



# Feature Reduction Using Rough Set Theory for Medical Applications

<sup>1</sup>Chaitrali Esave, <sup>2</sup>Omkar Pawar, <sup>3</sup>Sejal Doshi, <sup>4</sup>K.S.Tiwari

<sup>1,2,3,4</sup>Department of E&TC, MESCOE, Pune, Maharashtra, India.

Email: <sup>1</sup>chaitu.esave123@gmail.com, <sup>2</sup>omkarpawar608@gmail.com, <sup>3</sup>doshiseju@gmail.com, <sup>4</sup>kanchan.tiwari@mescoepune.org

**Abstract**— Rough Set Theory (RST) is a soft computing technique for data mining. It provides a mathematical approach for criticizing of significance of objects and attributes, reduction of data. In this paper, for generating a reduct RST plays an important role. For discovery of knowledge, reduction of the attribute is an important process. Using Binary Discernibility matrix algorithm attribute reduct vector is formed. In this paper we are generating a reduct vector for a binary information system using a discernibility matrix and reduct generation algorithm for Wilson Breast Cancer database (WBCD). Discernibility matrix algorithm is used for computing the proximate reduct. Results are explained using the MATLAB software.

**Keywords**— Rough Set Theory, information system, decision table, discernibility matrix, reduct

## I. INTRODUCTION

Rough Set Theory (RST) is developed by Z. Pawlak in 1985. Nowadays data is increasing rapidly, so the data mining is important for dealing with such a huge data. Data mining is the process of extracting a pattern of data from large data set and transform it into a suitable structure, so that it will be useful for future work. RST is the mathematical approach for data redundancy. RST is mainly use to reduce the original data. RST does not required any extra information about the data. RST is the efficient algorithm for finding the patterns hidden in the data. RST is easy to understand and it also provides the evaluation of the significance of the data. RST is mainly use for approximation of concepts. Fundamental of RST is mainly the upper and lower approximation of spaces of sets. RST can be used for feature extraction, feature selection, data reduction, etc. Reduct is a set of important feature among all. RST is also a important tool for the solution of various problems such as knowledge analysis, quality of information, evaluation of data reduct of data, etc. Nowadays RST is widely used for analysis of data attributes and information.

In this paper we are generating a reduct vector for a binary information system using a discernibility matrix and reduct generation algorithm for Wilson Breast Cancer database (WBCD). Reducts are useful for classification of the information system. The results are obtained in the MATLAB software.

## II. ROUGH SET THEORY

RST is developed by Z. Pawlak in 1985. It is most efficient method which supports the attribute reduction. RST is mathematical approach for data redundancy. RST algorithm is mainly use for evaluation of significance of data. Some of the main preliminaries of RST are as follows:

### 1) INFORMATION SYSTEM

Representation of data in physical or the logical form in RST is nothing but the information system. It represented in terms of table.

#### Decision Table

All the condition attributes and decision attributes of Information system together forms a decision table.

#### Inconsistent Decision Table

Is the decision table which is not compatible if the two objects are same in condition attributes but different in decision attribute.

#### Condition Attributes

The attributes that does not decides to which class the object is belong and that are called as condition attributes.

#### Desion Attributes

The attributes which are definitely decides the class to which the object is belong are the decision attributes.

#### Reduct

Reduct is the set of specific attributes. Reduct is nothing but the important feature among all. It is an essential [part of an object which make the object discernible by the information system. It reduces computation cost for rule generation.

#### Core

Core is the intersection of all the reduct in the information system. Core can be expanded using reducts.

#### Rough set and Crisp set

The rough set is the subset which is defined through its upper and lower approximation and when the boundary region is non-empty. Whereas when the boundary region is empty that set is consider as crisp set.

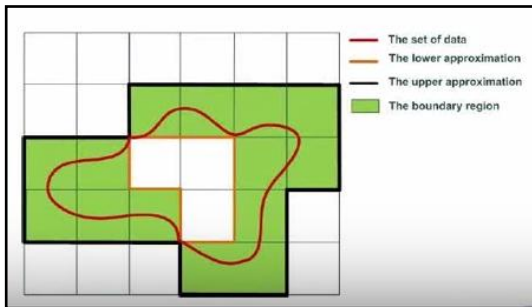


Figure 1: Regions of RST

**Upper Approximation**

Upper approximation is the set of all the elements which are possibly belongs to the subgroup of interest.

**Lower Approximation**

Lower Approximation is the set of objects that are certainly belonging to the X that is POS(X).

**Boundary region**

Boundary region is the set of elements that helps to decide that set is rough set or the crisp set.

**Block Diagram**

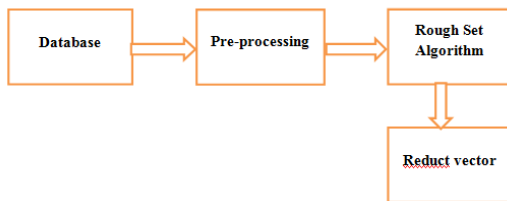


Figure 2: Block Diagram

**Algorithm for feature reduction**

1. Input information system.
2. Form a decision table.
3. Generate discernibility matrix.
4. Generate reduct vector.

**Algorithm for Discernibility matrix**

**Input:** Simplified Decision table

**Output:** Discernibility matrix

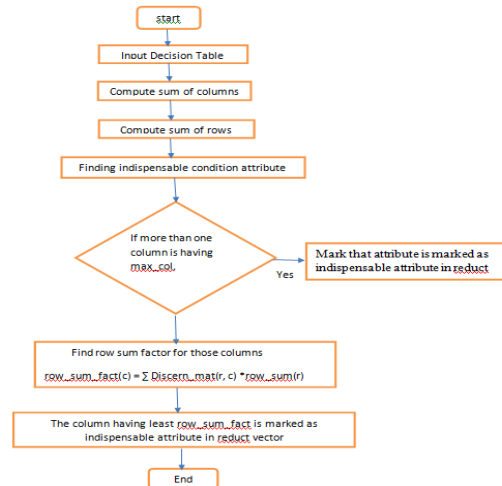
1. Input decision table.
2. Compare the rows.
3. If two row having different decision then X\_OR the two rows. Resulted row is the row of discernibility matrix.

4. IF decision of two row is different make the row 0 in discernibility matrix

**Algorithm for reduct generation**

1. Take the discernibility matrix.
2. Take the sum of column
3. take the sum of rows.
4. Find maximum sum from coloumn\_sum.
5. If only one column sum is maximum value among all the sum then that attribute is the indispensable attribute in reduct vector.
6. If more than one column is having maximum sum then find the row factorial for those coloumn
7.  $row\_fact(c) = \sum Discern\_matrix(r, c) * row\_sum(r)$
8. The column having lowest row\_\_fact is the indispensable attribute in the reduct vector.

**Flowchart for Reduct Generation**



**Algorithm for reduct vector generation**

1. AND the partially generated reduct and discernibility matrix.
2. If the result is non zero elimante that row from discernibility matrix.
3. If the result is zero keep those rows of the discernibility matrix.
4. This forms a new discernibility matrix.
5. Apply the reduct generation algorithm on the newly generated discernibility matrix.
6. Repeat the procedure till discernibility matrix becomes empty.

Result of reduct generation for binary information system

	c1	c2	c3	c4	c5	c6	c7	c8	c9	d
x1	1	1	0	0	1	1	0	0	1	1
x2	0	0	1	1	1	1	1	0	0	2
x3	1	0	1	1	1	1	0	0	0	3
x4	1	0	0	0	1	1	1	1	0	4
x5	1	1	1	1	0	0	0	1	1	2
x6	1	0	1	0	0	1	1	1	0	3
x7	1	1	1	0	0	0	1	1	1	4
x8	1	1	0	0	1	1	0	0	1	3

Figure 3: Decision Table

C =

1	1	1	1	0	0	1	0	1
0	1	1	1	0	0	0	0	1
0	1	0	0	0	0	1	1	1
0	0	1	1	1	1	0	1	0
0	1	1	0	1	0	1	1	1
0	0	1	0	1	1	1	1	0
0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0
1	0	1	1	0	0	0	1	0
0	0	0	0	0	0	0	0	0
1	0	0	1	1	0	0	1	0
1	1	0	1	1	1	0	1	1
1	1	1	1	0	0	1	0	1
0	0	1	1	0	0	1	1	0
0	1	0	0	1	1	0	1	1
0	0	0	0	0	0	0	0	0
0	1	0	1	1	1	1	1	1
0	0	0	0	0	0	0	0	0
0	1	1	1	1	1	1	0	1
0	0	1	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	1	1	1
0	1	0	1	0	1	1	0	1
0	0	0	1	0	0	1	0	0
0	0	1	1	1	1	0	1	0
0	1	0	0	0	1	0	0	1
0	0	0	0	0	0	0	0	0
0	0	1	0	1	1	1	1	0

Figure 4 : Binary discernibility matrix

reduct\_vector =

0	0	0	1	0	0	0	0	0
---	---	---	---	---	---	---	---	---

Figure 5: Partially generated reduct vector

### III. CONCLUSION

As we use the discernibility matrix and reduct vector algorithm for reduct generation the number of iterations are reduced.

Future work



All the algorithms will implemented on the hardware with the help of FPGA.

### REFERENCES:

- [1] Jiang Yu Du Bin , “A Efficiency Complete Algorithm for Attribute Reduction”, in Computer science and information technology .ICCSIT .2<sup>nd</sup> IEEE International Conference , pp. 377 - 379, Aug 2009,
- [2] Syed Hasnain Ali, Dr. Usman Qamar, Madiha Guftar, Abdul Wahab Muzaffar,” A Feature Reduction Framework based on Rough Set for Biomedical Data Sets”, in SAI Intelligent Systems Conference London, UK , pp.343 – 349 ,2015
- [3] Ming Hu and Qiang Zhang,” Application of Rough Sets to Image Pre-processing for Face Detection ”, Proceedings of the IEEE International Conference on Information and Automation Zhangjiajie, China, pp. 545 - 548,2008
- [4] Renu Vashist SMVDU, Katra, Jammu, INDIA, “An Algorithm for Finding the Reduct and Core of the Consistent Dataset”, International Conference on Computational Intelligence and Communication Networks, pp. 741 – 745,2015
- [5] Jing Tian, Quan Wang, Bing Yu,” A Rough Set Algorithm for Attribute Reduction via Mutual Information and Conditional Entropy”,10th International Conference on Fuzzy Systems and Knowledge Discovery , pp. 567 – 571,2013
- [6] Guifeng Zhang, Lina Yi,” Feature Selection Using Rough Set Theory for Object-Oriented Classification of Remote Sensing Imagery”, National High Technology Research and Development Program of China,pp. 1 – 7,2012
- [7] Dai Zhe, Liu Jianhui,” A Dimensionality Reduction Based On Rough Set Theory for Complex Massive Data”, 8th International Congress on Image and Signal Processing, pp. 1520 – 1528, 2015