DATASET VERTICAL PARTITIONING FOR ROUGH SET BASED CLASSIFICATION

QASEM A. AL-RADAIDEH

Department of Computer Information Systems
Faculty of Information Technology and Computer Sciences
Yarmouk University Irbid 21163, Jordan
qasemr@yu.edu.jo

Abstract

Dataset partitioning problem involves the vertical partitioning of the classification datasets into suitable subsets that preserve or enhance the classification quality of the original datasets. Typical classification model needs to be constructed for each subset and all generated models are then combined to form the classification model. This paper presents a dataset partitioning approach for rough set based classification. In this approach, the dataset is partitioned into two mutually exclusive subsets. Local reduct set is generated for each attribute subset which is then combined and used to generate the set of classification rules. A preliminary experimental result using the partitioning approach over some standard medical datasets showed that the approach preserves the classification accuracy.

***Key words*:**
Data Mining; Rough Sets; Feature Set Decomposition; Classification.

1. Introduction

Data dimensionality reduction is one of the main open problems in the field of data mining. There are two main forms of attribute driven data dimensionality reduction, namely feature subset selection and feature set partitioning (decomposition). Several approaches have been proposed and implemented for feature subset selection [H. Liu & L.Yu. 2002]. Feature subset selection, as a preprocessing step for data mining is the task of focusing the attention of an induction algorithm on subset of the given input attributes while ignoring the rest attributes. It discards those attributes that are irrelevant to the learning target.

For decomposition, instead of focusing on a single subset, the set of all input attributes is partitioned and the induction algorithm is run over each subset without ignoring any attribute from the input set of attributes. In this case the dataset is partitioned into individual useful parts rather than reduce the attribute set to a single useful group. Some approaches have been proposed for attribute partitioning [Bohanec 2002, A. Kusiak, 2000]

Zupan *et al*. [1997, 2002] proposed an approach for attribute decomposition based on function decomposition method used in the area of switching circuits design. According to this approach, attributes are transformed into new concepts in an iterative manner and creates a hierarchy of concepts. To find the new concept, the problem is transformed into a Graph Coloring problem, which is an NP hard problem.

Another approach proposed by Kusiak [A. Kusiak, 2000] is to use one of the features of the original dataset as an intermediate decision class for the first partitioned dataset and use another feature for the next partition. This type of decomposition is useful where the dataset was collected from environments that have incremental decisions.

This paper presents the problem of attribute partitioning for data mining tasks and proposes an approach for vertical partitioning of attributes set for rough set based classification to overcome the problem of large number of attributes in the decision tables. Moreover, the approach could be used for even small attributes set to enhance the quality of the output and to reduce the execution time of the reducts generation process.

The approach partitions the decision table into smaller sub decision tables where the decision class is attached to each sub table and local reducts are generated for each sub table. Combining the set of local reducts of the subsets will generate the global reduct set. The set of rules are generated for the global set of reducts. The classification accuracy is used to measure the quality of the output of classification process. A preliminary experimental results using two subset partitioning over some standard machine learning medical datasets, show that the approach preserve the classification accuracy with less computation time.

The rest of the paper is organized as follows: In section 2, for the convenience, we present briefly some notion needed in the paper from Rough Set Theory. The classification task in data mining is introduced in general

and presented in section 3. Section 4 presents the attribute partitioning problem. The Proposed approach of for attribute partitioning is presented in section 5. The experimental work and discussion is presented in section 6. Conclusion and future work are presented in section 7.

2. Rough Set Preliminaries

Rough set theory [Pawlak Z. 1991] was developed in Poland in the early 1980s as a mathematical tool for knowledge discovery and data analysis, and concerns itself with the classificatory analysis of imprecise, uncertain or incomplete expressed in terms of data acquired from experience. The notion of classification is central to the approach; the ability to distinguish between objects, and consequently reason about partitions of the universe. Rough set theory has been adopted in many researches for several data mining tasks including classification [Bazan J. et al, 2000]

In rough set theory, objects are perceived through the information that is available about them, that is, through their values for a predetermined set of attributes. In the case of inexact information, one has to be able to give and reason about rough classifications of objects.

The structure of data is represented in the form of **Decision System/Table** (**DS**). The decision system is a pair of the form **DS** = (U, **A**$\cup${**d**}), where U is a nonempty finite set of objects called the Universe, while **A** is a nonempty finite set of attributes. Every attribute a$\in$**A** is a total function a:U$\rightarrow V_a$, where V_a is the set of allowable values for the attribute a (i.e, its values range). The attributes belonging to **A** are called conditional attributes while **d** is called decision attribute.

For each possible subset of attributes **B** $\subseteq$ **A**, a decision table gives rise to an equivalence relation called an **Indiscernibility Relation IND(B)**, where two objects (x,y) are members of the same equivalence class if and only if they cannot be discerned from each other on the basis of the set of attributes **B**. The formal definition of **IND(B)** can be expressed as: **IND(B)={(x,y) $\in$ |U|×|U| : $a(x) = a(y)$ $\forall$ $a \in$ B}**. The **discernibility matrix (M)** of a decision system is a symmetric |U| × |U| matrix with entries c_{ij} defined as $\{a \in A | a(x_i) \neq a(x_j)\}$ if $d(x_i) \neq d(x_j)$, Φ otherwise.

A **Reduct (R)** of **A** refers to the minimal selection of attributes that can be used to represent all classes of the decision system. A reduct has two main properties: (1) **C(R) = C(A)**, i.e., **R** produces the same classification (**C**) of objects as the collection **A** of all attributes. (2) for any attribute $a \in$ **R, C(R-{**a**}) $\neq$ C(R),** i.e., a reduct is a minimal subset with respect to property (1).

3. Classification in Data Mining

The classification task concentrates on predicting the value of the decision class for an object among a predefined set of classes values given the values of some given attributes for the object. In the literature many classification approaches have been proposed and implemented by researchers, such as, decision tree based classification, statistical classification, neural network based classification, genetic algorithms classifiers and rough set based classification [K. Cios et al, 1998].

In general, data classification is a two-step process where in the first step, which is called the learning step, a model that describes a predetermined set of classes or concepts, is build by analyzing a set of training database objects. Each object is assumed to belong to a predefined class. In the second step, the model is tested using a different data set. The classification accuracy is estimated by computing the number of the correctly classified objects. If the accuracy of the model is considered acceptable, the model can be used to classify future data objects for which the class label is not known.

4. Attribute Set Partitioning

As mentioned earlier, decomposition is used for dimension reduction of data that is used in the data mining task in hand. Practically there are two forms of data set decomposition: (1) Feature set decomposition, where a data set is partitioned vertically basing on features. (2) Object set decomposition, where a data set is partitioned horizontally basing on objects. The two types of decomposition offer several benefits to data mining algorithms in term of computation and structure [Maimon O. and Rokach L. 2002; A. Kusiak, 2000]. Some benefits of using decomposition include: (1) Conceptual simplification of the problem by reducing the dimensionality. (2) Achieving clearer results for each sub set where in the case of classification we may have a small number of rules with compact size. (3) Reducing run time by solving smaller problems. (4) Suitable to be used in Distributed and Parallel environments. (5) Allowing different solution techniques for individual problems. (6) Suitable for Visualization of the output. (7) Allowing different solution techniques for individual sub problem.

Considering all partitions in search space is highly intractable. Optimization techniques or restriction based techniques are highly recommended. Some restricted decomposition approaches have been discussed in [A. Kusiak, 2000]. The ways that could be used for choosing the partitioning attribute for binary partition could include: (1) Manually based on expert's knowledge on a specific domain. (2) Arbitrarily based on a random chosen attribute. (3) Due to some restrictions. (4) Induced by a suitable algorithm.

5. The Proposed Approach

One of the simplest set partitioning approaches is to partition the dataset into two random partitions. The partitioning is done assuming that the attributes are not correlated. Each disjoint subset is independently used in the generation of classification model. The partitions could be Balanced or Imbalanced. Since the main objective is to reduce the exponential computational time of reduct generation, in our approach we search for the best-balanced partitions that enhance or preserve the classification accuracy of a dataset.

In the proposed partitioning approach, attributes are grouped into subsets along with intermediate concept class. The approach considers the decision class as an intermediate class for each subset. In this way we can measure the dependency of the decision class to each partitioned subset.

To demonstrate the effectiveness of the idea of partitioning on the classification task, we adopted the binary partitioning approach by limiting the number of partitions to two partitions i.e. the data set is partitioned into only two data subsets (P&P′). The partitioned attribute set is chosen randomly within some predefined constraints. A threshold is used to limit the size of the partitioned subset to be not less that 25% of the original dataset size. Some previous experiment showed that small size of subset does not give good classification accuracy. Moreover, in partitioning we are looking for a balanced partitioning as much as possible. If a small subset (P) is chosen the complement subset (P′) will be large. Another constraint considered is to choose neighboring attributes by choosing random chops from the attribute set rather than choosing single attribute to form the subset P. The steps of the approach is presented below:

Input:
S = Set of attributes $\{a_1,..,a_n\}$,
where n = |S| which denote the cardinality of S.
T_r =Training Dataset & T_t = Test Dataset.

Output:
Two partitions with Best Classification Accuracy
Classifier of the two partitions.

Steps:
Initialize Parameters.
CA_s = Classify and get Classification Accuracy (S)
While more possible binary partitions *I*
and *Best-CA* < CA_s do
Choose subset of S => P_i
{Where P_i satisfies some predefined constrains.}
P_i = set of attributes in first part.
P_i' = set of attributes in second part
(i.e., the complement of P_i)
R_i = Get Set of Reducts (P_i)
R_i' = Get Set of Reducts (P_i')
Red = $R_i \cup R_i'$
$RuleSet_i$ = Generate Rules (Red)
CA_i = Classify and get Classification Accuracy
(T_t, $RuleSet_i$)
If CA_i > Best_CA Then
Best_CA = CA_i
Best_Rules = $RuleSet_i$
End while
Generate Classifier (Best_Rules)

---

6. Experiments and Discussions

The main purpose of this experiment is to investigate the influence of attribute partitioning on the Rough Set base classification. The Rosetta rough sets based toolkit [Ohrn 1998] was used to test the proposed decomposition approach. Rosetta is a toolkit application, which allows for the analysis of tabular data using the rough set methodology. It is a Windows based application with a GUI front-end and computational kernel. Using Rosetta, the entire rough set experimental process can be carried out from data completion to data classification.

The experiments reported here used three real world standard medical dataset. The datasets are obtained from the machine learning data repository at the University of California at Irvine (UCI) [Blake C. et al. 1998]. The three data sets along with their characteristics are presented in Table 1. The output of classification process using the complete feature set is also presented in Table 1. The output includes the number of generated reducts, number of generated rules, and the total classification accuracy. To generate reducts, there are several proposed approaches [Qasem et al 2003; Azura et al 2002]. For consistency, in all experiments and for all partitions we have adopted the Genetic Algorithm based reduct generation technique implemented in Rosetta. The experiments showed that we could achieve good classification accuracy by generating between 10 to 30 possible partitioned subsets.

Table 1: Datasets features and classification accuracy using Rough Sets

No	Dataset	#Cond. Attr.	#Objects	#Classes	#Reducts	#Rules	Accuracy %
1	Heart disease	13	270	2	1739	3344	81.5
2	Cleveland	13	303	2	1896	6372	83.5
3	Lymphography	18	148	4	1467	2744	81.8

To evaluate each subset, we use a wrapper approach by adopting the classification accuracy base on rough set classification as an evaluation measure. The classification accuracy measure used in this experiment is computed using the confusion matrix. A confusion matrix contains information about actual and predicted classifications done by a classification algorithm. Performance of such algorithms is commonly evaluated using the data in the matrix. The classification accuracy is derived from the confusion matrix. The *accuracy* (*AC*) is the proportion of the total number of predictions that were correct. Table 2 shows the confusion matrix for the ***Heart Disease*** dataset. The total Accuracy of the classification is computed in the right lower corner cell, which is equal to 81.5%

*Table 2: The Confusion Matrix of the **Heart** Disease dataset*

		Predicted		
		1	**2**	
Actual	**1**	**38**	**8**	**82.61**
	2	**7**	**28**	**80.00**
		84.44	**77.78**	***81.48***

The preliminary results showed that using partitioning we can achieve most of partitioning benefits presented earlier. The main achievement is enhancing or at least preserving classification accuracy with leas computation time for reduct computation. Sample of the partitions for the ***Heart Disease***, ***Cleveland***, *and* ***Lymphography*** datasets are found and presented in Table 3, Table 4, and Table 5 respectively.

We have evaluated the presented partitioning based classification approach over the three medical datasets. In term of classification accuracy, the approach preserves and even enhances the accuracy over the accuracy of using the complete features set. The partitions that give the highest accuracy are highlighted in the tables in bold. It is observed that the classification accuracy of the partitioned model performs well on all datasets. It shows significantly good results in the Heart dataset. The comparison of the classification accuracy using the complete set of attributes and using the partitioned sets is depicted in Figure 1.

*Table 3: Classification of the **Heart Disease** dataset using Partitioning.*

No	Feature Set Partitions	#Reducts	Accuracy %
1	{1..4,d},{5..13,d}	263	79.1
2	{1..5,d},{6..13,d}	157	79.1
3	{1..6,d},{7..13,d}	102	80.3
4	{1..7,d},{8..13,d}	107	81.5
5	{1..8,d},{9..13,d}	161	76.6
6	{1..9,d},{10..13,d}	280	77.8
7	{1..10,d},{11..13,d}	510	79.1
8	{1,3,5..13,d},{2,4,6..12,d}	118	79.1
9	{1..5,10..13,d},{6..9,d}	298	**90.1**
10	{1..4,11..13,d},{5..10,d}	112	81.5

*Table 4: Classification of the **Cleveland** dataset using Partitioning.*

No	Feature Set Partitions	#Reducts	Accuracy %
1	{1,2,d},{3,..,14,d}	841	81.3
2	{1,..,4,d}{5,..,14,d}	330	78.0
3	{1,..,5,d},{6,..,14,d}	196	80.2
4	{1,..,6,d},{7,..,14,d}	148	82.4
5	{1,..,7,d},{8,..,14,d}	143	85.7
6	{1,..,10,d},{11,..,14,d}	513	80.2
7	{1,3,5,..,13,d},{2,4,6,..,14,d}	153	**84.6**
8	{1,3,5,7,d},{2,4,6,8,9,..,14,d}	280	78.0
9	{2,4,6,8,d},{1,3,5,7,9,10,..,14,d}	346	82.4
10	{4,..,11,d},{1,2,3,12,13,14,d}	205	80.2

*Table 5: Classification of the **Lymphography** dataset using Partitioning.*

No	Feature Set Partitions	#Reducts	Accuracy %
1	{1,..,5,d},{6,..,18,d}	645	77.3
2	{1,..,7,d},{8,..,18,d}	406	79.6
3	{1,..,9,d},{10,..,18,d}	293	**81.8**
4	{1,..,11,d},{12,..,18,d}	250	72.7
5	{1,..,13,d},{14,..,18,d}	438	75.0
6	{1,..,4,15..18,d},{5,..,14,d}	258	77.3
7	{1,..,5,14,..,18,d},{6,..,13,d}	178	77.3
10	{1,3,5,..,15,d},{2,4,6,..,18,d}	266	77.3

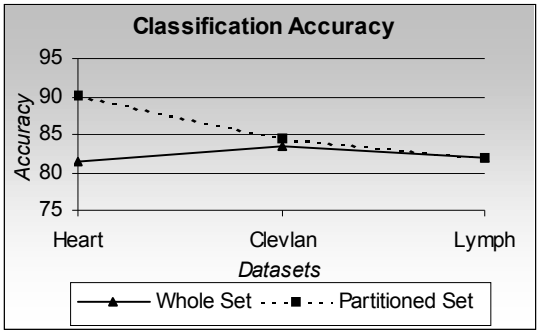

Figure 1: Classification Accuracy using whole set of attributes and partitioned sets.

7. Conclusion

The problem of vertical dataset partitioning for data mining tasks is presented in this paper. Dataset vertical partitioning is used to overcome the problem of large number of attributes in the decision table. An approach for vertical partitioning of attributes set for rough set based classification is presented in the paper. In the approach the dataset is partitioned into two mutually exclusive subsets and local reduct set is generated for each attribute subset. The experimental results showed that the approach could be used for even small attributes set to enhance the quality of the output and to reduce the execution time of the reducts generation process.

More research and experiments need to be conducted using the concept of attribute decomposition. One approach is being investigated and experimented for attribute partitioning based on sorting the attributes according to their relevance to the decision class. A

measure of feature relevance would allow eliminating irrelevant combination of features, which will reduce the computational efforts.

References

[1] Azuraliza A. B., Nasir S., Mohd H. S., Mohammad O. 2002, *Propositional Satisfiability Algorithms to Find Minimal Reducts for Data Mining*, International Journal of Computer Mathematics, Vol 79(4), pp. 379-389.

[2] Bazan J., Nguyen H.S., Nguyen S.H., Synak P., Wróblewski J., 2000. *Rough Set Algorithms in Classification Problem.* In: L. Polkowski, S. Tsumoto and T.Y. Lin (eds.): Rough Set Methods and Applications. Physica-Verlag, Heidelberg, New York, pp. 49 – 88

[3] Bohanec M, Zupan B, Bratko I, Cestnik B, 2002, *A function-decomposition method for development of hierarchical multi-attribute decision models*, Decision Support Systems (1035).

[4] Cios K., Predrycs W., Swiniarski R. 1998. *Data Mining Methods for Knowledge Discovery*. Kluwer Academic Publishers.

[5] Kusiak A., 2000, *Decomposition in Data Mining: An Industrial Case Study*. IEEE Transactions on Electronics Packaging manufacturing, Vol. 23, No. 4.

[6] Liu H. & Yu L. *2002, Feature Selection for Data Mining.* Collected from citeseers.nj.nec.com/ on 3 Jan 2003.

[7] Maimon O and Rokach L., 2002. *Improving supervised learning by feature decomposition*, Proc. of the 2nd Inter. Symposium on Foundations of Information and Knowledge Systems, LNCS, Springer, pp. 178-196.

[8] Pawlak Z, 1991. *Rough Sets: Theoretical Aspects of Reasoning about Data*. Kluwer, Dordrecht. ISBN 0-7923-1472-7.

[9] Qasem A. A. Nasir S., Mohd H. S., Hamidah I. 2003**,** *An Empirical Comparison of Reduct Generation Approaches in the context of Rough Set based Classification*. Proc. of ICITNS'2003, Amman, Jordan.

[10] Rokach L and Maimon O., 2001. *Theory and Application of Attribute Decomposition*, Proc. of the 1st IEEE International Conference on Data Mining, IEEE Computer Society Press, pp. 473-480.

[11] Son H. Nguyen & Hoa S. Nguyen, 1998. *Discretization Methods in Data Mining*. In: L. Polkowski & A. Skowron (Eds.): Rough Sets in Knowledge Discovery. Physica-Verlag, Heidelberg, pp. 451-482.

[12] Zupan B, Bohanec M, Bratko I, Cestnik B. 1997 (a). *A dataset decomposition approach to data mining and machine discovery*. In Proc. of the 3rd International Conference on Knowledge Discovery and Data Mining, Newport Beach, CA, pp 299-302.

[13] Zupan B, Bohanec M, Bratko I, Demsar J. 1997(b), *Machine learning by function decomposition*, In: Machine Learning: Proc. of the 14th International Conference (ICML'97), Nashville, Tennessee, pp 421-429.

[14] Ohrn A., *The design and implementation of a knowledge discovery toolkit based on rough sets-The ROSETTA system.* In: L. Polkowski & A. Skowron (Eds.), Rough Sets in Knowledge Discovery 1: Methodology and applications, Physica-Verlag, Heidelberg, 1998, pp. 376-399