www.icgst.com

Application of Clustering for Feature Selection Based on Rough Set Theory Approach

K.Thangavel[1], Qiang Shen [2], A. Pethalakshmi[3*]

[1] *Department of Mathematics, Gandhigram Rural Institute-Deemed University,* Gandhigram-624 302, Tamil Nadu, India. E-mail: drktvelu@yahoo.com.
[2] *Department of Computer Science, University of Wales,* Aberysiwyth, Ceredigion, SY23 3DB, Wales, U.K, E-mail:qqs@aber.ac.uk
[3] *Department of Computer Science, M.V.M Government Arts College(W),* Dindigul-624 001, Tamil Nadu, India. E-mail: Pethalakshmi@yahoo.com

Abstract

Unsupervised clustering is an essential technique in Datamining. Since feature selection is a valuable technique in data analysis for information preserving data reduction, researchers have made use of the rough set theory to construct reducts by which the unsupervised clustering is changed into the supervised reduct. Rule identification involves the application of Datamining techniques to derive usage patterns from the information system. Knowledge extraction from data is the key to success in many fields. Knowledge extraction techniques and tools can assist humans in analyzing mountains of data and to turn the information contained in the data into successful decision making. This paper proposes, to consider an information system without any decision attribute. The proposal is useful when we get data, which contains only input information (condition attributes) but without decision (class attribute). K-Means algorithm is applied to cluster the given information system for different values of K. Decision table could be formulated using this clustered data as the decision variable. Then Quickreduct and VPRS algorithms are applied for selecting features. Ultimately, Rule Algorithm is used for obtaining optimum rules. The experiments are carried out on data sets of UCI machine learning repository and the HIV data set to analyze the performance study.

Keywords: *Datamining, K-Means Clustering, Rough set, Reduct and Rule induction.*

1 Introduction

1.1 Datamining

Datamining refers to extracting or "mining" knowledge from large amounts of data. There are many other terms carrying a similar or slightly different meaning to Datamining, such as knowledge mining from databases, knowledge extraction, data pattern analysis, data archaeology, and data dredging. Datamining treats as synonym for another popularly used term, Knowledge Discovery in Databases (KDD) [9]. KDD consists of the following steps to process it such as Data cleaning, Data integration, Data selection, Data transformation, Datamining, Pattern evaluation and Knowledge presentation.

KDD is the nontrivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data. Datamining is not a single technique, some commonly used techniques are: Statistical Methods, Case-Based Reasoning (CBR), Neural Networks, Decision Trees, Rule Induction, Bayesian Belief Networks (BBN), Genetic Algorithms, Fuzzy Sets and Rough Sets. Datamining relates to other areas, including machine learning, cluster analysis, regression analysis, and neural networks. Both neural network and regression approaches create the same model based on a training data set. This model normally uses a predetermined set of features. A machine learning algorithm of data mining generates a number of models (usually in the form of decision rules) capturing relationships between the input features and the decisions. In an extreme case, the set

of features included in each rule could be independent from all other rules, which is similar to the result produced by cluster analysis. Neural network and regression models can be viewed as "population based" as a single model is formed for the entire population (training data set), while the datamining approach follows an "individual (data object) based" paradigm. The "population based" tools determine features that are common to a population (training data set). The models (rules) created by datamining are explicit.

One of the new datamining theories is the rough set theory[28] that can be used for

(i) reduction of data sets
(ii) finding hidden data patterns
(iii) generation of decision rules

M. Goebel et.al [8] provided an overview of common knowledge discovery tasks and approaches to solve these tasks. S. Shah et.al [35] applied Datamining to predict the Survival of Kidney Dialysis Patients. In this paper we consider an information system without decision attribute which can be found in many real life cases. A novel clustering technique is applied to cluster the information system in order to get class attribute. The overview of the clustering technique is elaborated here under.

1.2 Clustering Overview

Clustering [9] is the process of grouping the data into classes or clusters so that objects within a cluster have high similarity in comparison to one another, but are very dissimilar to objects in other clusters. Clustering has its roots in many areas, including Datamining, statistics, biology, and machine learning. Clustering is a type of multivariate statistical analysis also known as cluster analysis, unsupervised classification analysis, or numerical taxonomy. Cluster analysis is based on a mathematical formulation of a measure of similarity. In other words cluster analysis is an exploratory data analysis tool which aims at sorting different objects into groups in a way that the degree of association between two objects is maximal if they belong to the same group and minimal otherwise. In other words, cluster analysis simply discovers structures in data without explaining why they exist.

Clustering is an excellent example of an unsupervised learning technique [6,7] and we cannot observe the (real) number of clusters in the data. However, it is reasonable to replace the usual notion (applicable to supervised learning) of "accuracy" with that of "distance." In general, we can apply the v-fold cross-validation method to a range of numbers of clusters in K-Means clustering, and observe the resulting average distance of the observations (in the cross-validation or testing samples) from their cluster centers (for K-Means clustering). L. Shen et al. [36] described the stock prediction for use of investors. In his paper, the original decision table together with a new decision attribute obtained by Self-Organizing Maps (SOM) is reconstructed. The SOM is applied as a cluster method. Questier et al. [32] described the uses of rough set theory to construct the reducts in a supervised way for reducing the number of features in an unsupervised clustering. S. Susanto et.al [39] developed a new approach for the allocation of the students using fuzzy clustering algorithm. T. Maciag et al. [24,25] applied K-Means Clustering and Rough Set Exploration System (RSES) for feature selection and decision making.

1.3 Rough Set Based Feature Reduction

In 1982, Pawlak introduced the theory of Rough sets [28,29]. This theory was initially developed for a finite universe of discourse in which the knowledge base is a partition, which is obtained by any equivalence relation defined on the universe of discourse. In rough sets theory, the data is organized in a table called decision table. Rows of the decision table correspond to objects, and columns correspond to attributes. In the data set, a class label to indicate the class to which each row belongs. The class label is called as decision attribute, the rest of the attributes are the condition attributes. Here, C is used to denote the condition attributes, D for decision attributes, where $C \cap D = \Phi$, and t_j denotes the j^{th} tuple of the data table. Rough sets theory defines three regions based on the equivalent classes induced by the attribute values: lower approximation, upper approximation, and boundary. Lower approximation contains all the objects, which are classified surely based on the data collected, and Upper approximation contains all the objects, which can be classified probably, while the boundary is the difference between the upper approximation and the lower approximation. Hu et al., [10] presented the formal definitions of rough set theory. A. Kusiak [19] described the basic concepts of rough set theory, and other aspects of Datamining. The other aspects of datamining are Equivalence classes, Atoms, Approximation accuracy, Boundary approximation, Classification accuracy, Classification quality, Sensitivity, Specificity, Positive predicted value, Negative predicted value, Rule length, Rule strength, Exact rule, Approximate rule, Rule support, Rule coverage, Rule acceptance and Discrimination level.

Let U be any finite universe of discourse. Let R be any equivalence relation defined on U, which partitions

U. Here, (U, R) which is the collection of all equivalence classes, is called the approximation space. Let W_1, W_2, W_3, ..., W_n be the elements of the approximation space (U, R). This collection is known as knowledge base. Then for any subset A of U, the lower and upper approximations are defined as follows:

$$\underline{R}A = \cup\{W_i / W_i \subseteq A\}$$

$$\overline{R}A = \cup \{W_i / W_i \cap A \neq \varnothing\}$$

The ordered pair $(\underline{R}A, \overline{R}A)$ is called a rough set. Once defined these approximations of A, the reference universe U is divided into three different regions: the positive region $POS_R(A)$, the negative region $NEG_R(A)$ and the boundary region $BND_R(A)$, defined as follows:

$$POS_R(A) = \underline{R}A$$
$$NEG_R(A) = U - \overline{R}A$$
$$BND_R(A) = \overline{R}A - \underline{R}A$$

Hence, it is trivial that if $BND(A) = \Phi$, then A is exact. This approach provides a mathematical tool that can be used to find out all possible reducts. However, this process is NP-hard [17, 45], if the number of elements of the universe of discourse is large. As there is a one-to-one correspondence between the knowledge base and knowledge representation, the theory can be adopted for the decision tables in information systems.

Feature selection process refers to choose a subset of attributes from the set of original attributes. Feature selection has been studied intensively for the past one decade [15,16,21,26]. Besides that the brief introduction given here, the extensive literature of Rough sets theory can be referred to Orlowska [27], Peters and Skowron [30], Polkowski et al. [31] for recent comprehensive and overviews of developments.

The purpose of the feature selection is to identify the significant features, eliminate the irrelevant of dispensable features to the learning task, and build a good learning model such as web categorization discussed in [13]. The benefits of feature selection are twofold: it considerably decreased the computation time of the induction algorithm and increased the accuracy of the resulting mode.

All feature selection algorithms fall into two categories: (1) the filter approach and (2) the wrapper approach. In the filter approach, the feature selection is performed as a preprocessing step to induction. The filter approach is ineffective in dealing with the feature redundancy. Some of the algorithms in the Filter approach methods are Relief, Focus, Las Vegas Filter (LVF), Selection Construction Ranking using Attribute Pattern (SCRAP), Entropy-Based Reduction (EBR), Fractal Dimension Reduction (FDR). In Relief [16] each feature is given a relevance weighting that reflects its ability to discern between decision class labels. Focus [1], conducts a breadth-first search of all feature subsets to determine the minimal set of features that can provide a consistent labeling of the training data. LVF employs an alternative generation procedure – that of choosing random features subsets, accomplished by the use of a Las Vegas algorithm [22]. SCRAP [33] is an instance based filter, which determines feature relevance by performing a sequential search within the instance space. EBR [11] based on the entropy heuristic employed by machine learning techniques such as C4.5. EBR is concerned with examining a dataset and determining those attributes that provide the most gain in information. FDR [43] is a novel approach to feature selection based on the concept of fractals – the self-similarity exhibited by data on different scales.

In the wrapper approach [15], the feature selection is "wrapped around" an induction algorithm, so that the bias of the operators that defined the search and that of the induction algorithm interact mutually. Though the wrapper approach suffers less from feature interaction, nonetheless, its running time would make the wrapper approach infeasible in practice, especially if there are many features, because the wrapper approach keeps running the induction algorithms on different subsets from the entire attributes set until a desirable subset is identified. We intend to keep the algorithm bias as small as possible and would like to find a subset of attributes that can generate good results by applying a suite of datamining algorithms. Some of the Wrapper approach methods are Las Vegas Wrapper (LVW) and Neural network-based feature selection. The LVW algorithm [23] is a wrapper method based on LVF algorithm. This again uses a Las Vegas style of random subset creation which guarantees that given enough time, the optimal solution will be found. Neural network-based feature selection [34] is employed for backward elimination in the search for optimal subsets.

A decision table may have more than one reduct. Anyone of them can be used to replace the original table. Finding all the reducts from a decision table is NP-Hard [20]. Fortunately, in many real applications it is usually not necessary to find all of them and it is enough to compute one such reduct is sufficient [10]. A natural question is which reduct is the best if there exist more than one reduct. The selection depends on the optimality criterion associated with the attributes. If it

is possible to assign a cost function to attributes, then the selection can be naturally based on the combined minimum cost criteria. In the absence of an attribute cost function, the only source of information to select the reduct is the contents of the data table [26]. For simplicity, we adopt the criteria that the best reduct is the one with the minimal number of attributes and that if there are two or more reducts with same number of attributes, then the reduct with the least number of combinations of values of its attributes is selected. N. Zhong and A. Skowron [46] have applied Rough Sets with Heuristics (RSH) and Rough Sets with Boolean Reasoning (RSBR) for attribute selection and discretization of real-valued attributes.

In Section 2 the Datamining techniques are studied and implemented using MATLAB for the various data sets obtained from UCI machine learning repository [5] and the real HIV data set. Section 3 describes the experimental analysis of Quickreduct and VPRS, Section 4 states the conclusion of this paper and the directions for further research are proposed herein.

2 Datamining Process

The block diagram of the datamining methodology is depicted in the following figure.

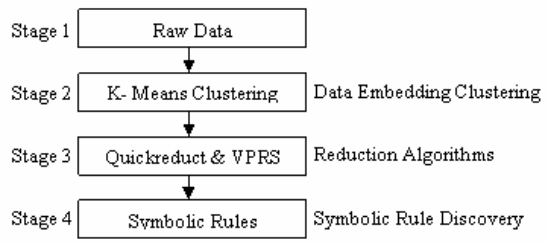

Figure 1. Datamining Process

2.1 Data Preparation

In the first stage, the data sets viz., Iris, Zoo, and Soybean (small) obtained from UCI machine learning repository [5] and the real HIV data set are considered for this study and it is tabulated in the Table 6. The HIV database consists of information collected from the HIV Patients at Voluntary Counseling and Testing Centre (VCTC) of Government Hospital, Dindigul District, Tamilnadu, India, a well-known centre for diagnosis and treatment of HIV. The advantage of this data set is that it includes a sufficient number of records of different categories of people affected by HIV. The set of descriptors presents all the required information about patients. It contains the records of 500 patients. The record of every patient contains 49 attributes and this has been reduced to 22 attributes after consulting the Physician. The details of attributes are given as follows: The continuous attributes are Age, Sex, Marital-Status, Occupation, Area, Loss-of-Weight, Continuous-Fever, Continuous-Cough, Skin-Disease, Oral-Thrush, Tuberculosis, Diarrahoea, Anaemia, Sexual–Transmission-Disease, Swelling-on-Neck, Different-Count, Total-Count, Erythrocyte-Rate, Creatinine, Loss-of-Appetite, Lymphodenopathy and the decision attribute Result (Positive, Negative, Suspect). In this study, decision attributes are omitted to analyze the proposed methodology, since there are possibilities to obtain the information system without decision attribute in the real life cases.

2.2 The K-Means Clustering Algorithm

In stage 2, the data set without decision attribute obtained from stage 1 is partitioned into K clusters, where each cluster comprises data-vectors with similar inherent characteristics. The overall outcome of this stage is the availability of K-number of data clusters, which forms the basis for subsequent discovery of symbolic rules that define the structure of the discovered clusters.

The K-Means Algorithm Process

- The dataset is partitioned into K clusters and the data points are randomly assigned to the clusters resulting in clusters that have roughly the same number of data points.
- For each data point, calculate the distance from the data point to each cluster.
- If the data point is closest to its own cluster, leave it where it is. If the data point is not closest to its own cluster, move it into the closest cluster.
- Repeat the above step until a complete pass through all the data points results in no data point moving from one cluster to another. At this point the clusters are stable and the clustering process ends.
- The choice of initial partition can greatly affect the final clusters that result, in terms of inter-cluster and intracluster distances and cohesion.

The cluster labels obtained by using K-Means algorithm are used to label the objects to get a decision system and then the reduct algorithms can be applied to reduce the condition attributes.

2.3 Reduct Algorithms

The third stage applies the following two different types of reduction algorithms for finding the smallest set of attributes after reconstructing the decision table.

2.3.1 Quickreduct algorithm

The reduction of attributes is achieved by comparing equivalence relations generated by sets of attributes. Attributes are removed so that the reduced set provides the same predictive capability of the decision feature as the original. A reduct is defined as a subset of minimal cardinality R_{min} of the conditional attribute set C such that $\gamma_R(D) = \gamma_C(D)$.

$$R = \{X : X \subseteq C; \gamma_X(D) = \gamma_C(D)\}$$
$$R_{min} = \{X : X \in R; \forall Y \in R; |X| \leq |Y|\}$$

The intersection of all the sets in R_{min} is called the core, the elements of which are those attributes that cannot be eliminated without introducing more contradictions to the dataset. In this method a subset with minimum cardinality is searched for.

The problem of finding a reduct of an information system has been the subject of much research in [2,40]. R. Jenson and Q. Shen [12,13,14] have developed the Quickreduct algorithm to compute a minimal reduct without exhaustively generating all possible subsets and also they developed Fuzzy-Rough attribute reduction with application to web categorization. K. Thangavel et al. [41, 42] applied Rough Sets for feature selection in Medical databases like Mammograms, HIV etc. Q. Shen and A. Chouchoulas [37] developed a potentially powerful fuzzy-rule induction algorithm with a rough set-assisted feature reduction method. They were also developed a modular approach to generating fuzzy rules with reduced attributes for the monitoring system of urban treatment plant [38].

The most basic solution to locating such a subset is to simply generate all possible subsets and retrieve those with a maximum rough set dependency degree. Obviously, this is an expensive solution to the problem and is only practical for very simple datasets. Most of the time only one reduct is required as, typically, only one subset of features is used to reduce a dataset, so all the calculations involved in discovering the rest are pointless.

To improve the performance of the above method, an element of pruning can be introduced. By noting the cardinality of any pre-discovered reducts, the current possible subset can be ignored if it contains more elements. However, a better approach is needed - one that will avoid wasted computational effort. The pseudo code of the Quickreduct is given below:

QUICKREDUCT(C,D)
C, the set of all conditional features;
D, the set of decision features.
(a) $R \leftarrow \{\}$
(b) Do
(c) $T \leftarrow R$
(d) $\forall x \in (C-R)$
(e) if $\gamma_{R \cup \{x\}}(D) > \gamma_T(D)$
where $\gamma_R(D)=card(POS_R(D)) / card(U)$
(f) $T \leftarrow R \cup \{x\}$
(g) $R \leftarrow T$
(h) until $\gamma_R(D) = = \gamma_C(D)$
(i) return R

2.3.2 Variable Precision Rough Set (VPRS)

Variable precision rough sets (VPRS) [47] extend rough set theory by the relaxation of the subset operator. It was proposed to analyze and identify data patterns which represent statistical trends rather than functional. The VPRS approach may also be found in [3,4,18]. As yet, there have been no comparative experimental studies between rough set methods and the VPRS method. The main idea of VPRS is to allow objects to be classified with an error smaller than a certain predefined level. This introduced threshold relaxes the rough set notion of requiring no information outside the dataset itself. Let $X,Y \subseteq U$, the relative classification error is defined by

$$c(X,Y)=1-\{|X \cap Y|/|X|\}$$

Observe that $c(X,Y) = 0$ if and only if $X \subseteq Y$. A degree of inclusion can be achieved by allowing a certain level of error, , β in classification:

$$X \subseteq_\beta Y \text{ iff } c(X,Y) \leq \beta, 0 \leq \beta < 0.5$$

Using $\subseteq_\beta$ instead of $\subseteq$, the β-upper and β-lower approximations of a set X can be defined as:

$$\underline{R}_\beta X = \cup \{ [x]_R \in U/R \mid [x] \subseteq_\beta X\}$$
$$\overline{R}_\beta X = \cup \{ [x]_R \in U/R \mid c([x]_R, X) < 1-\beta \} \quad (3)$$

Note that $\underline{R}_\beta X = \underline{R}X$ for $\beta=0$. The positive, negative and boundary regions in the original rough set theory can now be extended to:

$$POS_{R,\beta}(X) = \underline{R}_\beta X$$

$$NEG_{R,\beta}(X) = \overline{U} - R_\beta X$$
$$BND_{R,\beta}(X) = \overline{R_\beta X} - \underline{R_\beta X}$$

Consider a decision table A = (U, C ∪ D), where C is the set of conditional attributes and D the set of decision attributes. The β-positive region of an equivalence relation Q on U may be determined by

$$POS_{R,\beta}(Q) = \cup X \in U / Q\underline{R_\beta}X \quad (8)$$

where R is also an equivalence relation on U. This can then be used to calculate dependencies and thus determine β-reducts. The dependency function becomes:

$$\gamma_{R,\beta}(Q) = | POS_{R,\beta}(Q) | / |U| \quad (9)$$

It can be seen that the QUICKREDUCT algorithm outlined previously can be adapted to incorporate the reduction method built upon the VPRS theory. By supplying a suitable β-value to the algorithm, the β-lower approximation, β-positive region, and β-dependency can replace the traditional calculations. This will result in a more approximate final reduct, which may be a better generalization when encountering unseen data. However, the variable precision approach requires the additional parameter β which has to be specified from the start. By repeated experimentation, this parameter can be suitably approximated. However, problems arise when searching for true reducts as VPRS incorporates an element of inaccuracy in determining the number of classifiable objects.

2.4 Rule Extraction

In this stage, reduced data obtained from stage 3 is applied to the rule extraction algorithm [44] to formulate the efficient rules (Table 6). The rule extraction algorithm uses the following Heuristic Approach:

(i) Merge identical rows that are rows with similar condition and decision attribute values.
(ii) Compute the core of every row.
(iii) Merge duplicate rows and compose a table with reduct value.

2.5 Worked example

A system of 8 data points consisting four condition attributes with no decision attribute is taken into consideration and it is presented in Table 1.

Table 1: Data Set

Object	Weight	Door	Size	Cylinder
1	LOW	2	COM	4
2	LOW	4	SUB	6
3	MEDIUM	4	COM	4
4	HIGH	2	COM	6
5	HIGH	4	COM	4
6	LOW	4	COM	4
7	HIGH	4	SUB	6
8	LOW	2	SUB	6

In Table 1, the following substitutions LOW=1, MEDIUM=2, HIGH=3, COM=1 and SUB=2 can be used. Applying K-Means Clustering algorithm with K=2. The clustered rows are {1, 3, 5, 6} and {2, 4, 7, 8}. Then the above table is reconstructed using the clustered rows as the decision value, presented in Table 2.

Table 2: Data set after K-means Clustering

Object	Weight	Door	Size	Cylinder	Mileage
1	1	2	1	4	1
2	1	4	2	6	2
3	2	4	1	4	1
4	3	2	1	6	2
5	3	4	1	4	1
6	1	4	1	4	1
7	3	4	2	6	2
8	1	2	2	6	2

Applying the Quickreduct algorithm in Table 2, the final reduct attributes {WEIGHT, DOOR, SIZE} is obtained. Hence, Table 2 can be reduced into Table 3 using the attribute reduct {WEIGHT, DOOR, SIZE}.

Table 3: Attribute Reduction

Object	Weight	Door	Size	Mileage
1	1	2	1	1
2	1	4	2	2
3	2	4	1	1
4	3	2	1	2
5	3	4	1	1
6	1	4	1	1
7	3	4	2	2
8	1	2	2	2

2.5.1 Rule Extraction

Merge identical objects of Table 3. In this step, take the condition attributes of {WEIGHT, DOOR, SIZE} as presented in Table 3. If any identical pair occurs, merge it, otherwise compute the core of every object in Table 3 and present it as in Table 4.

Table 4: Core

Object	Weight	Door	Size	Mileage
1	1	*	1	1
2	1	*	2	2
3	*	4	1	1
4	3	*	*	2
5	*	4	1	1
6	1	*	1	1
7	3	*	*	2
8	1	*	2	2

In the next step, merge duplicate objects with same decision value and compose a table with the reduct value. That is, the merged rows are {1, 6},{2, 8}, {3, 5}and{4, 7} as presented in Table 5.

Table 5: Merged Rows

Object	Weight	Door	Size	Mileage
1	1	*	1	1
2	1	*	2	2
3	*	4	1	1
4	3	*	*	2

Table 5 shows the new set of objects which contains the rules of Table 2. Decision rules are often presented as implications and are often called "if….then…" rules. We can express the rules as follows:

(i) If SIZE = 1 THEN MILEAGE = 1
(ii) If SIZE = 2 THEN MILEAGE = 2
(iii) If DOOR = 4 and SIZE = 1 THEN MILEAGE = 1
(iv) If WEIGHT = 3 THEN MILEAGE = 2

3 Experimental Analysis

The K-Means Clustering, Quickreduct, VPRS and Rule extraction algorithm have been implemented using MATLAB for databases available in the UCI data repository and the HIV data directly collected from the 500 HIV patients. The Comparative Analysis of Quickreduct and VPRS is tabulated in Table 6 as given below. It is observed that less number of rules are generated for the reduct set obtained by using VPRS than the reduct set generated by using Quickreduct.

4 Conclusion

In the rule extraction process, almost all the researchers have framed the rules after applying any one of the reduct algorithms based on rough set theory approach or statistical approach. In this paper, the K-means algorithm has been used to cluster the data set into K-clusters. Then applying the Quick and VPRS reduct algorithms to get the best reduct set of attributes, it was found that the VPRS produces the best reduct for the large data set. The VPRS generates the reduct set which consists of 12 and 14 attributes for K = 2 and 3 respectively, whereas Quickreduct generates a reduct set with 15 attributes for K = 2 and K = 3 in the case of HIV data set. It was observed that less number of rules were produced when the VPRS reduct applied for K = 3 compared to Quickreduct. The unsupervised technique was applied for clustering in this work. The proposed work can be improved by introducing the Neural Network in order to train the system and this is the direction for further research work.

References

[1] H. Almuallim and T.G.Dietterich. Learning with many irrelevant features. *9th National Conference on Artificial Intelligence, MIT Press*, 547-552, 1991.

[2] J.J. Alpigini, J.F. Peters, J. Skowronek and N. Zhong (Eds.). Rough Sets and Current Trends in Computing. *Third International Conference, RSCTC, Malvern, PA, USA*, 2002.

[3] M. J. Beynon. An investigation of β-reduct selection within the variable precision rough set model. *Proceedings of Second International Conference on Rough Sets and Current Trends in Computing*, 114-122, 2000.

[4] M.J. Beynon. Reducts within Variable Precision Rough Sets Model: A Further Investigation. *European Journal of Operational Research*, 134(3):592-605, 2001.

[5] C. L. Blake and C. J. Merz. UCI Repository of machine learning databases. *Irvine, University of California*, 1998, http://www.ics.uci.edu/~mlearn/.

[6] P. Dayan. Unsupervised learning. *The MIT*

Table 6: Comparative Analysis

Datasets	No. of Records	Features	Quickreduct				VPRS β =0.4			
			K = 2		K =3		K = 2		K= 3	
			Reduct	Rule	Reduct	Rule	Reduct	Rule	Reduct	Rule
Car	8	4	3	4	3	5	3	2	2	2
Iris	150	4	2	31	3	38	2	24	2	27
Zoo	101	18	6	7	5	6	5	6	5	6
Soybean (small)	47	35	4	6	4	5	4	5	4	5
HIV	500	21	15	16	15	24	12	14	14	22

Encyclopedia of the Cognitive Science, 1999.

[7] J.G. Dy and C.E. Brodley. Feature selection for unsupervised learning. *Journal of Machine Learning Research*, 5, 2004.

[8] M. Goebel and L. Gruenwal. A Survey of Datamining and Knowledge Discovery Software Tools. *SIGKDD Explorations*, 11(1), 1999.

[9] J. Han and M. Kamber. *Datamining: Concepts and Techniques*, Morgan Kaufmann Publishers, 1995.

[10] X. Hu, T.Y. Lin and J. Jianchao. A New Rough Sets Model Based on Database Systems. *Fundamenta Informaticae*, 1-18, 2004.

[11] R.Jensen and Q.Shen. A Rough Set – Aided system for Sorting WWW Book-marks. In N.Zhong et al.(Eds.), *Web Intelligence: Research and Development*, 95-105, 2001.

[12] R. Jensen. *Combining Rough and Fuzzy Sets for Feature Selection.* Ph.D Thesis, School of Informatics, University of Edinburgh, 2005.

[13] R. Jensen and Q. Shen. Fuzzy-rough attribute reduction with application to web categorization. *Fuzzy Sets and Systems*, 141(3): 469-485, 2004.

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

[15] G.H. John, R. Kohavi and K. Pfleger. Irrelevant Features and the Subset Selection Problem. *Proceedings of 11th International Conference on Machine Learning*, 121-129, 1994.

[16] K. Kira and L.A. Rendell. The Feature selection Problem: Traditional Methods and a New Algorithm. *Proceedings of AAAI, MIT Press*, 129-134, 1992.

[17] R.Kohavi. Useful feature subsets and Rough set reducts. *Proceedings of the 3rd International Workshop on Rough Set and Soft Computing*, 310-317, 1994.

[18] M. Kryszkiewicz. Maintenance of reducts in the variable precision rough sets model. *ICS Research Report, 31/94, Warsaw University of Technology*, 1994.

[19] A. Kusiak. Rough Set Theory: A Datamining Tool for Semiconductor Manufacturing. *IEEE Transactions on Electronics Packaging Manufacturing*, 24(1), 2001.

[20] T.Y. Lin, N. Cercone (Eds.). *Rough sets and Datamining: Analysis of Imprecise Data*. Kluwer Academic Publishers, 1997.

[21] H. Liu and H. Motoda (Eds.). *Feature Extraction Construction and Selection: A Datamining Perspective*, Kluwer International Series in Engineering and Computer Science, Kluwer Academic Publishers, 1998.

[22] H.Liu and R.Setiono. A probabilistic approach to feature selection – a filter solution. *Proceedings of the 9th International conference on Industrial and Engineering Applications of AI and ES*, 284-292, 1996.

[23] H.Liu and R.Setiono. Feature selection and classification – a probabilistic wrapper approach. *Proceedings of the 9th International Conference on Industrial and Engineering Applications of AI and ES*, 419-424, 1996.

[24] T.Maciag and D.H.Hepting. Analysis of User Classifiers for Personalization of Environmental Decision Support System Interfaces. *Artificial Neural Networks in Engineering, St.Louis, Missouri*, 2005.

[25] T.Maciag, D.H.Hepting and D.Slezak. Personalizing User Interfaces for Environmental Decision Support Systems. *International Conference on Web Intelligence-Rough Set Workshop, Compiegne, France*, 2005.

[26] M. Modrzejewski. Feature Selection Using Rough Sets Theory. *Proceedings of the 11th International Conference on Machine Learning*, 213-226, 1993.

[27] E. Orlowska. *Incomplete Information:Rough Set Analysis*. Physica-Verlag, Heidelberg, 1998.

[28] Z. Pawlak. Rough Sets. *International Journal of Computer and Information Sciences*, 11(5): 341-356, 1982.

[29] Z. Pawlak. *Rough Sets:Theoritical Aspects and Reasoning about Data*. Kluwer Academic Publishers, 1991.

[30] J. F. Peters and A. Skowron(eds.). Transactions on Rough Sets 1, *Springer-Verlag, Berlin*, 2004.

[31] L. Polkowski, S. Tsumoto and T.Y. Lin,(eds.). Rough Set Methods and Applications: New Developments in Knowledge Discovery in Information Systems. *Physica-Verlag, Heidelberg*, 2000.

[32] F. Questier, A. Rollier, B. Walczak and D.L. Massart. Application of rough set theory to feature selection for unsupervised clustering. *ScienceDirect-Chemometrics and Intelligent Laboratory Systems*, 63(2):155-167, 2002.

[33] B.Raman and T.R.Loerger. Instance-based filter for feature selection. *Journal of Machine Learning Research*, l:1-23, 2002.

[34] R.Setiono and H.Liu. Neural Network feature selector. *IEEE Transactions on Neural Networks*, 8(3):645-662, 1997.

[35] S. Shah, A. Kusiak and B. Dixon. Datamining in predicting Survival of Kidney Dialysis Patients-Invariant object approach. *Proceedings of Photonics West- Bios*, 4949, 2003.

[36] L. Shen and H. Tong Loh. Applying rough sets to market timing decisions. *Decision Support Systems*, 37(4):583-597, 2004.

[37] Q. Shen and A. Chouchoulas. A rough-fuzzy approach for generating classification rules. *Pattern Recognition*, 35:2425-2438, 2002.

[38] Q. Shen and A. Chouchoulas. A modular approach to generating Fuzzy Rules with Reduced Attributes for the Monitoring of Complex Systems. *Engineering Applications of Artificial Intelligence*, 13(3):263-278, 2002.

[39] S. Susanto, I. Suharto and P. Sukapto. Using fuzzy clustering algorithm for allocation of students. *World Transaction of Engineering and Technology Education, UICEE*, 1:245-248, 2002.

[40] R.W. Swiniarski, and A. Skowron. Rough Set Methods in Feature Selection and Recognition. *Pattern Recognition Letters*,24(6):833, 2003.

[41] K. Thangavel and A. Pethalakshmi. Feature Selection for Medical Database Using Rough System. *International Journal on Artificial Intelligence and Machine Learning*,5(4), 2005.

[42] K. Thangavel, M. Karnan and A. Pethalakshmi. Performance Analysis of Rough Reduct Algorithms in Mammograms. *International Journal on Graphics Vision and Image Processing*, 8:13-21, 2005.

[43] C.Traina Jr, A.Traina, L.Wu and C.Faloutsos. Fast Feature selection using the fractal dimension. *Proceedings of the 15th Brazilian Symposium on Databases(SBBD)*, P.158, 2000.

[44] Y. Yang and T. C. Chiam. Rule Discovery Based On Rough Set Theory. *Proceedings of ISIF, TuC4*, 11-16, 2000.

[45] N.Zhong, J.Dong and S.Ohsuga. Using Rough Sets with Heuristics for Feature Selection. *Journal of Intelligent Information systems*, 16:199-214, 2001.

[46] N. Zhong and A. Skowron. A Rough Set-Based Knowledge Discovery Process. *International Journal of Applied Mathematics and Computer Science*, 11(3):603-619, 2001.

[47] W. Ziarko. Variable Precision Rough Set Model. *Journal of Computer and System Sciences*, 46(1):39-59, 1993.

***Corresponding Author**
Name: A.Pethalakshmi
Postal Address: 2/38, "Alagappa Illam", Rajakkapatti, Collectorate Post, Dindigul, TamilNadu, India-624 001. E-mail: pethalakshmi@yahoo.com
Telephone No: 0451-2460285, 0451-2452511 , 0541-2452371, Fax No: 0451-2453071.