iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** II **Month of publication:** February

DOI: http://doi.org/10.22214/ijraset.2019.2099

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue II, Feb 2019- Available at www.ijraset.com

A Novel Method to Reduce Attributes from Reducts

Remesh K M [1], Latha R Nair [2]
[1]Division of Computer Engineering, School of Engineering, CUSAT, Kochi, India
[2]Division of Computer Engineering, School of Engineering, CUSAT, Kochi, India

Abstract: A major challenge in present day data mining is the high dimension of datasets. Because of the high dimensionality of data, feature selection becomes a necessary pre-processing step to data mining, which removes superfluous features and enhances performance of classification. Thus, attribute reduction becomes an important pre-processing step for data mining. The concepts of rough set theory have been successfully used for attribute reduction. Reducts in rough set theory provide a minimal set of attributes preserving the knowledge. Removal of an attribute from a reduct creates information loss. The severity of information loss due to removal of an attribute depends on the significance of that attribute. In many cases, the variation in classification accuracy when an attribute that has negligible significance is removed from a reduct may not be significant. Removal of insignificant attributes is useful when dealing with high dimensional data sets. In this paper, a novel method is proposed to identify attributes with negligible significance which can be removed from a reduct. Experimental results for various datasets are presented. The results prove that the accuracy of classification do not vary much after the removal of the insignificant attributes from the reduct.

Keywords: Rough set, reduct, core, attribute dependency, significance, dimensionality reduction

I. INTRODUCTION

Present day data mining is highly challenging as it involves mining of high dimensional data. Mining of high dimensional datasets provides many mathematical challenges thereby becoming complex and time consuming. Moreover, most of the machine learning and data mining techniques are not efficient in handling high dimensional data. The attribute reduces the dimensionality of the datasets which in turn speeds up the mining process. High dimensional datasets may have many measured variables having useful information relevant to the problem. At the same time, it may contain many irrelevant attributes and less important or even unwanted features which are redundant to represent the essential knowledge content in the information system. Thus, the intrinsic dimension of the dataset may be very small[1][13]. It is of interest for many researchers to reduce the dimension of the original dataset.

There are many techniques that are popularly used for dimensionality reduction. Two of the widely used techniques are[15]

1) *Feature Selection:* where only a subset of the original set of attributes is used for representing the knowledge involved [2][15].
2) *Feature Extraction:* where the original high dimensional data is mapped into a space with lower dimension [15].

Rough Set Theory (RST) is a mathematical approach or tool that can be effectively used in data analysis especially to deal with incomplete and uncertain data [1]. Reducts in rough set theory is the minimal set of attributes that hold the complete knowledge of the dataset [1][15]. Significance of attributes can be determined using RST [15]. One of the major advantages of RST is in association rule generation. RST generates the minimal set of association rules. Association rules generated using the rough set approach is much lesser in number as compared to the conventional methods.

Rough set theory proposed by Z. Pawlak during the early 90's can be effectively used for dimensionality reduction especially for attribute reduction of datasets. Attribute reduction in rough set theory works by selecting only the information rich attributes from a dataset. With the reduced set of attributes, the whole knowledge of the dataset can be represented without any loss of information that is required for classification. The rough set approach is very simple and highly efficient for feature selection as compared to the conventional methods such as Principle Component Analysis (PCA).

The paper is organized as follows: Section I contains a general introduction about data mining and the advantages of reducing attributes from datasets. Section II introduces rough set theory and its basic concepts such as indiscernibility, recuct, core, attribute dependency and attribute significances. Section III narrates the recent studies in feature selection using rough set theory. Section IV introduces a novel method to reduce attributes from a reduct and an algorithm for the same. Section V discusses experimental results. It gives a comparison of classification accuracies obtained for different datasets for the reducts and the reduced datasets.

II. PRELIMINARY CONCEPTS OF ROUGH SET THEORY

Rough Sets is an effective mathematical tool to deal with imprecise or vague information contained in data sets [1]. The basic concept of rough set theory lies with the indiscernibility relation [4]. The objects characterized by the same information are called indiscernible and the set of all objects which are indiscernible is called the elementary set. The equivalence relation *IND(P)* is defined as: $IND(P) = \{(x,y) \in U^2 \mid \forall a \in P, a(x) = a(y)\}$ (1)

The target set $X \subseteq P$ can be approximated using the information in P by constructing the lower and upper approximations of X as:

$$\underline{P}X = \{ x \mid [x]_P \subseteq X \} \quad (2)$$

$$\overline{P}X = \{ x \mid [x]_P \cap X \neq \phi \} \quad (3)$$

The tuple ($\underline{P}X$, $\overline{P}X$) composed of the lower and upper approximations is called the rough set.

A. Reduct and Core

All the measured variables do not contribute to knowledge content in an information system. By removing unwanted or unimportant superfluous attributes from a dataset, a subset of attributes is obtained which can fully characterize the knowledge contained in the information system. Such a subset of minimal attributes that fully characterize the whole information involved in the original information system is called a reduct [15]. In other words, the attributes of a reduct are sufficient to represent the knowledge contained in the dataset. The equivalence class structure obtained using the attributes of a reduct and the attributes of the original dataset will be the same [15]. There may be more than one reduct in an information system.

The reducts of an information system may contain a set of common attributes known as core [15].

B. Attribute Dependency and Significance

The ability of rough set theory to evaluate the attribute dependency among a set of attributes, and the significance of attributes to determine the decision attributes makes the theory very beneficial. The attribute set 'Q' depends on the attribute set 'P' of a given information system in a degree k is :$k = \gamma_P(Q) = | POS_P(Q) | / | U | \quad 0 <= k <= 1$ (4)

Degree of dependence may be used to find the reducts of an information system. It can be concluded that the subset is a reduct if the degree of dependence of the subset of attributes is the same as the degree of dependence of the whole set of attributes.

The significance of an attribute $a \in P$ to determine Q is defined as:$Sig_P(Q, a) = \gamma_P(Q) - \gamma_{P-\{a\}}(Q)$ (5)

Where P is the set of conditional attributes and Q is the decision attribute.

III.RELATED STUDY

Bing Zhou and Yiyu Yao in their study [2] strongly opine that confirmation-theoretic rough set models give better results for feature selection. T Sridevi and Murugan in their study [16] proposed a new method for attribute reduction in breast cancer diagnosis. The result underlines better attribute reduction when the rough set based algorithms are used. Using this algorithm, number of attributes could be reduced from 10 to 7 in the dataset used. Shraddha Sarode and Jayant Gadge in their study[17] of attribute reduction, used a rough set based quick reduct algorithm for web page classification. They had applied Naïve Bayesian classifier along with dimensionality reduction for classifying the documents. Wojciech Ziarko in his study [11] focused on discovering the dependencies among attributes and according to this study, a number of alternative reducts were computed and the one with the lowest total cost was selected; it will be more regular as it reflects stronger data patterns. M. Zhang and J. T. Yao in the study "a rough set approach to feature selection" [18], uses two heuristics - Average Support Heuristic (ASH) and Parameterized Average Support Heuristic (PASH) for feature selection. As PASH takes the overall quality of potential rules, it produces a balanced supported distribution over all decision classes. Huilian Fan, Yuanchang Zhong in their study "A Rough Set Approach to Feature Selection Based on Wasp Swarm Optimization" [19], rough set based wasp swarm optimization algorithm for attribute reduction has been proposed. This algorithm obtains the feature space for minimum attributes by using the significance of the feature as probability information. This algorithm gives better results compared to other intelligent swarm algorithms for attribute reduction. Aijun An1, Yanhui Huang, Xiangji Huang, and Nick Cercone [20] in their study present a feature selection method using rough sets for web page classification. Experiments based on this study prove that the approach improves the classification accuracy. Yan Wang and Lizhuang Ma [21] propose a feature forest algorithm for feature selection and found that this method improves classification accuracy. Caballero Yailé, etal [23] in their paper propose two algorithms RS Red and MRS Reduct based on rough sets to determine good reducts in reasonable time. It is seen that better performance can be obtained for feature selection using rough sets. Removing insignificant attributes from reducts is a challenging problem. Deviations in the classification accuracy on datasets on removal of attributes are the concern of this paper.

IV. ALGORITH FOR REDUCT SIZE MINIMIZATION

Reduct is the minimal set of attributes that holds the same knowledge as that of the original data set[15]. Removing attributes from a reduct lead to loss of knowledge in the information system. All the attributes in a reduct may not be highly significant in representing the knowledge in the data set. Removing an attribute that has negligible significance can speed up dataset processing and reduce its storage requirement [2]. It can improve the overall mining performance of high dimensional data. However, the classification accuracy may vary from that of the original dataset. A reduct being a minimal set of attributes that preserves the whole knowledge of the data set, the attributes which have considerable significance cannot be removed. Removal of attributes that have considerable significance causes loss of substantial knowledge in the information system. Identifying attributes from a reduct that has negligible effect in classification is a difficult task.

In this paper a novel method is proposed that determines a threshold such that attributes whose significance falling below the threshold can be removed. The threshold is selected with a view that any attribute whose significance falling below the threshold has negligible role in the classification process. Such attributes, if any, can be removed from the reduct of the data set. Better performance in classification can be obtained by removing the attribute having negligible significance.

This approach makes use of the population mean and standard deviation of the significances of the attributes of a reduct. The degree of variation from the average value (mean) of a population can be identified by its standard deviation. Here, the significances of attributes are considered as the population. It characterizes a normal distribution as it distributes almost symmetrically around the mean of significances of attributes. The “empirical rule” of probability distribution says that 68 % of the population are in the range, [$\boldsymbol{\mu} - \boldsymbol{\sigma}$, $\boldsymbol{\mu} + \boldsymbol{\sigma}$], where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are the population mean and standard deviation respectively. Values above the higher range, [$\boldsymbol{\mu} + \boldsymbol{\sigma}$] , have the greater significance and are the most important attributes in the data set. But values below the lower range, [$\boldsymbol{\mu} - \boldsymbol{\sigma}$], have minor significance and can be discarded. Thus for a normal distribution, only a few objects from the data sets are discarded in the knowledge extraction process. Obviously, as we take the population mean and standard deviation of significances of attributes as basis of distribution, the distribution would be normal.

A. Preprocessing

There can be outliers present among the set of attributes. Such outliers are to be kept away for calculating the population mean and standard deviation. The method to remove outliers from a reduct has the following steps [13]:

1) Calculate mean and standard deviation of significances of attributes.
2) Remove all the attributes whose significances lie outside the range (mean – standard deviation, Mean + standard deviation)
3) Recalculate mean and standard deviation of the subset of attributes obtained and apply this to determine the lower threshold.

The algorithm used to determine the lower threshold for the elimination of attributes from a reduct is given below:

Algorithm Reduct sizer eduction

a) *Input:* Decision Table T
b) *Output:* A reduced set of attributes

i) Determine reducts
ii) Select any reduct (preferably with minimum attributes)
iii) Determine the significance of all the conditional attributes
iv) Preprocess attribute significances
v) Determine the $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ of significance of the attributes
vi) Calculate the lower threshold of significance as

$$\mathbf{LT} = \text{abs}\,(\ \boldsymbol{\mu} - \boldsymbol{\sigma}\ )$$

vii) Identify the attributes whose significances are less than **LT.** If there are more than one attribute, identify the combination of attributes whose sum of significances is less than **LT**
viii) Remove the attributes from the reduct
ix) Output the reduced set of attributes

The above algorithm selects only the attributes of a reduct whose significance fall above the lower threshold. Attributes falling below the lower threshold have meager significance and negligible effect in classification. These attributes can be removed from the data set reducing the dimension of the information system.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue II, Feb 2019- Available at www.ijraset.com

V. RESULTS AND DISCUSSION

Experiments based on the proposed algorithm were conducted on many popular datasets downloaded from UCI Machine Learning Data Repository, out of which results of five data sets are shown in tables 1, 2 and 3. The classification accuracies of the datasets with reduced attributes were evaluated with different algorithms. The results thus obtained were compared with the classification accuracy of the reducts. A comparative chart of the results evaluated when the algorithms - Naïve Bayes and J48, were applied on the reduced and reducts datasets is plotted. Table 3 shows the classification accuracy of both the reduct and the reduced dataset when these algorithms were used for classification. On analysis of the results, it is observed that the classification accuracies do not differ considerably thereby preserving the knowledge content of the dataset. It is observed in certain case that, removal of attributes from reducts improves classification accuracy. This is observed for the Car and Sick datasets when J48 algorithm was used for classification. Enhancement in classification accuracy was obtained when Naïve Bayes algorithm was used on phishing dataset for classification. In certain cases, the proposed algorithm could not find any attribute that could be removed from the reduct.

The total number of attributes of each original dataset and its reduct are given in Table 1 along with the number of attributes in the reduced dataset. The ROC areas after classification of the reduct and reduced dataset are also shown in table 1. It is observed that in most of the cases, reduct has only less number of attributes than the original dataset. It is also observed that the difference in the ROC areas of the reduct and reduced datasets are almost negligible in most of the cases. There is no difference in the ROC area of reduct and reduced Mushroom dataset.

TABLE 1
ROC Area of Reduct and Reduced set

Sl. No.	Data Set	No. of Attributes	Reduct			Reduced set		
			No. of attributes	*ROC area*		*No. of attributes*	*ROC area*	
				Naïve Bayes	*J48*		*Naïve Bayes*	*J48*
1	Mushroom	22	5	0.999	1	4	0.999	1
2	Car	7	7	0.976	0.976	6	0.974	0.978
3	Lymphography	19	7	0.887	0.818	6	0.892	0.793
4	Sick	30	4	0.925	0.951	3	0.922	0.916
5	Phishing	10	10	0.948	0.96	8	0.949	0.961

From Table 2, it is understood that the number of correctly classified instances when both the algorithms were used on the reduct and reduced datasets also gives analogous results.

TABLE 2
Correctly classified objects

Sl. No.	**Data Set**	**Number of objects**	***Correctly Classified Objects***					
			No. of attributes	***Reduct***		***No. of attributes***	***Reduced set***	
				Naïve Bayes	***J48***		***Naïve Bayes***	***J48***
1	Mushroom	8124	5	8011	8124	4	8031	8100
2	Car	1728	7	1478	1596	6	1475	1611
3	Lymphography	148	7	119	111	6	115	106
4	Sick	3772	4	3493	3727	3	3523	3697
5	Phishing	1353	10	1228	1228	8	1142	1227

TABLE 3
Classification Accuracy

Sl. No.	Data Set	*Classification accuracy of Algorithms*					
		No. of attributes	*Reduct*		*No. of attributes*	*Reduced set*	
			Naïve Bayes	*J48*		*Naïve Bayes*	**J48**
1	Mushroom	5	98.60	100	4	98.85	99.70
2	Car	7	85.53	92.36	6	85.35	93.22
3	Lymphography	7	80.40	75.00	6	77.70	71.62
4	Sick	4	95.73	97.45	3	93.39	98.01
5	Phishing	10	84.10	90.76	8	84.48	90.68

Fig. 1 shows the graphical representation of classification accuracies of the data sets given in table 3. It can be seen that classification accuracies of the datasets do not differ much for the algorithms used for the study.

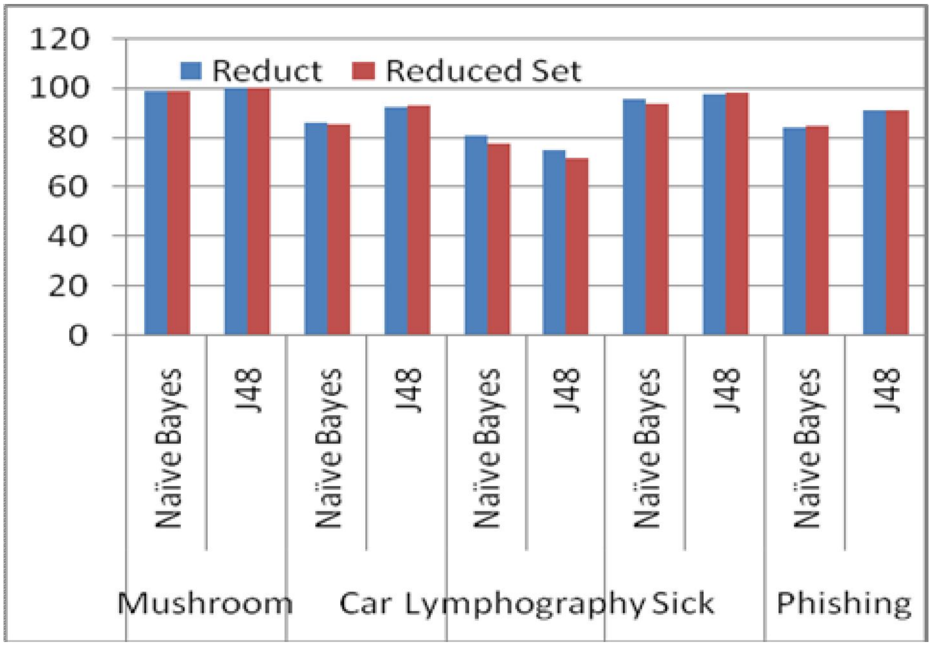

Fig1. Classification Accuracy

VI. CONCLUSION

Reducts generate the minimal set of attributes preserving the knowledge in the information system. Minimizing the number of attributes improves mining performance for high dimensional datasets. A Reduct is a minimal set having the whole knowledge of the information system. Further reduction of attributes from a reducts can be done, if required, by removing attributes with low significance. The attributes having less significance have negligible effect in classification. The proposed algorithm can be used to remove attributes, having negligible significance, from the reducts. Experimental results show only negligible variation in classification accuracy after the removal of attributes with low significance. This method can be effectively used for reducing attributes from a reduct.

REFERENCES

[1] Alex Sandro Aguiar Pessoa, Stephan Stephany, " An Innovative Approach for Attribute Reduction in Rough Set Theory", Intelligent Information Management, Vol. 6, P 223-239, 2014
[2] Bing Zhou, Yiyu Yao, " Feature Selection Based on Confirmation-Theoretic Rough Sets", RSCTC, P 181-188, 2014
[3] Xiuyi Jia, Wenhe Liao, Zehenmin Tang, Lin Shang, "Minimum Cost Attribute Reduction in Decision-Theoretic Rough Set Models", Information Sciences 219 P 151-167, Elsevier Inc. 2013
[4] Mert Bal, " Rough Sets Theory as Symbolic Data Mining Method: An Application on Complete Decision Table", Information Science Letters, P 35-47, 2013
[5] Prasanta Gogoi, Dhruba K Bhattacharyya, Jugal K Kalita, " A Rough Set Based Effective Rule Generation method for Classification with an Application of Intrusion Detection", International Journal of Security and Networks, Vol . 8 , No.2, 2013
[6] Prasanta Gogoi, Ranjan Das, B Borah, D K Bhattacharyya, " Efficient Rule Set Generation using Rough Set Theory for Classification of High Dimensional Data", International Journal of Smart Sensors and Ad Hoc Networks, ISSN No. 2248-9738, Vol. 1, Issue 2, 2011
[7] Arirra Roy, Rajdeep Chatterjee, " Introducing New Hybrid Rough Fuzzy Association Rule Mining Algorithm", International Conference on Recent Trends in Information, Telecommunication and Computing, DOI: 02.ITC.2014.5.101
[8] Pawlak Z , "Rough Sets Theory and its Applications"
[9] Bo Wang, Kenning Gao, Bin Zang, " Algorithm for Feature Selection for Inconsistent Data Preprocessing based Rough Set", International Journal of Information and System Sciences, Vol. 1, P 311-319, 2005
[10] Shrikant Brajesh Sagar, Akhilesh Tiwari, " Rough Set and Genetic based Model for Extracting Weighted Association Rules", International Journal of Hybrid Information Technology, ISSN 1738-9968, Vol. 8, No. 11, pp. 121-138, 2015
[11] Wojciech Ziarko, " Rough Set Approaches for Discovery of Rules and Attribute Dependencies", Handbook of Data Mining and Knowledge Discovery, Oxford University Press, pp. 328-338, 2002
[12] Pawlak Z, " Rough Sets", International Journal of Information and Computer Sciences", Vol. 11, No. 5, pp.341-356, 1982
[13] Sabu MK, Raju G, "Dimensionality Reduction – A Rough Set Approach", International Journal of Machine Intelligence , ISSN 0975-2927 & E-ISSN 0975-9166, Vol 3, Issue 4, 2011
[14] Sabu MK, Raju G, " A Rough Set Based Feature Selection Approach for the prediction of Learning Disabilities", International Journal of Advanced Computational Engineering and Networking, ISSN 2320-2106, Vol 2, Issue 12, 2014
[15] Sabu M K, " Studies on Rough Sets Theory with Applications to Data Mining ", Phd thesis from internet (2014)
[16] T. Sridevi and A. Murugan, "Rough set theory based attribute reduction for breast cancer diagnosis" Indian Journal of Innovations and Developments, vol. 1, no. 5, 2012.
[17] Shraddha Sarode and Jayant Gadge, "Approach for Dimensionality Reduction in Web Page Classification ", International Journal of Computer Applications, vol. 99, no. 14, 2014.
[18] M. Zhang and J. T. Yao, "A rough sets based approach to feature selection", Proceedings of IEEE Annual Meeting of the Fuzzy Information Processing, pp. 27-30, 2004.
[19] Huilian FAN , Yuanchang ZHONG, "A Rough Set Approach to Feature Selection Based on Wasp Swarm Optimization" , Journal of Computational Information Systems, vol. 8, no. 3, 2012.
[20] Aijun An, Yanhui Huang, Xiangji Huang, and Nick Cercone, "Feature Selection with Rough Sets for Web Page Classification", Transactions on Rough Sets II, Springer Berlin Heidelberg, pp. 1-13.
[21] Y Wang and L. Ma, "Feature selection for medical dataset using rough set theory", Proceedings of 3rd WSEAS international conference on Computer engineering and applications, pp. 68-72, 2009.
[22] D. Asir Antony Gnana Singh, E. Jebamalar Leavline, E. Priyanka and C. Sumathi, " Feature Selection Using Rough Set For Improving the Performance of the Supervised Learner" , International Journal of Advanced Science and Technology Vol.87, pp.1-8, 2016
[23] Caballero Yailé, Álvarez Delia, Bello Rafael, García María M, "Feature Selection Algorithms using Rough Set Theory", IEEE Xplore, November 2007, DOI: 10.1109/ISDA.2007.70.