Dimensionality Reduction: Rough Set Based Feature Reduction

T.R.JeraldBeno*, M.Karnan**

*Research Scholar, Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu, India.

** Prof and Head, Department of Computer Science and Engineering,
Tamilnadu College of Engineering, Coimbatore Tamilnadu, India.

***Abstract*-** Feature selection refers to the problem of selecting those input features that are most predictive of a given outcome; a problem encountered in many areas such as machine learning, pattern recognition and image processing. In particular, this has found successful application in tasks that involve datasets containing huge numbers of features which would be impossible to process further. Recent examples include cluster analysis and image classification. Rough set theory has been used as such a dataset pre processor with much success, but current methods are inadequate at tending minimal reductions. This paper proposes a new feature selection mechanism based on fuzzy forward and backward reduct. It also presents a new entropy- based modification of the original rough set-based approach. These are applied to the problem finding minimal rough set reducts, and evaluated experimentally.

Index Terms- Knowledge Discovery, Clustering, optimization, fuzzy forward and backward reduct.

I. INTRODUCTION

Data mining refers to an extracting or "mining" knowledge from large amounts of data. There are many other terms carrying a similar or slightly different meaning to Data mining, such as knowledge mining from databases, knowledge extraction, data pattern analysis, data archaeology and data dredging. Data Mining is one of the steps in Knowledge Discovery in Databases (KDD). KDD and is defined as the nontrivial process of identifying valid, novel, potentially useful and ultimately understandable patterns of interest in data. KDD consists of the following steps to process it such as Data cleaning, Data integration, Data selection, Data transformation, Data mining, Pattern evaluation and Knowledge presentation.

A. Data Cleaning

This phase is used to remove noise and inconsistent data. The goal of this phase is to improve the overall quality of any information that may be discovered. Data Integration: This phase combines multiple data sources. Data Selection: A target dataset is selected or created. Several existing datasets may be joined together to obtain an appropriate example set. Data transformation: Where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, for instance. Data mining: An essential process where intelligent methods are applied in order to extract data patterns. Pattern evaluation: To identify the truly interesting patterns representing knowledge based on some interestingness measures. Knowledge presentation: Where visualization and knowledge representation techniques are used to present the mined knowledge to the user [2- 5].

Clustering is a machine learning technique used to place data elements into related groups without advance knowledge of the group definitions. Popular clustering techniques include k-means clustering and Expectation Maximization (EM) clustering. The notion of a "cluster" varies between algorithms and is one of the many decisions to take when choosing the appropriate algorithm for a particular problem. At first the terminology of a cluster seems obvious: a group of data objects. However, the clusters found by different algorithms vary significantly in their properties, and understanding these "cluster models" is key to understanding the differences between the various algorithms.

Typical cluster models include

* Connectivity models: hierarchical clustering builds models based on distance connectivity.
* Centroid models: k-means algorithm represents each cluster by a single mean vector.
* Distribution models: clusters are modeled using statistic distributions, such as multivariate normal distributions used by the Expectation-maximization algorithm.
* Density models: DBSCAN and OPTICS defines clusters as connected dense regions in the data space.
* Subspace models: in Bi clustering, clusters are modeled with both cluster members and relevant attributes.
* Group models: provide the grouping information.
* Graph-based models: a graph theory, i.e., a subset of nodes in agraph such that every two nodes in the subset are connected by an edge can be considered as a prototypical form of cluster. Relaxations of the complete connectivity requirement (a fraction of the edges can be missing) are known as quasi-cliques.

A "clustering" is essentially a set of such clusters, usually containing all objects in the data set. Additionally, it may specify the relationship of the clusters to each other, for example a hierarchy of clusters embedded in each other. Clusterings can be roughly distinguished in:

* Hard clustering: each object belongs to a cluster or not
* Soft clustering or fuzzy clustering: each object belongs to each cluster to a certain degree there are also finer distinctions possible, for example:
* Strict partitioning clustering: here each object belongs to exactly one cluster
* Strict partitioning clustering with outliers: objects can also belong to no cluster, and are considered
* Overlapping clustering :a hard clustering, objects may belong to more than one cluster.

* Hierarchical clustering: objects that belong to a child cluster also belong to the parent cluster
* Subspace clustering: while an overlapping clustering, within a uniquely defined subspace, clusters are not expected to overlap.

Rough Set theory is an effective tool to deal with vagueness and uncertainty information to select the most relevant attributes for a decision system. However, to find the minimum attributes is a NP-hard problem. In this paper, we describe a method to decrease the scale of the problem by filtering core attributes, and then employ the checking tree to test the rest attributes from bottom to top by using peer-to-peer technique [7].

Furthermore, we utilize pruning method to enhance the speed and discard the node when one of its child node superset of certain attribute reduction found before. Experimental results show that our parallel algorithm has the high speed-up ratio while the attribute reductions are distributed in the bottom of the tree. In a peer-to-peer network, our algorithm will amortize the required memory on client computers. Accordingly, this algorithm can be applied to deal with larger data set in a distributed environment.

B. Hierarchical clustering

Connectivity based clustering, also known as "hierarchical clustering", is based on the core idea of objects being more related to nearby objects than to objects farther away. As such, these algorithms connect "objects" to form "clusters" based on their distance. A cluster can be described largely by the maximum distance needed to connect parts of the cluster [21-24]. At different distances, different clusters will form, which can be represented using a dendrogram, which explains where the common name "hierarchical clustering" comes from: these algorithms do not provide a single partitioning of the data set, but instead provide an extensive hierarchy of clusters that merge with each other at certain distances. In a dendrogram, the y-axis marks the distance at which the clusters merge, while the objects are placed along the x-axis such that the clusters don't mix.

While these methods are fairly easy to understand, the results are not always easy to use, as they will not produce a unique partitioning of the data set, but a hierarchy the user still needs to choose appropriate clusters from. The methods are not very robust towards outliers, which will either show up as additional clusters or even cause other clusters to merge (known as "chaining phenomenon", in particular with single-linkage clustering.

C. Centroids-based clustering

In centroid-based clustering, clusters are represented by a central vector, which may not necessarily be a member of the data set. When the number of clusters is fixed to k, k-means clustering "k"-means clustering gives a formal definition as an optimization problem: find the cluster centers and assign the objects to the nearest cluster center, such that the squared distances from the cluster are minimized.

The optimization problem itself is known to be NP-hard, and thus the common approach is to search only for approximate solutions. A particularly well known approximate method is Lloyd's algorithm often actually referred to as ""k-means algorithm"". It does however only find a local optimum, and is commonly run multiple times with different random initializations. Variations of k-means often include such optimizations as choosing the best of multiple runs, but also restricting the centroids to members of the data set k-medoids, choosing medians k-medians clustering, choosing the initial centers less randomly K-means or allowing a fuzzy cluster assignment Fuzzy clustering and Fuzzy c-means.

Most k-means-type algorithms require the Determining the number of clusters in a data set number of clusters to be specified in advance, which is considered to be one of the biggest drawbacks of these algorithms. Furthermore, the algorithms prefer clusters of approximately similar size, as they will always assign an object to the nearest centroid. This often leads to incorrectly cut borders in between of clusters (which is not surprising, as the algorithm optimized cluster centers, not cluster borders).

K-means has a number of interesting theoretical properties. On one hand, it partitions the data space into a structure known as Voronoi diagram. On the other hand, it is conceptually close to nearest neighbor statistical classification and as such popular in machine learning. Third, it can be seen as a variation of model based classification and Lloyd's algorithm as a variation of the Expectation-maximization algorithm for this model.

D. Distribution-based clustering

The clustering model most closely related to statistics is based on Probability distribution and distribution models. Clusters can then easily be defined as objects belonging most likely to the same distribution. A nice property of this approach is that this closely resembles the way artificial data sets are generated: by sampling random objects from a distribution.

While the theoretical foundation of these methods is excellent, they suffer from one key problem known as over fitting, unless constraints are put on the model complexity. A more complex model will usually always be able to explain the data better, which makes choosing the appropriate model complexity inherently difficult.

The most prominent method is known as expectation-maximization algorithm (or short: "EM-clustering"). Here, the data set is usually modeled with a fixed (to avoid over fitting) number of Gaussian distributions that are initialized randomly and whose parameters are iteratively optimized to fit better to the data set. This will converge to a local optimum, so multiple runs may produce different results. In order to obtain a hard clustering, objects are often then assigned to the Gaussian distribution they most likely belong to, for soft clustering this is not necessary [17-20].

Distribution-based clustering is a semantically strong method, as it not only provides you with clusters, but also produces complex models for the clusters that can also capture correlation and dependence of attributes. However, using these algorithms puts an extra burden on the user: to choose appropriate data models to optimize, and for many real data sets, there may be no mathematical model available the algorithm is able to optimize. (e.g. assuming Gaussian distributions is a rather strong assumption on the data).

Clusters are defined as areas of higher density than the remainder of the data set. Objects in these sparse areas - that are required to separate clusters - are usually considered to be noise and border points.

In contrast to many newer methods, it features a well-defined cluster model called "density-reach ability". Similar to linkage based clustering; it is based on connecting points within certain distance thresholds. However, it only connects points that satisfy

a density criterion, in the original variant defined as a minimum number of other objects within this radius. A cluster consists of all density-connected objects (which can form a cluster of an arbitrary shape, in contrast to many other methods) plus all objects that are within these objects' range. Another interesting property of DBSCAN is that its complexity is fairly low - it requires a linear number of range queries on the database - and that it will discover essentially the same results (it is deterministic algorithm and deterministic for core and noise points, but not for border points) in each run, therefore there is no need to run it multiple times.

The key drawback of DBSCAN and OPTICS is that they expect some kind of density drop to detect cluster borders. Moreover they can not detect intrinsic cluster structures which are prevalent in the majority of real life data. Efficiently detects such kinds of structures. On data sets with, for example, overlapping Gaussian distributions - a common use case in artificial data - the cluster borders produced by these algorithms will often look arbitrary, because the cluster density decreases continuously. On a data set consisting of mixtures of Gaussians, these algorithms are nearly always outperformed by methods such as EM clustering that are able to precisely model this kind of data [9-12].

II. ROUGH SETS: FOUNDATIONS

The rough set approach to approximation of sets leads to useful forms of granular computing that are part of computational intelligence .The basic idea underlying the rough set approach to information granulation is to discover to what extent a given set of objects approximates another set of objects of interest. Objects are compared by considering their descriptions.

Due to space limitations we provide only a brief explanation of the basic framework of rough set theory, along with some of the key definitions. A more comprehensive review can be found in sources.

Rough sets theory provides a novel approach to knowledge description and to approximation of sets. In rough sets theory, feature values of sample objects are collected in what are known as information tables. Rows of such a table correspond to objects and columns correspond to object features.

Rough Set based Feature Reduction: an Overview

In 1982, Pawlak introduced the theory of Rough Sets [1]. 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 the rough sets theory, the data is collected in a table called the decision table. Rows of the decision table correspond to objects and columns correspond to features.

In the data set, a class label indicates the class to which each row belongs. The class label is called a decision feature and the rest of the features are the condition features. Consider that the data set (condition-features, decision-features) is stored in a relational table with the form Table. C is used to denote the condition features, D for decision features, 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 feature 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. presented the formal definitions for rough sets theory

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

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

$RA = \cup \{Wi / W_i \cap A \neq \varnothing\}$

The ordered ($\underline{R}A$, RA) pair is called a rough set. Once defined these approximations of A, the reference universe U is divided in 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:

$POS_R(A) = \underline{R}A$

$NEG_R(A) = U - RA$

$BND_R(A) = RA - \underline{R}A$

Hence, it is trivial that if BND (A) = Φ, then A is exact. This approach provides a mathematical tool that can be used to find out all possible reduces. Two kinds of features are generally perceived as being unnecessary: features that are irrelevant to the target concept (like the row ID, customer ID) and features that are redundant, given other features. In actual applications, these two kinds of unnecessary features can exist at the same time but the latter redundant features are more difficult to eliminate because of the interactions between them. In order to reduce both kinds of unnecessary features to a minimum, feature selection is used [13-16].

Feature selection is a process to choose a subset of features from the original features. Feature selection has been studied intensively in the past decades. The purpose of the feature selection is to identify the significant features, eliminate the features that are irrelevant or dispensable and build a good learning model. The benefits of feature selection are twofold: it considerably decreases the computation time of the induction algorithm and increases the accuracy of the resulting mode.

The purpose of this study is to develop theoretical background and practical aspects of feature extraction (FE) as means of

(1) Dimensionality reduction, and

(2) representation space improvement, for supervised learning (SL) in knowledge discovery systems.

The focus is on applying metaheuristic techniques, conventional Principal Component Analysis (PCA) and two class-conditional approaches for two targets:

(1) for a base level classifier construction, and

(2) for dynamic integration of the base level classifiers

Theoretical bases are derived from classical studies in data mining, machine learning and pattern recognition.

The different aspects of the experimental study on a number of benchmark and real-world data sets include analyses of

(1) Importance of class information use in the FE process;

(2) Advantages of using either extracted features or both original and extracted features for SL;

(3) Applying FE globally to the whole data and locally within natural clusters;

(4) The effect of sampling reduction on FE for SL; and
(5) the problems of FE techniques selection for SL for a problem at consideration

The main contributions of the abstract can be divided into contribution (1) to current theoretical knowledge and (2) to development of practical suggestion on applying FE for SL.

SOFTWARE: MAT LAB Version 6.5/7.0

III. DIMENSIONALITY REDUCTION

Data sets for analysis may contain hundreds of attributes, many of which may irrelevant to the mining task, or redundant, For example, if the task is to classify customers as to whether or not they are likely to purchase a popular new CD at All Electronics when notified of a sale, attributes such as age or music taste. Although it may be possible for a domain expert to pick out some of the useful attributes, this can be a difficult and time-consuming task, especially when the behavior of the data is not well known. Leaving out relevant attributes or keeping irrelevant attributes may be detrimental, causing confusion for mining algorithm employed.

Dimensionality reduction [8] reduces the data size by removing such attributes from it. Typically, methods of attribute subset selection are applied. The goals of attribute subset selection are applied. The goal of attribute subset selection is to find a minimum set of attributes such that the resulting probability distribution of the data classes is as close as possible to the original distribution obtained using all attributes. Miming on a reduced set of attributes has an additional benefit. It reduces the number of attributes appearing in the discovered patterns, helping to make the patterns easier to understand.

An exhaustive search for the optimal subset of attributes can be prohibitively expensive, especially as d and the number of data classes increase. Therefore, heuristic methods that explore a reduced [6] search space are commonly used for attribute subset selection. These methods are typically greedy in that, while searching through attribute space, they always make what looks to be the best choice at the time. Their strategy is to make a locally optimal choice in the hope that this will lead to a globally optimal solution. Such greedy methods are effective in practice and may come close to estimating an optimal solution.

Basic heuristic methods of attribute subset selection include the following techniques.

Stepwise forward selection: The procedure starts with an empty set of attributes. The best of the original attributes is determined and added to the set. At each subsequent iteration or step, the best of the remaining original attributes is added to the set.

Step wise backward elimination: The procedure starts with the full set of attributes. At each step, it removes the worst attribute remaining in the set.

Combination of forward selection and backward elimination: The stepwise forward selection and backward elimination methods can be combined so that at each step, the procedure selects the best attribute and removes the worst from remaining attributes.

IV. FORWARD, REDUCT AND QUICK REDUCT EXPERIMENT RESULTS

To check and compare the efficiency of this new method a series of experiments was run with different test problems. Specifically car dataset were used. These data sets can be found in the well-known data repository of the University of California, UCI,

The following database were used

1. Breast cancer Database: 1500 instance, 9 attributes and 4 classes. The dataset of the cases we have selected 1-200 first set, 1-500 second set, 1-750 third set, 1-1000 fourth set and 1-1500 finally set of test dataset.
2. Real Dataset: Breast cancer database information collected from the private hospital, Coimbatore District, Tamilnadu, India, 500instance, 9 attributes and 4 classes. The entire database were included the test dataset.

The forward selection algorithm and relative reduct algorithm has been implemented using MATLAB for car databases and breast cancer database. The relative reduct algorithm lower and upper approximations are used to improve the efficiency of the algorithm and the Comparative Analysis of relative reduct algorithm and quick relative reduct algorithm is tabulated in Table 1 and Table 2.

Table 1: Comparison Results (breast cancer data base)

data set		breast cancer dataset 1-200	breast cancer dataset 1-500	breast cancer dataset 1-750	breast cancer dataset 1-1000	breast cancer dataset 1-1500
Instances		150	200	350	450	550
No. of Attributes		9	9	9	9	9
Forward selection	No. of reduct	4	4	4	4	4
	Times in (sec)	49	68	67	59	46
Relative Reduct	No. of reduct	3	1	1	1	1
	Times in (sec)	41	52	57	49	39
Quick Relative reduct	No. of reduct	3	1	1	1	1
	Times in (sec)	33	23	19	23	24

Table 2: Comparison Results (Cancer data base)

Medical dataset		Breast cancer 1-100	Breast cancer 1-200	Breast cancer 1-300	Breast cancer 1-360
Instances		100	150	250	350
No. of Attributes		8	8	8	8
Forward selection	No. of reduct	2	2	2	2
	Times in (sec)	42.34	60.25	61.26	53.8
Relative Reduct	No. of reduct	5	3	3	1
	Times in (sec)	69.48	91.99	78.56	103

Quick Relative reduct	No. of reduct	5	3	3	1
	Times in (sec)	40.19	35.49	42.3	31.81

V. FUZZY FORWARD REDUCTEXPERIMENT RESULTS

Real Dataset: 360 instance, 8 attributes and 4 classes. The entire database were included the test dataset. The fuzzy forward selection algorithm and relative reduct algorithm has been implemented using MATLAB for breast cancer database. Comparative Analysis of fuzzy forward and relative reduct algorithm is tabulated in Table 3.

Table 3: Comparison Results (Fuzzy)

Medical dataset	Instances	No. of Attributes	Fuzzy forward selection (Number of reduct)	Fuzzy Relative Reduct (Number of reduct)
Breast cancer 1-100	100	8	2	3
Breast cancer 1-200	150	8	3	3
Breast cancer 1-300	250	8	4	5
Breast cancer 1-360	350	8	1	2

VI. FORWARD SELECTIONREDUCT AND QUICK REDUCT PERFORMANCE

The Performance Analysis of the forward selection, relative reducts and the quick relative reduct in medical are getting the following points: The relative reduct algorithm and quick relative reduct algorithm are obtained same number of reducts. When see the time comparison between both the algorithms, the quick relative reduct algorithm take very less time in same number of reducts.

VII. CONCLUSION

This paper has highlighted the shortcomings of conventional hill-climbing approaches to feature selection. These techniques often fail to find minimal data reductions. Some guiding heuristics are better than others for this, but as no perfect heuristic exists there can be no guarantee of optimality. From the experimentation, it appears that the entropy-based measure is a more useful hill-climbing heuristic than the rough set- based one. However, the entropy measure is a more costly operation than that of dependency evaluation which may be an important factor when processing large datasets. Due to the failure of hill-climbing methods and the fact that complete searches are not feasible for even medium-sized datasets, stochastic approaches provide a promising feature selection mechanism. This paper proposed a new technique based on fuzzy-rough set for this purpose. The initial results are promising, but more experimentation and further investigation into its associated parameters is required. Work is being carried out into the application of this to fuzzy-rough set-based feature selection, where the problem is further compounded by the non-monotonicity of the fuzzy-rough , an algorithm for relative reduct, quick relative attribute reduction, Fuzzy forward and relative reduct. It is based on the rough set theory using backward elimination. Illustration of the algorithm processing and experiment results indicate that the algorithm proposed by this paper is effective and efficient. The technique was originally proposed to avoid the calculation of discernibility functions or positive regions, which can be computationally expensive.

REFERENCES

[1] Pawlak, Z. (1991) Rough Sets, Theoretical Aspects of Reasoning about Data, Kluwer Academic, Dordrecht.

[2] Pelleg, D. and Moore, A. (2000) X-means: Extending K-means with Efficient Estimation of the Number of Clusters, Proceedings of the 17th International Conference on Machine Learning, P. Langley, Ed. Morgan Kaufmann Publishers, San Francisco, CA, Pp.727-734.

[3] Peters, G. (2006) Some refinements of rough k-means clustering, Pattern Recognition, Vol. 39, Pp.1481–1491.

[4] Phung, D., Adams, B., Tran, K., Venkatesh, S. and Kumar, M. (2009) High Accuracy Context Recovery using Clustering Mechanisms, In proceedings of the seventh international conference on Pervasive Computing and Communications, PerCom Galveston, USA, Pp. 122-130.

[5] Raiser, S., Lughofer, E., Eitzinger, C. and Smith, J.E. (2010) Impact of object extraction methods on classification performance in surface inspection systems, Special Issue Paper, Machine Vision and Applications, Springer-Verlag, PP. 627~641.

[6] Ray, S. and Turi, R.H. (1999) Determination of number of clusters in K-means clustering and application in colour image segmentation, Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques (ICAPRDT_99).

[7] Schwarz, G. (1978) Estimating the Dimension of a Model, The Annals of Statistics, Vol. 6, No.2, Pp.461-464.

[8] Silver, M. Sakata, T. Su, H.C. Herman, C. Dolins, S.B. and O'Shea, M.J. (2001) Case study: how to apply data mining techniques in a healthcare data warehouse. Journal of Healthcare Information Management, 15(2), 155-164.

[9] Tajunisha, N. and Saravanan, V (2011) A new approach to improve the clustering accuracy using informative genes for unsupervised microarray data sets, International Journal of Advanced Science and Technology, Vol. 27, Pp. 85-94.

[10] Zhang, T., Ramakrishnan, R. and Livny, M. (1996) BIRCH: an efficient data clustering method for very large data bases, Proceedings of the ACM SIGMOD InternationalConference on Management of Data, Montreal, Canada, Vol. 25, No. 2, Pp.103–114.

[11] C. Velayutham and K. Thangavel**"**Improved Rough Set Algorithms for Optimal Attribute Reduct", JOURNAL OF ELECTRONIC SCIENCE AND TECHNOLOGY, VOL. 9, NO. 2, pp 108-117 JUNE 2011

[12] H. Almuallim and T.G. Dietterich, "Learning with ManyIrrelevant Features," Proc. Ninth Nat'l Conf. Artificial Intelligence,pp. 547-552, 1991.

[13] "Rough Sets and Current Trends in Computing," Proc. Third Int'lConf., J.J. Alpigini, J.F. Peters, J. Skowronek, and N. Zhong, eds.,2002.

[14] T. Beaubouef, F.E. Petry, and G. Arora, "Information Measures for Rough and Fuzzy Sets and Application to Uncertainty in Relational Databases," Rough-Fuzzy Hybridization: A New Trend in Decision Making, 1999.

[15] C.L. Blake and C.J. Merz UCI Repository of Machine Learning Databases, University of California at Irvine, 1998.

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

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

[18] R. Jensen, Q. Shen (2004) Semantics-preserving dimensionality reduction: Rough and fuzzy rough based approaches. IEEE Transactions on Knowledge and Data Engineering, vol. 16, pp.1457–1471

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

[20] Z. Pawlak. Rough Sets: Theoretical Aspects and Reasoning about Data. Kluwer Academic Publishers, Dordrecht, 1991.

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

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

[23] H. Liu and H. Motoda. Feature Extraction Construction and Selection: A Data mining Perspective, Kluwer Academic Publishers, 1998.

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

AUTHORS

First Author – T.R.JeraldBeno, Research Scholar, Department of Computer Science and Engineering, ManonmaniamSundaranar University, Tirunelveli, Tamilnadu, India.

Second Author –M.Karnan, Prof and Head, Department of Computer Science and Engineering, Tamilnadu College of Engineering, Coimbatore Tamilnadu, India.