Proceedings of the 4th National Conference; INDIACom-2010
Computing For Nation Development, February 25 – 26, 2010
Bharati Vidyapeeth's Institute of Computer Applications and Management, New Delhi

Dimension Reduction Using Rough Set Theory For Intrusion Detection System

Nandita Sengupta[1] and **Jaya Sil**[2]
[1]University College of Bahrain, Bahrain, [2]Bengal Engineering and Science University Shibpur
E-Mail: [1]ngupta@ucb.edu.bh, [2]js@cs.becs.ac.in

ABSTRACT
This paper briefly describes Information Detection System management by reducing dimension of the information system(IS) for network traffic both row wise and column wise. For dimension reduction, Rough Set Theory (RST) is used where determination of reduct and core plays a major role. But finding reduct is difficult task for a huge IS. Our paper suggests a technique where we propose to divide the complete IS table into different subdivisions as per the values of decision attribute. Then in each subdivided IS table, some conditional attributes are combined to reduce column size of the subdivided IS table. In this reduced subdivided IS table significance of the attribute is determined and reduct is found out.

KEYWORDS
Dimension reduction, RST, reduct, core

1. INTRODUCTION

In every twenty months or so the amount of information in the world gets doubled. In every field it becomes very important to retrieve the information in proper time. This is the reason a lot of research scholars are working for efficient information retrieval. In this paper we have concentrated on dimension reduction technique for classification of network traffic for intrusion detection system by using some soft computing technique. We have also reviewed current research work on dimension reduction using rough set theory and other techniques which we have highlighted in this paper. Classification is one of the important key for information retrieval. But before classification reduction of dimension is a bottle neck because of enormous electronic data in any field. Finding out reduct and core is a must for reducing IS table while doing classification through set approximation of RST. We are proposing a method where before finding out reduct we need to reduce the attributes for the entire IS table. We would suggest to reduce the number of attributes by using correlation feature selection method then finally Rough Set Theory would be applied by calculating the significance of each remaining attribute.

2. SET APPROXIMATIONS OF ROUGH SET THEORY

Rough Set Theory (RST) is an extension of set theory which helps in classification and decision making. In RST, classification is done by using the concept of Lower and Upper Approximation. Lower approximation is a set of system objects that are known with certainty to belong to the subset of interest, whereas the upper approximation is a description of the objects that possibly belong to the subset. An information system is considered as table which is considered as decision system. Information system is consisting of rows (objects) and columns (attributes). Mathematically, information system can be represented as I = (U,A) where U is a nonempty finite set of objects and A is a nonempty finite set of attributes such that $a: U \rightarrow V_a$ for every $a \in A$. V_a is the set of values that attribute a may take. For decision systems, $A = \{C \cup D\}$, where C is the set of conditional attributes and D is the set of decision attributes.

3. POSITIVE, NEGATIVE AND BOUNDARY REGION

Rough set is a pair of set approximations, lower approximation and upper approximation. Lower approximation is also known as positive region. $\underline{C}X$ is the set of indiscernibility classes which are subsets of X. If X is characterized by a particular decision value, it means that all indiscernibility classes in $\underline{C}X$ contain objects with that value. There are also indiscernibility classes which contain only some tuples in X. In this case, we cannot classify them. These are the objects in boundary region, mathematically, it can be represented as $\overline{C}X - \underline{C}X$. The elements which are in U-$\overline{C}X$, which contain tuples not in X, exist in negative region. If $\underline{C}X = \overline{C}X$, then boundary region is empty, and X is said to be crisp or precise.

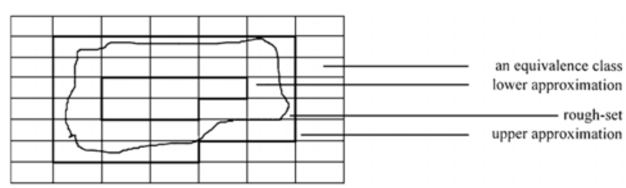

Figure 1:Rough Set with Upper Approximation and lower Approximation

4. PRESENT WORK ON DIMENSION REDUCTION

M Zhang et.al. proposed a new rough set based approach PASH (Parameterized Average Support Heuristic) in their paper[1]. PASH selects features causing high average support of rules over all decision classes. As of 1997, when a special issue on relevance including several papers on variable and feature selection was published (Blum and Langley, 1997, Kohavi and John, 1997), few domains explored used more than 40 features. The situation has changed considerably in the past few years and, in this special issue, most papers explore domains with hundreds to tens of thousands of variables or features [2]. New techniques are proposed to address these challenging tasks involving many irrelevant and redundant variables. Keyun Hu et. al. proposed a novel feature ranking technique using discernibility matrix in their paper [3]. By making use of attribute frequency information in discernibility matrix, we develop a fast feature ranking mechanism. Based on the mechanism, two heuristic reduct computation algorithms are proposed. One is for optimal reduct and the other for approximate reduct. They have also reported empirical results. In paper [4], a pipelined approach using two clustering algorithms in combination with rough sets is investigated for the purpose of discovering important combination of attributes in high dimensional data. Several k-means algorithms are used as fast procedures for simplification of the attribute set of the information systems presented to the rough sets algorithms. The data submatrices described in terms of these features are then discretized w.r.t the decision attributes according to different rough set based schemes. From them, the reducts and their derived rules are extracted, which are applied to test data in order to evaluate the resulting classification accuracy. An exploration of this approach (using Leukemia gene expression data) was conducted in a series of experiments within a high-throughput distributed-computing environment. They led to subsets of genes with high discrimination power. Good results were obtained with no preprocessing applied to the data. Jennifer G. Dy et.al. explore the issues involved in developing automated feature subset selection algorithms for unsupervised learning in their paper [5]. Aijun An et.al. present a feature reduction method based on the rough set theory and investigate the effectiveness of the rough set feature selection method on web page classification in their paper [6]. Their experiments indicate that rough set feature selection can improve the predictive performance when the original feature set for representing web pages is large. Huan Liu et. al. described in their paper [7] that feature selection is associated with dimension reduction, irrelevant data removal, learning accuracy increasing and improvement of comprehensibility. In this paper, they consider the problem of active feature selection in a filter model setting. They described active feature selection called selective sampling. They demonstrated it by applying it to a widely used feature selection algorithm, "Relief", and they have shown how it realizes active feature selection and reduces the required number of training data for "Relief" to achieve time savings without performance deterioration. In paper [8], Rajen B. Bhatt et. al. have shown that the fuzzy-rough set attribute reduction algorithm [Jenson, R., Shen, Q., 2002. Fuzzy-rough sets for descriptive dimensionality reduction. In: Proceedings of IEEE International Conference on Fuzzy Systems, FUZZ-IEEE'02, May 12-17, pp. 29-34] is not convergent on many real datasets due to its poorly designed termination criteria; and the computational complexity of the algorithm increases exponentially with increase in the number of input variables and in multiplication with the size of data patterns. Based on natural properties of fuzzy t-norm and t-conorm, they have put forward the concept of fuzzy-rough sets on compact computational domain, which is then utilized to improve the computational efficiency of FRSAR algorithm. They have achieved speed up factor as high as 622 with this concept with improved accuracy. They also restructure the algorithm with efficient termination criteria to achieve the convergence on all the datasets and to improve the reliability of selected set of features. In paper [9], Grzegorz Gora et.al. described a method combining two widely-used empirical approaches to learning from examples: rule induction and instance-based learning. In their algorithm (RIONA), decision is predicted not on the basis of the whole support set of all rules matching a test case, but the support set restricted to a neighbourhood of a test case. The size of the optimal neighbourhood is automatically induced during the learning phase. According to them, the empirical study shows the interesting fact that it is enough to consider a small neighbourhood to achieve classification accuracy comparable to an algorithm considering the whole learning set. The combination of k-NN and a rule-based algorithm results in a significant acceleration of the algorithm using all minimal rules. Moreover, they presented classifier which has high accuracy for both kinds of domains: more suitable for k-NN classifiers and more suitable for rule based classifiers. Application of rough sets and statistical methods is applied for feature reduction and pattern recognition in the paper [10]. The role of rough set reducts in feature selection and data reduction in pattern recognition is emphasized by the description of rough set theory presented in this paper. An algorithm is described for feature selection and reduction based on rough sets methods proposed jointly with Principal Component Analysis. In the paper [11], the authors proposed an algorithm which is using rough set theory with greedy heuristics for feature selection. Selecting features is similar to the filter approach, but the evaluation criterion is related to the performance of induction. That is, they select the features that do not damage the performance of induction. In paper [12], Lawrence J. Mazlack et. al. suggest that by progressively partitioning the data to reduce intra-item dissonance within the resulting partitions, cohesion will be enhanced. The rationale is that by recursively partitioning the dataset, dissonance will be discovered and be eliminated within the subpartitions. Thus, they mentioned that the coherence of the resulting partitions will be greater than the coherence of the initial data set. Jensen et. al. specified in their paper [13] that fuzzy rough sets can be used as a tool to discover the data dependencies and reduce number of attributes contained in a data set by purely structural methods. The Fuzzy Rough Attribute Reduction (FRAR) can be applied to datasets where conditional and decision attribute values are crisp or fuzzy. Fan Jinsong et. al. mentioned in their paper [14] that an integrated approach combined Rough Set Theory and SVM have been applied for classification. The first step is classified roughly with Rough Set. The second step should be classified precisely based on SVM algorithm. In paper [15], it is mentioned that Rough Sets Theory is a powerful mathematical tool to handle vagueness and uncertainty inherent in making decisions. In paper [16], authors have presented an expressive language for representing knowledge about vague concepts. It is based on the rough set formalism and it caters for implicit definition of rough relations by combining different regions (e.g. upper approximation, lower approximation, and boundary) of other rough relations. They mentioned that the semantics of the proposed language is obtained by translating it to the language of extended logic programs whose meaning is captured by paraconsistent stable models. A query language is also discussed to retrieve information about the defined rough relations. Y.Y. Yao mentioned in his paper [17] that granulation of the universe and approximation of concepts in the granulated universe are two related fundamental issues in the theory of rough sets. He specified in his paper that the basic ideas of information granulation have appeared in fields, such as interval analysis, quantization, rough set theory, the theory of belief functions, divide and conquer, cluster analysis, machine learning, databases, and many others. There is a fast growing and renewed interest in the study of information granulation and computations under the umbrella term of Granular Computing (GrC), covering theories, methodologies, techniques, and tools that make use of granules in the process of problem solving. Günther Gediga et. al. described an algorithm in their paper [18] to impute missing values from given data alone and analyse its performance. Their approach is based on non–numeric rule based data analysis. They mentioned that their analysis offers no straightforward way to define loss functions or a likelihood function; these are based on statistical assumptions, which are not given in rule based data analysis. Therefore, other optimisation criteria must be used. A simple criterion is the demand that the rules of the system should have a maximum in terms of consistency, which means if we fill a missing entry with a value, we should result in a rule which is consistent with the other rules of the system. Haixuan Yang et. al. proposed in their paper [19] a generalized dependency degree, Γ, between two given sets of attributes, which counts both deterministic and indeterministic rules while γ counts only deterministic rules. To understand Γ better, they investigated its various properties. To show its advantage, they made a comparative study with the conditional entropy and γ in a number of experiments. Experimental results showed that the speed of the new C4.5 using Γ is greatly improved when compared with the original C4.5R8 using conditional entropy, while the prediction accuracy and tree size of the new C4.5 are comparable with the original one. Moreover, Γ achieves better results on attribute selection than γ. A formal model of machine learning by considering user preference of attributes is proposed in paper [20]. According to their opinion, the model seamlessly combines internal information and external information. They also mentioned that this model can be extended to user preference of attribute sets. By using the user preference of attribute sets, user preferred reducts can be constructed.

5. PROPOSED WORK ON DIMENSION REDUCTION

In our paper we propose that before going for determination of reducts, attributes are subdivided as per decision attributes. Total information system table is subdivided as per some conditional and decision attributes. Once the size of the table is reduced, degree of dependency is calculated as per Rough Set Theory and reduct is calculated. In our proposed work we have considered network data set [21] for Intrusion Detection System[22] , where in the training data set 42 attributes are considered for 25192 traffic data as mentioned in table 1. We would like to combine some attributes which are not affecting the classification. We have used software "Tanagra" to represent this data. Here, we want to apply correlation based feature selection and we want to reduce some of these attributes (like duration,land,wrong_fragment,urgent,num_failed_loginsetc.) which are not contributing in traffic determination (whether traffic is normal or anomaly). After reducing some attributes we need to determine "reduct" by using rough set theory.

Attribute	Category	Informations
duration	Continue	-
protocol_type	Discrete	3 values
service	Discrete	66 values
flag	Discrete	11 values
src_bytes	Continue	-
dst_bytes	Continue	-
land	Discrete	2 values
wrong_fragment	Continue	-
urgent	Continue	-
hot	Continue	-
num_failed_logins	Continue	-
logged_in	Discrete	2 values
num_compromised	Continue	-
root_shell	Continue	-
su_attempted	Continue	-
num_root	Continue	-
num_file_creations	Continue	-
num_shells	Continue	-
num_access_files	Continue	-
num_outbound_cmds	Continue	-
is_host_login	Discrete	1 values
is_guest_login	Discrete	2 values
count	Continue	-
srv_count	Continue	-

serror_rate	Continue	-
srv_serror_rate	Continue	-
rerror_rate	Continue	-
srv_rerror_rate	Continue	-
same_srv_rate	Continue	-
diff_srv_rate	Continue	-
srv_diff_host_rate	Continue	-
dst_host_count	Continue	-
dst_host_srv_count	Continue	-
dst_host_same_srv_rate	Continue	-
dst_host_diff_srv_rate	Continue	-
dst_host_same_src_port_rate	Continue	-
dst_host_srv_diff_host_rate	Continue	-
dst_host_serror_rate	Continue	-
dst_host_srv_serror_rate	Continue	-
dst_host_rerror_rate	Continue	-
dst_host_srv_rerror_rate	Continue	-
class	Discrete	2 values

Table 1: Network data for Intrusion Detection System

Dataset Description: 42 attributes and 25192 examples

$x \in U$	a	b	c	d	$\Rightarrow$	e
0	S	R	T	T		R
1	R	S	S	S		T
2	T	R	R	S		S
3	S	S	R	T		T
4	S	R	T	R		S
5	T	T	R	S		S
6	T	S	S	S		T
7	R	S	S	R		S

Table 2: An Example data set

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

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

Note that this relation corresponds to the equivalence relation for which two objects are equivalent if and only if they have the same vectors of attribute values for the attributes in P. The partition of U, determined by IND (P), is denoted by U/IND(P), or U/P, which is simply the set of equivalence classes generated by IND(P):

$$U / IND(P) = \otimes \{U / IND(\{a\}) | a \in P\}$$

Where $A \otimes B = \{X \cap Y \,|\, X \in A, Y \in B, X \cap Y \neq \phi\}$

If $(x, y) \in IND(P)$, then x and y are indiscernible by attributes from P. In the above example, if we consider $P = \{b, c\}$, then $U / IND(P)$ will be calculated as follows:

$$U / IND(P) = U / IND(\{b\}) \otimes U / IND(\{c\})$$
$$= \{\{0,2,4\}, \{1,3,6,7\}, \{5\}\} \otimes \{\{0,4\}, \{1,6,7\}, \{2,3,5\}\}$$
$$= \{\{0,4\}, \{1,6,7\}, \{2\}, \{3\}, \{5\}\}$$

The equivalence classes of the indiscernibility relation with respect to P are denoted $[x]_p, x \in U$.

A. ATTRIBUTE DEPENDENCIES: For data analysis it is very important to find out the dependencies between attributes. If a set of attributes Q depends totally on a set of attributes P, which is denoted as $P \Rightarrow Q$, which means all attribute values from Q are uniquely determined by values of attributes from P. If there exists a functional dependency between values of Q and P, then Q depends totally on P. In Rough Set Theory, dependency is defined in the following way:
For $P, Q \subset A$, it is said that Q depends on P in a degree k $(0 \le k \le 1)$, denoted $P \Rightarrow_k Q$, if $k = \gamma_P(Q) = \frac{|POS_P(Q)|}{|U|}$, where $|S|$ stands for the cardinality of the set S.
If k=1, Q depends totally on P, if 0<k<1, Q depends partially in a degree k. If k=0, Q does not depend on P. In the example, the degree of dependency of attribute {e}, on the attributes {b,c} is

$$\gamma_{\{b,c\}}(\{e\}) = \frac{|POS_{\{b,c\}}(\{e\})|}{|U|} = \frac{|\{2,3,5\}|}{|\{0,1,2,3,4,5,6,7\}|} = \frac{3}{8}$$

Significance of the attribute can be calculated and from that it is possible to derive that which attribute/attributes is/are indispensable. When a feature is removed from the set of features, an estimate of the significance of the feature can be calculated by calculating the change in dependency. If the change in dependency is higher, the feature will be more significant. If the significance is 0, then the feature is dispensable. Dependency of the attribute Q on the attribute P is defined as $\gamma_P(Q)$. If the feature "a" is removed from the feature set P, then dependency of the attribute Q on the attribute (P-a) would be $\gamma_{P-\{a\}}(Q)$. Therefore, the significance of the attribute "a" is calculated as

$$\sigma_P(Q, a) = \gamma_P(Q) - \gamma_{P-\{a\}}(Q)$$

If P={a,b,c} and Q={e}, then

$$\gamma_{\{a,b,c\}}(\{e\}) = \frac{|\{2,3,5,6\}|}{|\{0,1,2,3,4,5,6,7\}|} = \frac{4}{8}$$

$$\gamma_{\{b,c\}}(\{e\}) = \frac{|\{2,3,5\}|}{|\{0,1,2,3,4,5,6,7\}|} = \frac{3}{8}$$

$$\gamma_{\{a,b\}}(\{e\}) = \frac{|\{2,3,5,6\}|}{|\{0,1,2,3,4,5,6,7\}|} = \frac{4}{8}$$

$$\gamma_{\{a,c\}}(\{e\}) = \frac{|\{2,3,5,6\}|}{|\{0,1,2,3,4,5,6,7\}|} = \frac{4}{8}$$

Significance of the three attributes can be calculated as follows:

$$\sigma_P(Q,a) = \gamma_P(Q) - \gamma_{P-\{a\}}(Q) = \frac{4}{8} - \frac{3}{8} = \frac{1}{8}$$

$$\sigma_P(Q,b) = \gamma_P(Q) - \gamma_{P-\{b\}}(Q) = \frac{4}{8} - \frac{4}{8} = 0$$

$$\sigma_P(Q,c) = \gamma_P(Q) - \gamma_{P-\{c\}}(Q) = \frac{4}{8} - \frac{4}{8} = 0$$

Therefore, from the values of significance it is obvious that attribute b and c are dispensable and attribute a is indispensable.

B. REDUCT: A reduct is the minimal attribute set which can preserve the classification of the entire data set[23]. Reduct is defined as a minimal subset R of the initial attribute set C such that for a given set of attributes D, $\gamma_R(D) = \gamma_C(D)$. R is a minimal subset if $\gamma_{R-\{a\}}(D) \neq \gamma_R(D)$ for all $a \in R$, it means that if any attribute is removed from R, then degree of dependency will be affected. Global reduct is defined as the reduct which has smallest cardinality. A dataset can have many reducts. Collection of all reducts is defined as follows

$$R_{all} = \{X \mid X \subseteq C, \gamma_X(D) = \gamma_C(D); \gamma_{X-\{a\}}(D) \neq \gamma_X(D), \forall a \in X\}$$

The intersection of all sets in R_{all} is called the core. If any element from core is eliminated, inconsistency will arise in the representation of dataset. Objective is to find out the global reduct, i.e., to find out $R_{\min}$, where $R_{\min} \subseteq R_{all}$ and

$$R_{\min} = \{X \mid X \in R_{all}, \forall Y \in R_{all}, |X| \leq |Y|\}$$

It is very difficult task to find out the reduct $(R_{\min})$ of an information system. This is a potential area of research.

C. DISCERNIBILITY MATRIX: Discernibility matrix is used for finding reducts. A discernibility matrix of a decision table is a symmetric $|U| \times |U|$ matrix where the matrix elements are defined by

$$c_{ij} = \{a \in C \mid a(x_i) \neq a(x_j)\}$$

$$i,j = 1,2,\ldots\ldots, |U|$$

c_{ij} is the set which contains those attributes which are different between objects i and j. This discernibility matrix considers only those object discernibilities that occur when the corresponding decision attributes differ.

$x \in U$	0	1	2	3	4	5	6	7
0								
1	a,b,c,d							
2	a,c,d	a,b,c						
3	b,c		a,b,d					
4	d	a,b,c,d		b,c,d				
5	a,b,c,d	a,b,c		a,b,d				
6	a,b,c,d		b,c		a,b,c,d	b,c		
7	a,b,c,d	d		a,c,d			a,d	

Table 3: Discernibility Matrix

A discernibility function f_D is a Boolean function of m Boolean variables $a_1^*, a_2^*, a_3^*, \ldots.., a_m^*$

$$f_D(a_1^*, a_2^*, \ldots\ldots, a_m^*) = \wedge\{\vee c_{ij}^* \mid 1 \leq j \leq i \leq |U|, c_{ij} \neq \phi\}$$

The discernibility function is

$$f_D(a^*, b^*, c^*, d^*) = (a^* \vee b^* \vee c^* \vee d^*) \wedge (a^* \vee c^* \vee d^*) \wedge (a^* \vee b^* \vee c^*) \wedge (b^* \vee c^*)$$
$$\wedge (d^*) \wedge (a^* \vee b^* \vee d^*) \wedge (b^* \vee c^* \vee d^*) \wedge (a^* \vee d^*)$$

From which discernibility function is simplified as follows:

$$f_D(a^*, b^*, c^*, d^*) = (b^* \vee c^*) \wedge (d^*)$$

The reducts of the dataset may be obtained by converting the expression above from conjunctive normal form to disjunctive normal form. Hence, the minimal reducts are $\{b, d\} and \{c, d\}$

D. DIMENSION REDUCTION: It is very important to reduce the dimension where huge data objects or objects features are available. In reality it becomes unmanageable with huge amount of data. Therefore, with minimum loss of information dimension reduction needs attention for improvement of efficiency. Before processing the data, redundant data or the data which is misleading should be identified and removed.

6. CONCLUSION

Here, in this paper we suggest a technique of dimension reduction which we would like to apply for Intrusion Detection System. Initially, we would like to apply some correlation feature selection method for reducing the size of the dimension table, column wise and finally we would like to determine the significance of attributes considering each attribute in the reduced information system table by using Rough Set Theory.

7. ACKNOWLEDGMENT

We are thankful to the developer of the software "Tanagra" , which is an efficient software in the filed of machine learning.

8. REFERENCES

[1] M. Zhang, J. T. Yao, "A Rough Sets Based Approach to Feature Selection", International Conference of the North American Fuzzy Information Processing Society No23, Banff AB , CANADA (2004) , pp. 434-439

[2] Isabelle Guyon, Andre Elisseeff, "An Introduction to Variable and Feature Selection", Journal of Machine Learning Research 3 (2003) 1157-1182

[3] Keyun Hu, Yuchang Lu and Chunyi Shi, "Feature Ranking in Rough Sets", AI Communications, Volume 16 , Issue 1 (May 2003), Pages: 41 - 50 , ISSN:0921-7126

[4] Julio J. Valdes and Alan J. Barton, "Relevant attribute discovery in high dimensional data based on rough sets and unsupervised classification: Application to Leukemia gene expressions", published at the Tenth International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC 2005) August 341 – September 3, 2005. Regina, Canada

[5] Jennifer G. Dy, Carla E. Brodley, "Feature Selection for Unsupervised Learning", Journal of Machine Learning Research 5 (2004) 845–889

[6] Aijun An, Yanhui Huang, Xiangji Huang, and Nick Cercone, "Feature Selection with Rough Sets for Web Page Classification", Transactions on Rough Sets II:Rough Sets and Fuzzy Sets Springer 2004, Vol. 3135, pp. 1-13,ISBN 3-540-23990-1

[7] Huan Liu, Hiroshi Motoda, Lei Yu, "Feature Selection with Selective Sampling", Proceedings of the Nineteenth International Conference on Machine Learning table of contents,Pages: 395 – 402,2002,ISBN:1-55860-873-7

[8] Rajen B. Bhatt, M. Gopal, "On fuzzy-rough sets approach to feature selection", Pattern Recognition Letters, Volume 26, , Issue 7 (May 2005) Pages: 965 - 975 ,ISSN:0167-8655, Elsevier Science Inc.

[9] Grzegorz Gora, ArkadiuszWojna, "RIONA: A New Classification System Combining Rule Induction and Instance-Based Learning", Fundamenta Informaticae XX (2002) 1–22,IOS Press

[10] Roman W. Swiniarski, "Rough Sets Methods in Feature Reduction and Classification", Int. J. Appl. Math. Comput. Sci., 2 001, Vol.11, No.3, 565-582

[11] Ning Zhong, Juzhen Dong, "Using Rough Sets with Heuristics for Feature Selection", Journal of Intelligent Information Systems, 16, 199–214, 2001

[12] L.J.Mazlack, Aijing He, Y Zhu, Sarah Coppock, "A Rough Set Approach in Choosing Partitioning Attributes", Proceedings of the ISCA 13th International Conference (CAINE-2000), November, 2000, pp.1-6, 2000

[13] Richard Jensen and Qiang Shen, "Aiding Fuzzy Rule Induction using Fuzzy Rough Attribute Reduction", Proceedings of the 2002 UK Workshop on Computational Intelligence,pp. 81-88,2002

[14] Fan Jinsong, Fang Tingjian, "Chinese Character Classification Based on Rough Set and SVM Algorithm", MVA200 IAPR Workshop on Machine Vision Applications, Nov 28-30 2000

[15] Francis E.H. Tay, Lixiang Shen, "Economic and financial prediction using rough sets model", European Journal of Operational Research 141 (2002) 641–659

[16] Aida Vit´oria, Carlos Viegas Dam´asio, Jan Małuszy´nski, "From Rough Sets to Rough Knowledge Bases", Fundamenta Informaticae, Volume 57 , Issue 2-4 (February 2003), Pages: 215 – 246

[17] Y.Y. Yao, "Information Granulation and Approximation in a Decision-theoretic Model of Rough Sets", Rough-neuro Computing: a Way to Computing with Words, Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003

[18] Günther Gediga, Ivo Düntsch, "Maximum consistency of incomplete data via non-invasive imputation", Artificial Intelligence Review, Volume 19 , Issue 1 (March 2003), Pages: 93 – 107

[19] Haixuan Yang, Irwin King and Michael R. Lyu, "The Generalized Dependency Degree Between Attributes", Journal of The American Society for Information Science and Technology, 58(14):2280–2294, 2007

[20] Yiyu Yao, Yan Zhao, Jue Wang and Suqing Han, "A Model of Machine Learning Based on User Preference of Attributes", Proceedings of the Fifth International Conference on Rough Sets and Current Trends in Computing (RSCTC'06), 587-596, 2006

[21] "Nsl-kdd data set for network-based intrusion detection systems."Available on: http://nsl.cs.unb.ca/KDD/NSL-KDD.html, March 2009

[22] Nandita Sengupta, Jaya Sil, "Network Intrusion Detection Using RST, k means and Fuzzy c means clustering", proceeding of the conference ICIP2009, Bangalore,India

[23] Nandita Sengupta, Jaya Sil, "An Integrated Approach to Information Retrieval using RST,FS and SOM", proceeding of the conference ICIS2008, Bahrain.