) \qquad (3)$$

Mutual information measures the decrease of uncertainty about P caused Q , and its inverse is the same. It measures the amount of information about P contained in Q or Q contained in P .The mount of information contained in P about itself is obviously $H(P)$, namely, $I(P;P)=H(P)$.

Having the definition of mutual information above, if we want to operate knowledge reduction, we must find the standard of importance for an attribute in a decision table. By calculating the change in mutual information when an attribute is added to the set of considered conditional attributes, the significance of the attribute can be obtained. The higher the change in mutual information, the more significant the attribute is.

Definition 5[4] (Significance of Conditional Attribute) Suppose $T=(U,C\bigcup D,V,f)$ be a decision table, $R\subseteq C$. For arbitrary attribute $a\in C-R$, the significant of attribute a is expressed as:

$$\begin{aligned} SGF(a,R,D) &= I(R\bigcup\{a\};D)-I(R;D) \\ &= H(D|R)-H(D|R\bigcup\{a\}) \end{aligned} \qquad (4)$$

If $R=\varnothing$, $SGF(a,R,D)$ becomes $SGF(a,D)=H(D)-H(D|a)=I(a;D)$ .That means the mutual information between conditional attribute a and decision attribute D. When the value of $SGF(a,R,D)$ is higher, it implies, under the known condition of R , attribute a is more important for decision attribute D.

2.2 Reduct

The **MIBARK** (Mutual Information-Based Algorithm for Reduction of Knowledge) algorithm given in Algorithm 1, attempts to calculate a minimal reduct without exhaustively generating all possible subsets. It starts off with relative core and adds in turn, one at a time, those attributes that result in the greatest increase in $SGF(a,R,D)$, until this produces its ending condition. It has been proved that this method does not always generate a minimal reduct, but it does result in a close-to-minimal reduct, which is still useful in greatly reducing data set dimensionality

Algorithm 1: **MIBARK**

(Mutual Information-Based Algorithm for Reduction of Knowledge)

Step1: Compute the mutual information $I(C;D)$ between conditional attribute C and decision attribute D in the decision table T ;

0-7695-2874-0/07 $25.00 © 2007 IEEE

Step2: Compute the relative core $C_0 = CORE_D(C)$; Generally speaking, $I(C_0;D) < I(C;D)$; Sometimes, relative core $C_0 = \varnothing$, meanwhile, $I(C_0;D) = 0$;

Step3: Let $B = C_0$, for conditional attribute sets $C - B$ repeat:

1. For every attribute $p \in C - B$, compute conditional mutual information $I(p;D|B)$;
2. Select the attribute which brings the maximum of conditional mutual information $I(p;D|B)$, then record it as p (if exists multi attributes achieving the maximum at the same time, choose one whose combination with B reaching least as p; then $B \Leftarrow B \bigcup \{p\}$;
3. if $I(B;D) = I(C;D)$, end; otherwise, go to 1;

Step4: Last, conditional attribute sets B is a relative reduction we need.

An intuitive understanding of **MIBARK** implies that, for a dimensionality of M, the complexity of this algorithm is $O(M^2)$.

3. Fuzzy-Rough Attribute Reduction

The **MIBARK** process described previously can only operate effectively with data sets containing discrete values. While for decreasing the lose of information most decision tables contain fuzzy attributes, Therefore, the method for fuzzy rough set attribute reduction is urgently needed. Here the fuzzy information entropy and fuzzy conditional entropy are exploited in the process of dimensionality reduction. By using fuzzy–rough sets, it is possible to use this information to better attribute reduction. This forms the central contribution of this paper.

3.1 Mathematical Techniques for Fuzzy-Rough Attribute Reduction

One technique for fuzzy-rough attribute reduction based on the dependence degrees of attributes is studied [8]. Nevertheless, it lacks intuition, which is just the advantage of information entropy method. Now, we discuss mainly the expression of information entropy in fuzzy decision table.

Firstly, we rewrite Formula (1), (2):

Suppose $U = \{x_1, x_2, ..., x_n\}$, P, Q (viz. Knowledge) are two equivalence relation on U. The partitions on U being exported respectively by P, Q are X, Y : $X = \{X_1, X_2, ..., X_n\}$, $Y = \{Y_1, Y_2, ..., Y_m\}$. $\forall X_i \in X, Y_j \in Y$ are all the crisp sets. We use the membership function, then for $\forall X_i \in X$ and $x_k \in U$,

$$\mu_{X_i}(x_k) = \begin{cases} 1 & x_k \in X_i \\ 0 & x_k \notin X_i \end{cases}, \mu_{Y_j}(x_k) = \begin{cases} 1 & x_k \in Y_j \\ 0 & x_k \notin Y_j \end{cases}$$

Consequently, $p(X_i) = \frac{|X_i|}{|U|}$ can be expressed as $p(X_i) = \frac{\sum_{k=1}^{|U|} \mu_{X_i}(x_k)}{|U|}$, $i = 1, 2, ..., n$; the same means, $p(Y_j) = \frac{\sum_{k=1}^{|U|} \mu_{Y_j}(x_k)}{|U|}$, $j = 1, 2, ..., m$. Thereby, formula (1) can be rewritten:

$$H(P) = -\sum_{i=1}^{n} p(X_i) \log p(X_i)$$
$$= -\sum_{i=1}^{n} \frac{\sum_{k=1}^{|U|} \mu_{X_i}(x_k)}{|U|} \log \frac{\sum_{k=1}^{|U|} \mu_{X_i}(x_k)}{|U|} \qquad (5)$$

And formula (2) can be denoted as

$$H(Q|P) = -\sum_{i=1}^{n} p(X_i) \sum_{j=1}^{m} p(Y_j | X_i) \log p(Y_j | X_i)$$
$$= -\sum_{i=1}^{n} p(X_i) \sum_{j=1}^{m} \frac{p(Y_j \cap X_i)}{p(X_i)} \log \frac{p(Y_j \cap X_i)}{p(X_i)}$$
$$= -\sum_{i=1}^{n} \frac{\sum_{k=1}^{|U|} \mu_{X_i}(x_k)}{|U|} \sum_{i=1}^{m} \frac{\sum_{k=1}^{|U|} \mu_{X_i \cap Y_j}(x_k)}{\sum_{k=1}^{|U|} \mu_{X_i}(x_k)} \log \frac{\sum_{k=1}^{|U|} \mu_{X_i \cap Y_j}(x_k)}{\sum_{k=1}^{|U|} \mu_{X_i}(x_k)}$$

(6) According to the format conversion, we could apply them into fuzzy-rough set. We define a fuzzy decision table first.

Definition 6 Suppose $U = \{x_1, x_2, ..., x_N\}$, fuzzy attribute sets $\widetilde{A}$ are composed of a group of fuzzy attribute $\{\widetilde{A}^1, \widetilde{A}^2, ..., \widetilde{A}^M, \widetilde{A}^{M+1}\}$. $D = \{\widetilde{A}^{M+1}\}$ is fuzzy decision attribute. Others are fuzzy conditional attributes $C = \{\widetilde{A}^1, \widetilde{A}^2, ..., \widetilde{A}^M\}$. Each fuzzy attribute can partition the U into p_j fuzzy equivalence classes, namely,

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

$F(\widetilde{A}^j)=\{\widetilde{F}_1^j,\widetilde{F}_2^j,...,\widetilde{F}_{p_j}^j\}(j=1,2,...,M+1)$, $\widetilde{F}_i^j(1\le i\le p_j)$ is a fuzzy set. We call the information system $S=(U,\widetilde{A})$ a fuzzy decision table.

Next, we describe the information entropy and conditional entropy in fuzzy-rough set in the fuzzy decision table defined above.

Definition 7 Suppose a fuzzy decision table $S=(U,\widetilde{A})$. P,Q are fuzzy equivalence relation. $U/IND(P)=\{\widetilde{X}_1,\widetilde{X}_2,...,\widetilde{X}_n\}$, $U/IND(Q)=\{\widetilde{Y}_1,\widetilde{Y}_2,...,\widetilde{Y}_m\}$. $\forall \widetilde{X}_i\in U/IND(P),\widetilde{Y}_j\in U/IND(Q)$ are all fuzzy sets on U , then the entropy of knowledge P can be defined as

$$H(P)=-\sum_{i=1}^{n}p(\widetilde{X}_i)\log p(\widetilde{X}_i)$$

$$=-\sum_{i=1}^{n}\frac{\sum_{k=1}^{|U|}\mu_{\widetilde{X}_i}(x_k)}{|U|}\log\frac{\sum_{k=1}^{|U|}\mu_{\widetilde{X}_i}(x_k)}{|U|} \qquad (7)$$

The conditional entropy $H(Q\,|\,P)$ is expressed as

$$H(Q\,|\,P)=-\sum_{i=1}^{n}p(\widetilde{X}_i)\sum_{j=1}^{m}p(\widetilde{Y}_j\,|\,\widetilde{X}_i)\log p(\widetilde{Y}_j\,|\,\widetilde{X}_i)$$

$$=-\sum_{i=1}^{n}\frac{\sum_{k=1}^{|U|}\mu_{\widetilde{X}_i}(x_k)}{|U|}\sum_{i=1}^{m}\frac{\sum_{k=1}^{|U|}\mu_{\widetilde{X}_i\cap\widetilde{Y}_j}(x_k)}{\sum_{k=1}^{|U|}\mu_{\widetilde{X}_i}(x_k)}\log\frac{\sum_{k=1}^{|U|}\mu_{\widetilde{X}_i\cap\widetilde{Y}_j}(x_k)}{\sum_{k=1}^{|U|}\mu_{\widetilde{X}_i}(x_k)} \qquad (8)$$

$U/IND(P)=\otimes U/IND\{\tilde{A}^j\},\tilde{A}^j\in P$,

$U/IND(Q)=\otimes U/IND\{\tilde{A}^j\},\tilde{A}^j\in Q$. And

$$\widetilde{T}_1\otimes\widetilde{T}_2=\{\widetilde{X}\cap\widetilde{Y}:\forall\widetilde{X}\in\widetilde{T}_1,\forall\widetilde{Y}\in\widetilde{T}_2,\widetilde{X}\cap\widetilde{Y}\ne\varnothing\}$$

Moreover, $\mu(\cdot)$ is the membership function of fuzzy set. And

$\mu_{\widetilde{T}_1\cap\widetilde{T}_2\cap...\cap\widetilde{T}_n}(x)=\min\{\mu_{\widetilde{T}_1}(x),\mu_{\widetilde{T}_2}(x),...,\mu_{\widetilde{T}_n}(x)\}$. $\widetilde{T}_i$ is the fuzzy set on U .

Specially, when the fuzzy equivalence relation degenerates crisp equivalence relation, $H(P)$ also degenerates the information entropy of knowledge P in crisp rough sets and $H(Q\,|\,P)$ becomes the normal conditional entropy $H(Q\,|\,P)$ as well. As with classical rough sets, the information entropy of P is related to the proportion of objects that are discernible out of the entire data set. In the present approach, this corresponds to determining the fuzzy cardinality divided by the total number of objects in the universe.

Now, we extend the concept of mutual information to fuzzy-rough set, which is used to weigh the relative significance of fuzzy attribute in fuzzy decision table.

Suppose a fuzzy decision table $S=(U,\widetilde{A})$. $\Re$ is the set of fuzzy conditional attributes. Then, we add a fuzzy attribute $\tilde{A}^j$. The increment of mutual information is

$$I(\Re\cup\{\tilde{A}^j\};D)-I(\Re;D)=H(D\,|\,\Re)-H(D\,|\,\Re\cup\{\tilde{A}^j\}) \qquad (9)$$

Definition 8 Suppose a fuzzy decision table $S=(U,\widetilde{A})$. $\Re$ is the set of fuzzy conditional attributes. Then for all the $\tilde{A}^j\in C-\Re$, the significance $SGF(\tilde{A}^j,\Re,D)$ could be expressed as

$$SGF(\tilde{A}^j,\Re,D)=I(\Re\cup\{\tilde{A}^j\};D)-I(\Re;D)$$
$$=H(D\,|\,\Re)-H(D\,|\,\Re\cup\{\tilde{A}^j\}) \qquad (10)$$

If $\Re=\varnothing$, $SGF(\tilde{A}^j,\Re,D)$ can be written as $SGF(\tilde{A}^j,D)=H(D)-H(D\,|\,\tilde{A}^j)=I(\tilde{A}^j;D)$. It means the mutual information of fuzzy attribute $\tilde{A}^j$ and fuzzy decision attribute D . Fuzzy attribute $\tilde{A}^j$ is more important on fuzzy decision attribute D when the value of $SGF(\tilde{A}^j,\Re,D)$ increases.

3.3 Attribute Reduction Algorithm for Fuzzy-Rough Set

For these basic concepts above, next, we advance the detailed **MIBAFRAR** (Mutual Information-Based Algorithm for Fuzzy-Rough Attribute Reduction) method. It seeks the relative reduction from bottom to up while start with an empty-set. The process of this algorithm is: selecting the most significant attribute to add to relative potential reduct one by one, according to the significance of conditional attribute

$SGF(\tilde{A}^j,\Re,D)$ defined previously, until the ending condition is satisfied. As will be shown, the process becomes identical to the traditional approach when dealing with nominal well-defined attributes.

Algorithm 2: **MIBAFRAR**

(Mutual Information-Based Algorithm for Fuzzy-Rough Attribute Reduction)

Step1: Compute the mutual information $I(C;D)$ between fuzzy conditional attribute C and fuzzy decision

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

attribute D in the fuzzy decision table;

Step2: Let $\Re = \varnothing$, for conditional attribute sets $C - \Re$ repeat:

1. For every attribute $\tilde{A}^j \in C - \Re$, compute conditional mutual information $I(\tilde{A}^j; D \mid \Re)$;
2. Select the attribute which brings the maximum value of conditional mutual information $I(\tilde{A}^j; D \mid \Re)$, then record it as $\tilde{A}^j$ (if exists multi attributes achieving the maximum at the same time, choose one having the least number of equivalence classes as $\tilde{A}^j$); then $\Re \Leftarrow \Re \cup \{\tilde{A}^j\}$;
3. If $I(C; D) = I(\Re; D)$, end; otherwise, goto 1;

Step3: Last, conditional attribute set $\Re$ is a relative reduction we need.

3.4 An example

To illustrate the operation of fuzzy-rough set attribute reduction, an example is given here. In the new approach, membership degrees are used in calculating the fuzzy information entropy.

Table 1 **A Fuzzy Decision Table**

	Temperature			Humidity		Windy		Class	
	Hot	Mild	Cool	High	Normal	Flase	Ture	Positive	Negative
1	0.9	0.1	0	0.8	0.2	0.7	0.4	0.4	0.7
2	0.8	0.2	0.1	0.9	0.2	0.1	0.8	0.3	0.7
3	0.9	0.1	0.1	0.9	0.1	0.9	0.1	0.8	0.3
4	0.1	0.9	0	0.6	0.5	0.8	0.3	0.6	0.5
5	0	0.1	0.9	0	0.1	0.8	0.2	0.9	0.2
6	0	0.2	0.9	0.1	0.9	0.1	0.9	0.3	0.8

Using the fuzzy sets defined in Table 1, for all the conditional attributes (“Temperature”, “Humidity”, “Windy”), the following equivalence classes are obtained:

$U/"Temperature" = \{"Hot", "Mild", "Cool"\}$,

$U/"Humidity" = \{"High", "Normal"\}$,

$U/"Windy" = \{"False", "Ture"\}$.

 The first step is to calculate $I(C; D) = 0.1480$.

 Set $\Re = \varnothing$, for $\forall \tilde{A}^j \in C - \Re$, compute conditional mutual information $I(\tilde{A}^j; D \mid \Re)$ as follows: $I(\tilde{A}^1; D) = 0.0458$, $I(\tilde{A}^2; D) = 0.1205, I(\tilde{A}^3; D) = -0.0024$.

From this it can be seen that attribute “Humidity” will cause the greatest increase in mutual information. This attribute is chosen and added to the potential reduct $\Re = \{\tilde{A}^2\}$, then update $I(\Re; D) = I(\{\tilde{A}^2\}; \Re) = 0.1205$.

Similarly, the process iterates and we can attain that “Windy” is the maximum attribute of $I(\tilde{A}^j; D \mid \Re)$. Thus, renew $\Re = \{\tilde{A}^2, \tilde{A}^3\}$; compute $I(\Re; D) = 0.1472$.

A problem may arise when this approach is compared to the crisp approach. In conventional **MIBARK**, a reduct is defined as a subset B of the attributes which have the same mutual information content as the full attribute set C. In terms of the mutual information this means that the values $I(C; D)$ and $I(\Re; D)$ are identical absolutely if the data set is consistent. However, in the fuzzy–rough approach this is not satisfied strictly because of computing error in counting the value of fuzzy equivalence class membership degree which an object belong to. A possible way of overcoming this would be to determine the precise degree and use this as the terminative condition.

Here, we assume $|I(C; D) - I(\Re; D)| \leq 10^{-3}$. As this causes nearly no increase in mutual information, the algorithm stops and outputs the reduct {“Humidity”, “Windy”}. The steps taken by the fuzzy-rough set attributes reduction based on mutual information algorithm in this decision can be seen in Table 2. The data set can now be reduced to only those attributes appearing in the reduct.

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

Table 2 **Fuzzy Decision Table after Reduction**

	Humidity		Windy		Class	
	High	Normal	Flase	Ture	Positive	Negative
1	0.8	0.2	0.7	0.4	0.4	0.7
2	0.9	0.2	0.1	0.8	0.3	0.7
3	0.9	0.1	0.9	0.1	0.8	0.3
4	0.6	0.5	0.8	0.3	0.6	0.5
5	0	0.1	0.8	0.2	0.9	0.2
6	0.1	0.9	0.1	0.9	0.3	0.8

4. Conclusion

Rough Set theory is a kind of tool for dealing with imprecise knowledge but fails in managing knowledge under fuzzy condition and continuous attributes. In rough sets, knowledge is defined based on equivalence relation and inclusive relation of algebra so that the essence of knowledge is not easy to be understood. It is helpful to understand the rough set theory better by studying the concepts via information frame for rough sets. In this paper, we address knowledge expression and knowledge reduction in information frame for fuzzy decision table. Based on these new measures, the problem of knowledge reduction in fuzzy information systems is solved. By constructing an example, we prove that the technique is valid. The work of this paper extends the research of information frame in rough set to fuzzy rough set and establishes one direction for seeking efficient algorithm of knowledge acquisition in fuzzy information systems

5. Acknowledge

The research is supported by the National Science Foundation of China under grant No: 60475019

6. Reference

[1] Z.Pawlak. Rough Sets. *International Journal of Information and Computer Science*, 11:341-356, 1982

[2] M Banerjee, Sankar K Pal. Roughness of A Fuzzy Set. *Information and Computer Science*, 93:235-245, 1996

[3] Dubois D, Prade H. Rough Fuzzy Sets and Fuzzy Rough Sets. *Information and Computer Science*, 17:191-209, 1990

[4] D.Q Miao, G. R Hu. A Heuristic Algorithm for Reduction of Knowledge. *Journal of Computer Research and Development*, 36(6):681-684, 1999

[5] Dongyi Ye, Zhaotong Chen. A new difference matrix and its methods of core finding. *Electronic Journal*, 30(7):1086-1088, 2002

[6] D.Q Miao, J Wang. An information representation of the concepts and operations in rough set theory. *Journal of Software*, 10(2):113-116, 1999

[7] Xi Zhao Wang, Yan Ha, De Gang Chen. On The Reduction of Fuzzy Rough Sets. *In: Proceeding of the Third International Conference on Machine Learning and Cybernetics*, Guangzhou, 18-21:3175-3178, 2005

[8] R. Jensen, Q. Shen. Fuzzy–rough sets for descriptive dimensionality reduction. *Proc. 11th Internat. Conf. on Fuzzy Systems*, pp. 29–34, 2002

[9] Eric C C Tsang, De Gang Chen, John W T Lee, Daniel S Yeung. On The Upper Approximation of Covering Generalized Rough Sets. *In: Proceeding of the Third International Conference on Machine Learning and Cybernetics*, Shanghai, 26-29:4200-4203, 2004

[10] Daniel S Yeung, De Gang Chen, Eric C C Tsang, John W T Lee. On the Generalization of Fuzzy Rough Sets. *IEEE Transaction on Fuzzy System*, 13(3):343-361, June 2005

[11] Weizhi Wu, Jusheng Mi, Wenxiu Zhang. Generalized Fuzzy Rough sets. *Information Science*, 151:263-282,2003

[12] G. Y. Wang. *Rough Set Theory and Knowledge Discovery.* Xi'an JiaoTong University Press, Xi'an, 2001