Performance Analysis of Quickreduct, Quick Relative Reduct Algorithm and a New Proposed Algorithm

Ashima Gawar[1]

[1] GGSIPU

Received: 13 December 2013 Accepted: 1 January 2014 Published: 15 January 2014

Abstract

Feature Selection is a process of selecting a subset of relevant features from a huge dataset that satisfy method dependent criteria and thus minimize the cardinality and ensure that the accuracy and precision is not affected ,hence approximating the original class distribution of data from a given set of selected features. Feature selection and feature extraction are the two problems that we face when we want to select the best and important attributes from a given dataset Feature selection is a step in data mining that is done prior to other steps and is found to be very useful and effective in removing unimportant attributes so that the storage efficiency and accuracy of the dataset can be increased. From a huge pool of data available we want to extract useful and relevant information. The problem is not the unavailability of data, it is the quality of data that we lack in. We have Rough Sets Theory which is very useful in extracting relevant attributes and help to increase the importance of the information system we have. Rough set theory works on the principle of classifying similar objects into classes with respect to some features and those features may collectively and shortly be termed as reducts.

Index terms— data mining, rough set, quickreduct, quick relative reduct, feature selection, feature extraction.

1 Introduction a) Feature Selection and Feature Extraction

We have organized the remaining paper as follows : section 2 briefs about the data set used for the study. Section 3 describes the Quickreduct algorithm.

Author: MCA from Institute of Information Technology and Management, Janakpuri, New Delhi. e-mail: ashima_gavar@yahoo.in Section 5 explains the analysis of the comparison made between the Quickreduct and the Quick Relative Reduct algorithm. Section 6 suggests some improvement in the QuickReduct algorithm and finally Section 7 states the conclusion of the paper.

2 b) Reducts

The minimal set of attributes that will identify the other attributes of the dataset thus improving its accuracy and efficiency are called reducts.(Jothi and Inbarani , October 2012) Mathematically , a reduct of an algebraic structure that is calculated by removing some of the operations and relations of the mathematical structure we are using. In a reduct we keep only those attributes that are similar in nature and consequently have the goal of set approximation . Usually we can find several such subsets and those which are minimal among those are called reducts. Given an information table S, an attribute set R _At is called a reduct, if R satisfies the two conditions:1. IND(R) = IND(At); 2. For any a ? R, IND(R -{a}) ? IND(At).

3 c) Rough Sets

Rough set theory provide a novel methodological approach for approximation of large sets and describing the knowledge. In rough set theory firstly we collect a sample object set and store the feature values in information tables. Rough sets help us to find reducts without deteriorating the original quality of the dataset.

Characterization of Rough sets cannot be done in terms of information about the elements of rough sets. With every rough set a pair of precise sets, known as the lower and the upper approximations of the rough set. The lower approximation contains all the objects which definitely belong to the set and the upper approximation contains all objects which may possibly belong to the set. The difference between the upper and the lower approximation constitutes the boundary region of the rough set. Approximations are the fundamental concepts of rough set theory.

Rough set theory can be described as a formal methodology that can be employed to reduce the dimensions of datasets and is used as an preprocessing step to data mining. The reduced Section 4 describes Quick Relative Reduct algorithm.

4 Global Journal of Computer Science and Technology

Volume XIV Issue IV Version I1 (D D D D D D D D) Year 2014 c © 2014 Global Journals Inc. (US)

t to the process of finding out and select minimum subsets of attribute from a large set of original attributes and finally select the minimal one. The aim behind the process is to reduce dimensions across the datasets, remove the attributes which have no significance and identify the most important and useful attributes. (Zhang et al., 2003) It will help in improving and increasing accuracy and lessen the time that the algorithm will take for its computation. I dimensionality improves the runtime performance of an algorithm. Rough Set theory (Suguna and Thanushkodi , 2010) is a mathematical approach that is based on the principle that if the degree of precision in a dataset is lowered then we can more easily visualize the data patterns. The main aim is to approximate the lower and upper bounds. Rough set based data analysis initially analyses the data table called decision table in which the columns are labeled by attributes and rows represent the objects. The entries of the table will contain the value of the attributes . Attributes of the decision table are divided into two disjoint groups which are called decision and condition attributes respectively. Any rough set is associated with a pair of precise sets which are called the lower and upper approximations of the rough set is associated (Yiyu and Yan 2009).

5 II.

6 Data Preparation

We have manually performed analysis on the test datasets . The first dataset contains information about AUTOMOBILE and the second one contains information contains data about COMPUTER.

7 III.

Quickreduct Algorithm (QR) In Quickreduct algorithm we remove the attributes so that the set we get after reduction provides the same prediction of the decision feature as the original set which is achieved by comparing equivalence relations generated by sets of attributes. The attribute selected for the first time is to be included in the reduct set in the Quickreduct algorithm (Velayutham and Thangavel, September 2011) is the degree of dependency of that attribute which is not equal to zero..The algorithm tries to find out a minimal reduct without generating all possible subsets . Initially we take an empty set and add in the empty set R those attributes that will result in the greatest increase in dependency value one by one until we get the maximum possible value for the dataset.

8 Quick Relative Reduct Algorithm

In Quick Relative Reduct (Kalyani and karnan 2011)algorithm we find out the degree of relative dependency after removing the attributes from the set. If a attribute is removed and it causes the value of relative dependency to be one then that attribute is eliminated otherwise it is put in the core reduct. The process is repeated again and again till the value becomes one. The algorithm is explained below : Both Quick Reduct and Quick Relative reduct are reduct algorithms but the Quick Relative Reduct is a more efficient algorithm as it calculates reducts without expensive. It includes a simple approach using relative dependency.

9 Select next x

10 False

11 Proposed Algorithm

We propose a new algorithm to overcome the disadvantage of Quick Relative Reduct that in this algorithm we calculate relative dependency and the attribute is chosen with highest degree of dependency. When the highest relative dependency value is possessed by more than one attribute. For that purpose we can introduce a significance factor associated with every attribute and choose the attribute with greater significance. Significance factor (Jothi and Inbarani 2012) is defined as :

Assume X ? A is an attribute subset, x?A is an attribute, the importance of x for X is denoted by Sig X

12 Conclusion

In this we discussed the comparison analysis of the Quickreduct and the Relative QuickReduct algorithm. The Relative QuickReduct algorithm finds QuickRelative Reduct QuickReduct reducts based on backward elimination of attributes and the QuickReduct algorithm finds reducts based on forward elimination. We also found out that Quick Relative Reduct was better than the QuickReduct algorithm. Also the Relative QuickReduct algorithm can be modified further to improve the efficiency by introducing the concept of significance factor. Further work can be carried out on the defined algorithm to explore its efficiency and accuracy. The analysis was performed manually but the research can be carried out further for further suggestions and improvements.

13 Global Journal of Computer Science and Technology

Volume XIV Issue IV Version I Step 1: Take the R as the set of all conditional attributes.

Step 2 : Now select the conditional attribute.

Step 3 : Calculate the relative dependency of the attribute.

Step 4 : If relative dependency of the attribute is one then eliminate the attribute , Go to step 2.

Step 5: If relative dependency is not equal to one then select the highest dependency attribute ,if two attributes greater significance. [1]

11

Figure 1: Algorithm 1 .Figure 1 :

2

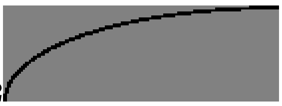

Figure 2: Figure 2 :

1

Year 2014
c

Figure 3: Table 1 :

2

Date set	Attributes	Instances	Selected Attributes	Reduct ?	Optimal ?
Automobile	4	8	3	Yes	Yes
Computer	6	20	3	Yes	Yes

Figure 4: Table 2 :

.1 This page is intentionally left blank

[Slowinski and Vanderpooten ()] 'A generalized definition of rough approximations based on similarity'. R Slowinski , D Vanderpooten . *IEEE Transactions on Knowledge and Data Engineering* 2000. p. .

[Suguna et al. (2010)] 'A Novel Rough Set Reduct Algorithm for Medical Domain Based on Bee Colony Optimization'. N Suguna , Dr , K Thanushkodi . *Journal Of Computing* June 2010. 2 (6) p. .

[Yiyu and Yan ()] 'Discernibility Matrix Simpli_cation for Constructing Attribute Reducts Discernibility matrix simplication for constructing attribute reducts'. Y Yiyu , Z Yan . *Information Sciences* 2009. 179 (5) p. .

[Pawlak ()] 'Rough sets'. Z Pawlak . *International Journal of Computer Information and Science* 1982. 11 p. .

[Yao et al. ()] *Rough Sets and Data Mining: Analysis for Imprecise Data*, Y Y Yao , S K M Wong , T Y Lin . 1997. p. . (A review of rough set models)

[Jothi and Inbarani (2012)] 'Soft Set Based Feature Selection Approach for Lung Cancer Images'. . G Jothi , H Inbarani . *International Journal of Scientific and Engineering Research* October 2012. 3 p. .

[Velayutham (2011)] 'Unsupervised Quick Reduct Algorithm Using Rough Set Theory'. C Velayutham . *Journal Of Electronic Science And Technology* September 2011. 9 (3) p. .

[Chandrasekhar et al. (2012)] 'Verdict Accuracy of Quick Reduct Algorithm using Clustering and Classification Techniques for Gene Expression Data'. T Chandrasekhar , K Thangavel , E N Sathishkumar . *IJCSI International Journal of Computer Science Issues* January 2012. 9 (1) p. .

[Zhang et al. ()] J Zhang , J Wang , H Huacan , JiaguangS . *A New Heuristic Reduct Algorithm Based on Rough Sets Theory , 4th International Conference*, (WAIM , Chengdu, China) August 17-19, 2003. 2003. 2762 p. .