ISSN: 2456-8848, Vol. I, Issue. I, June-2017

International Journal
of
Linguistics & Computing Research

Rough Set Theory: An Introduction

Girish Kumar Singh
Department of Computer Science and Application
Dr. Hari Singh Gour Central University
Sagar, Madhya Pradesh, India
gkrsingh@gmail.com

***Abstract*— In early 1980's Pawalak introduced the concept of Rough Set Theory. In very short span of time this theory become popular in soft computing. Rough set theory is associated with many other theories dealing with imperfect and vagueness data. This paper explains the Rough set theory and its basic concepts and has been illustrated with the help of example.**

***Keywords* — Decision System, Rough Set Theory, Reduct.**

I. INTRODUCTION

Rough Set theory introduced by Pawlak in early 1980's, is a technique for dealing with uncertainty and to identify cause-effect relationships in databases as a tool for data mining and database learning [1]. It has also been used for improved information retrieval and for uncertainty management in relational databases. Rough set Theory deals with those dataset in which some of the objects having same values for all conditional attribute but belongs to the two or more different classes. In this paper the concepts of RST are presented, defined and illustrated with the use of a representative Fruit dataset.

II. BASIC CONCEPTS OF RST

To understand the concepts of RST, we need to know some basic concepts. The basis of RST is information system and decision system. To explain the concepts fruit dataset which is given in table 1 has been used.

Fruit dataset has sixteen observations and each observation has five attributes namely skin, color, size, flesh and Eatable. Skin attribute describe the types of skin of the fruit, color attribute gives the color of the fruit. Size attribute give the size of fruit and by flesh attribute we can know the type flesh of the fruit. Eatable attribute inform us whether the fruit is café for eating or not. The value of eatable attribute is governed by the value of rest of rest of the attributes.

Table 1: Fruit Data

Object	Skin	Color	Size	Flesh	Eatable
O_1	hairy	brown	large	hard	safe
O_2	hairy	green	large	hard	safe
O_3	smooth	red	large	soft	danger
O_4	hairy	green	large	soft	safe
O_5	hairy	red	small	hard	safe
O_6	smooth	red	small	hard	safe
O_7	smooth	brown	small	hard	safe
O_8	hairy	green	small	soft	danger
O_9	smooth	green	small	hard	danger
O_{10}	hairy	red	large	hard	safe
O_{11}	smooth	brown	large	soft	safe
O_{12}	smooth	green	small	soft	danger
O_{13}	hairy	red	small	soft	safe
O_{14}	smooth	red	large	hard	danger
O_{15}	smooth	red	small	hard	safe
O_{16}	hairy	green	small	hard	danger

A. *Information System:*

A 3-tuple $S = (U, A, V_a)$ is called an *information system,* where U is a non-empty finite set of objects called the universe,

A is a non-empty finite set of attributes, for $\forall a \in A$, V_a is the value set of the attribute a. Table 2 is an example of information system. This table has only informative attributes no decision making attribute.

Table 2: Information System

Object	Skin	Color	Size	Flesh
O_1	hairy	brown	large	hard
O_2	hairy	green	large	hard
O_3	smooth	red	large	soft
O_4	hairy	green	large	soft
O_5	hairy	red	small	hard
O_6	smooth	red	small	hard
O_7	smooth	brown	small	hard
O_8	hairy	green	small	soft
O_9	smooth	green	small	hard
O_{10}	hairy	red	large	hard
O_{11}	smooth	brown	large	soft
O_{12}	smooth	green	small	soft
O_{13}	hairy	red	small	soft
O_{14}	smooth	red	large	hard
O_{15}	smooth	red	small	hard
O_{16}	hairy	green	small	hard

B. *Decision System:*

A *Decision system* is any information system of the form $S = (U, A \cup D, V_a)$, where $A \cap D = \phi$, D the set of decision attributes and A the set of conditional attributes. The fruit dataset as given in the table 3 is an example of a decision system. This table has decision making attribute in addition to informative attributes.

Table 3: Fruit Data as Decision System

Object	Skin	Color	Size	Flesh	Eatable
O_1	hairy	brown	large	hard	safe
O_2	hairy	green	large	hard	safe
O_3	smooth	red	large	soft	danger
O_4	hairy	green	large	soft	safe
O_5	hairy	red	small	hard	safe
O_6	smooth	red	small	hard	safe
O_7	smooth	brown	small	hard	safe
O_8	hairy	green	small	soft	danger
O_9	smooth	green	small	hard	danger
O_{10}	hairy	red	large	hard	safe
O_{11}	smooth	brown	large	soft	safe
O_{12}	smooth	green	small	soft	danger
O_{13}	hairy	red	small	soft	safe
O_{14}	smooth	red	large	hard	danger
O_{15}	smooth	red	small	hard	safe
O_{16}	hairy	green	small	hard	danger

Following is the description of the fruit dataset as a decision system:

U= { O_1 , O_2 , O_3 , O_4 , O_5 , O_6 , O_7 , O_8 , O_9 , O_{10} , O_{11} , O_{12} , O_{13} , O_{14} , O_{15} , O_{16} }
A = {Skin, Color, Size, Flesh}
D = {Decision}
V_{Skin}= {hairy, smooth}
V_{Color}= {brown, green, red}
V_{Size}= {small, large}
V_{Flesh}= {soft, hard}
$V_{Decision}$={safe, danger}

C. *Indiscernibility Relation:*

The indiscernibility relation is at the core of rough set theory. All concepts of rough set theory are based on indiscernibility relation. Any two objects are said to be indiscernible if the vectors representing the two objects are identical i.e. the two tuples are identical. Two objects may be indiscernible with respect to B⊆A if the attribute values of the attributes in B for the two objects are identical. Indiscernibility relation in on information system S denoted by $IND_S(B)$ for any B⊆A is a relation on U defined as,

$$IND_S(B) = \{(y, y') \in U \times U \mid \forall a \in B, a(y) = a(y')\} \quad \ldots (1)$$

If $(y, y') \in IND_S(B)$, then objects y and y′ are indiscernible from each other with respect to all attributes in B then $IND_S(B)$ is called the *B-indiscernibility* relation. It is trivial to prove that $IND_S(B)$ for any B⊆A satisfies the reflexivity, symmetricity and transitivity conditions. Therefore, using the equivalence relation $IND_S(B)$, the set of equivalence classes yields partition of the universe U denoted by $\frac{U}{IND_S(B)}$. The equivalence class of y∈ U with respect to B-*indiscernibility* relation is denoted by $[y]_B$. The relation $IND_S(B)$ when applied to the entire universe may also be indicated as IND(B).

Consider B= {Skin, Color, Size, Flesh}, B_1= {Skin, Color, Size}, B_2= {Skin, Color, Flesh}, B_3= {Skin, Color}. The indiscernibility relations corresponding to these sets of attributes are illustrated below. In the following examples of indiscernibility relation only the distinct pairs are exhibited while trivial cases, the pairs indicating reflexivity i.e. (O_i, O_i), are not included.

$IND_S(B)$ ={(O_6,O_{15}), (O_{15},O_6)}

$IND_S(B_1)$ ={(O_2,O_4), (O_3,O_{14}), (O_4,O_2), (O_5,O_{13}), (O_6,O_{15}), (O_8,O_{16}), (O_9,O_{12}), (O_{12},O_9), (O_{13},O_5), (O_{14},O_3), (O_{15},O_6), (O_{16},O_8) }

$IND_S(B_2)$ ={(O_2,O_{16}), (O_4,O_8), (O_5,O_{10}) (O_6,O_{14}) , (O_6,O_{15}), (O_8,O_4), (O_{10},O_5), (O_{14},O_6), (O_{14},O_{15}), (O_{15},O_6), (O_{15},O_{14}) (O_{16},O_2)}

$IND_S(B_3)$ ={(O_2,O_4), (O_2,O_8), (O_2,O_{16}), (O_3,O_6), (O_3,O_{14}), (O_3,O_{15}), (O_4,O_2), (O_4,O_8), (O_4,O_{16}), (O_5,O_{10}), (O_5,O_{13}), (O_6,O_3), (O_6,O_{14}), (O_6,O_{15}), (O_7,O_{11}), (O_8,O_2), (O_8,O_4), (O_8,O_{16}), (O_9,O_{12}), (O_{10},O_5), (O_{10},O_{13}), (O_{11},O_7), (O_{12},O_9), (O_{13},O_5), (O_{13},O_{10}),

(O_{14},O_3), (O_{14},O_6), (O_{14},O_{15}), (O_{15},O_3), (O_{15},O_6), (O_{15},O_{14}), (O_{16},O_2), (O_{16},O_4), $(O_{16},O_8)\}$

Once the concept of information system, decision system and indiscernibility relation is define we can define Rough Set Tehory.

III. ROUGH SET

Based on the indiscernibility relation we define lower and upper approximation and boundary region. If for any set (dataset) boundary region is not empty then such set is rough set.

A. Lower and Upper Approximation:

Consider a concept $X \subseteq U$. The dataset U is described by the values of all the attributes in A. A description of X may also be possible based on the information of $B \subseteq A$. The lower and upper approximations of the concept X with respect to the B offer a formulation for such a description. The B-lower approximation and B-upper approximation of X are represented as $\underline{B}(X)$ and $\overline{B}(X)$ respectively and are defined by,

$$\underline{B}(X) = \{x : [x]_B \subseteq X\} \quad \text{and,}$$

$$\overline{B}(X) = \{x : [x]_B \cap X \neq \phi\} \qquad \ldots\ldots (2)$$

The approximation regions $\underline{B}X$ and $\overline{B}X$ of the concept X are defined using the equivalence classes of the indiscernibility relation IND(B). The objects in $\underline{B}(X)$ with certainty are the members of X (certainly describe X) on the basis of the knowledge in B, while the objects in $\overline{B}(X)$ are possible members of X (possibly describe X) based on the knowledge in B.

Consider the decision system represented by fruit dataset and let X= {O_1, O_2, O_4, O_6, O_7, O_9, O_{11}, O_{14}} and let B= {Skin, color} then the equivalence classes of U = {O_1, O_2 , …, O_{16}} with respect to B are given by,

$[O_1]_B = \{O_1\}$
$[O_2]_B = \{O_2, O_4, O_8, O_{16}\}$
$[O_3]_B = \{O_3, O_6, O_{14}, O_{15}\}$
$[O_4]_B = \{O_2, O_4, O_8, O_{16}\}$
$[O_5]_B = \{O_5, O_{10}, O_{13}\}$
$[O_6]_B = \{O_3, O_6, O_{14}, O_{15}\}$
$[O_7]_B = \{O_7, O_{11}\}$
$[O_8]_B = \{O_2, O_4, O_8, O_{16}\}$
$[O_9]_B = \{O_9, O_{12}\}$
$[O_{10}]_B = \{O_5, O_{10}, O_{13}\}$
$[O_{11}]_B = \{O_7, O_{11}\}$
$[O_{12}]_B = \{O_9, O_{12}\}$
$[O_{13}]_B = \{O_5, O_{10}, O_{13}\}$
$[O_{14}]_B = \{O_3, O_6, O_{14}, O_{15}\}$
$[O_{15}]_B = \{O_3, O_6, O_{14}, O_{15}\}$
$[O_{16}]_B = \{O_2, O_4, O_8, O_{16}\}$

The partition of U with respect to the equivalence relation IND(B) for B = {Skin, color} is,

$$U / IND(B) = \{\{O_1\}, \{O_2, O_4, O_8, O_{16}\}, \{O_3, O_6, O_{14}, O_{15}\}, \{O_5, O_{10}, O_{13}\}, \{O_7, O_{11}\}, \{O_9, O_{12}\}\}$$

In the above example the equivalence classes which are certainly contained in the X are $[O_1]_B$, $[O_7]_B$ and $[O_{11}]_B$. Therefore, the lower approximation of X with respect to B is

$$\underline{B}X = \{O_1, O_7, O_{11}\}.$$

For the objects O_1, O_2 ,O_3,O_4,O_6,O_7,O_8,O_9,O_{11},O_{12},O_{14},O_{15}, and O_{16}, it may be observed that $[O_1]_B \cap X \neq \Phi$, $[O_2]_B \cap X \neq \Phi$, $[O_3]_B \cap X \neq \Phi$, $[O_4]_B \cap X \neq \Phi$, $[O_6]_B \cap X \neq \Phi$, $[O_7]_B \cap X \neq \Phi$, $[O_8]_B \cap X \neq \Phi$, $[O_9]_B \cap X \neq \Phi$, $[O_{11}]_B \cap X \neq \Phi$, $[O_{12}]_B \cap X \neq \Phi$, $[O_{14}]_B \cap X \neq \Phi$, $[O_{15}]_B \cap X \neq \Phi$ and $[O_{16}]_B \cap X \neq \Phi$ therefore, the upper approximation of X with respect to B is computed to be,

$$\overline{B}X = \{O_1, O_2, O_3, O_4, O_6, O_7, O_8, O_9, O_{11}, O_{12}, O_{14}, O_{15}, O_{16}\}$$

Properties of Lower and Upper Approximation:

1. $\underline{B}(X) \subseteq X \subseteq \overline{B}(X)$
2. $\underline{B}(\Phi) = \overline{B}(\Phi) = \Phi, \quad \underline{B}(U) = \overline{B}(U) = U$
3. $\overline{B}(X \cup Y) = \overline{B}(X) \cup \overline{B}(Y)$
4. $\underline{B}(X \cap Y) = \underline{B}(X) \cap \underline{B}(Y)$
5. $X \subseteq Y$ implies $\underline{B}(X) \subseteq \underline{B}(Y)$ and $\overline{B}(X) \subseteq \overline{B}(Y)$
6. $\underline{B}(X \cup Y) \supseteq \underline{B}(X) \cup \underline{B}(Y)$
7. $\overline{B}(X \cap Y) \subseteq \overline{B}(X) \cap \overline{B}(Y)$
8. $\underline{B}(\underline{B}(X)) = \overline{B}(\underline{B}(X)) = \underline{B}(X)$
9. $\overline{B}(\overline{B}(X)) = \underline{B}(\overline{B}(X)) = \overline{B}(X)$

B. Boundary Region:

The set $BN_B(X) = \overline{B}X - \underline{B}X$ is called the *boundary region* of X, which consists of those objects whose membership to X is not decisive on the basis of the knowledge in B. The set $U - \overline{B}X$ is said to be the *B-outside region* of X. It consists of objects which are with certainty classified as not belonging to X on the basis of knowledge in B.

In the previous example, X= {O_1, O_2, O_4, O_6, O_7, O_9, O_{11}, O_{14}} and B= {Skin, color}. Since the lower and upper approximations of X with respect to B are,

$$\underline{B}(X) = \{O_1, O_7, O_{11}\}$$
$$\overline{B}(X) = \{O_1, O_2, O_3, O_4, O_6, O_7, O_8, O_9, O_{11}O_{12}, O_{14}, O_{15}, O_{16}\}$$

The boundary region may be obtained,

$$BN_B(X) = \{O_2, O_3, O_4, O_6, O_8, O_9, O_{12}, O_{14}, O_{15}, O_{16}\}$$

C. Rough Set:

A set is said to be *rough* if the boundary region is non-empty and *crisp* otherwise.

In the above example since $BN_B(X) \neq \Phi$ therefore, the set X is a rough set.

IV. ROUGH SET THEORY

A. Positive Region:

Rough Set Theory offers tools to measure the degree of significance of attributes and the dependencies amongst them. For a given set of conditional attributes B, the *B-positive region* $POS_B(D)$ with respect to the relation IND(D) is defined as,

$$POS_B(D) = \bigcup \{ \underline{B}X : X \in [x]_D \} \quad \ldots\ldots (3)$$

The positive region $POS_B(D)$ contains all the objects in U that can be classified without any error into distinct classes defined by IND(D), based only on information in B. Greater the cardinality of $POS_B(D)$ higher is the significance of the attributes in the set B with respect to D.

B. Rough Membership Function:

The Rough membership function $\mu_X^B(y)$ is a tool to express how certainly an element *y* belongs to the concept X by the information about the element with respect to the set of attributes B. The Rough membership function is also used as a measure of significance of an attribute and is defined by,

$$\mu_X^B(y) = \frac{card\ (X \cap [y]_{IND(B)})}{card\ ([y]_{IND(B)})} \quad \ldots. (4)$$

C. Reduct:

A reduct is a minimal subset of attributes with the same capability of object classification as the set of all attributes.

V. REDUCT COMPUTATION

Reduct is one of the most important concepts in application of rough set theory in data mining. A reduct is the minimal set of attributes preserving classification accuracy of the original dataset. The problem to compute the reducts of a dataset is similar to the problem of feature selection. All the reducts of a dataset are obtained by constructing a discernibility function from the dataset [2]. It has been shown that the problems of finding minimal reduct and all reducts are NP-hard problems. Therefore, efficient methods to solve this NP-hard problem play an important role in the development of rough set-based data mining. Some efficient algorithms with heuristics, GA approach, etc. have also been proposed. Starzyk has used strong equivalence to simplify discernibility function [Starzyk1998]. However, this is still an open problem in rough set theory.

The conventional reduct computational algorithms fall into two categories: the reduction algorithms based on heuristic information and the reduction algorithms based on random strategies. Nevertheless, these algorithms do not guarantee to find a complete set of reducts for the dataset.

A. Heuristic Algorithms

Johnson's strategy [3] is based on Johnson approximation algorithm for computing minimal prime implicant of any Boolean function in conjunctive normal form (CNF) formula. The main idea of the algorithm is to find an attribute discerning the largest number of pairs of objects, i.e., an attribute that occurs most in the entries of discernibility matrix. This algorithm proceeds until a reduct set is found. The time complexity of this algorithm is $O(|A|^2\ |U|^2)$ and the space complexity of this algorithm is $O(|A|\ |U|^2)$, where A is the set of attributes and U is database.

Jue Wang [4] has proposed an attributes reduction algorithm based on significance of attributes in discernibility matrix. In this algorithm significance of attributes is define as the attributes frequency in discernibility matrix. Hence algorithm regards the number of occurrences of each attribute as the significance of each attribute. The algorithm selects the attribute with the largest frequency, and deletes the elements involved with the selected attribute in discernibility matrix. Then the frequency of other attributes is computed. The algorithm continues to select and compute the frequency of remaining attributes until a reduct set is found. The time complexity of this algorithm is also $O(|A|^2\ |U|^2)$ and the space complexity of this algorithm is also $O(|A|\ |U|^2)$.

By making use of attribute frequency information in discernibility matrix, Keyun Hu [5] has developed a feature ranking mechanism. Hu has proposed the algorithm using feature ranking as heuristics for reduct computation. The time complexity of this algorithm is $O((|A|+\log|U|)\ |U|^2)$ and the space complexity of this algorithm is $O(|A|\ |U|^2)$.

X. Hu et al. [6] have proposed a new rough sets model and defined the core and reducts based on relational algebra using efficient set-oriented database operations. They presented two new algorithms to calculate core and reducts respectively, for feature selections. However, the time complexity of the algorithm is $O(|A|^2|U|)$ for the best case in spite of the hashing and indexing mechanism provided by the database systems.

B. Random Reduct Algorithms:

Vinterbo [7] has formulated the rough set based attribute reduction as 'minimal hitting set' problem. He has defined an r-approximate hitting set as a set that intersects with at least a fraction r of given sets. Approximations of reducts from rough set theory are defined by means of minimal r-approximate hitting sets. In this method r-approximate hitting sets is computed using GA. The time complexity of the algorithm is $O(|A|^2\ |U|\ \log\ |U|)$ and the space complexity is $O(|U|)$. Obviously, reducts obtained by this algorithm are not guaranteed to be complete.

Bazan [8] opines that the above methods do not take into account the fact that part of reduct set is chaotic i.e. it is not stable in randomly chosen samples of a given decision table. He introduced the notion of dynamic reduct. Dynamic reducts are in some sense the most stable reducts of the given decision table, i.e., they are the most frequently appearing reducts in subtables created by random sampling of a given decision table. Computation of reduct of variable size dynamically can be extremely computationally intensive, even for decision tables of moderately size. This algorithm is quite stable in most cases, yet it does not compute all reducts.

QuickReduct algorithm [9, 10] is an attempt to calculate a minimal reduct without exhaustively generating all possible

subsets. Starting with an empty set the algorithm constructs the set P by adding the attributes with highest value of the attribute dependency $\gamma_p(D)$, for D the decision attribute, until a maximum possible value is reached for the dataset (usually 1). Where

$$\gamma_p(D) = \frac{|POS_p(D)|}{|U|} \quad \text{.... (5)}$$

CONCLUSION

This paper presents the basics of the Rough set theory introduced by Pawalak in early 1980's. Basic concepts of Rough Set theory have been illustrated with the help of example. the concept of Rough Set Theory. Rough set theory is associated with many other theories dealing with imperfect and vagueness data.

REFERENCES

[1]. Z. Pawlak, "Rough sets." International Journal of Computer and Information Sciences 11 (1982), pp. 341-356.

[2]. S. K. Pal, A. Skowron, Rough Fuzzy Hybridization- A new trend in decision making, Springer, 1999.

[3]. D. S. Johnson, "Approximation algorithms for combinatorial problems", Journal of Computer and System Sciences, pp.256-278, 1974.

[4]. J. Wang and J. Wang, "Reduction algorithms based on discernibility matrix: the ordered attributes method", Journal of Computer Science & Technology, vol.16, no.6, pp. 489-504, 2001.

[5]. K. Hu, Y. Lu and C. Shi, "Feature Ranking in Rough Sets", AI Communications, Special issue on Artificial intelligence advances in China, Volume 16 , Issue 1, pp. 41 – 50, May 2003.

[6]. Hu, X., T. Y. Lin and J. Han, A new rough set model based on database systems, Journal of Fundamental Informatics, vol.59, pp.135-152, 2004.

[7]. S. Vinterbo and A. Ohrn, "Minimal approximate hitting sets and rule templates", International Journal of Approximate Reasoning, vol.25, no.2, pp.123-143, 2000.

[8]. J. G. Bazan, A comparison of dynamic and non-dynamic rough set methods for extracting laws from decision tables, in Rough Sets in Knowledge Discovery 1: Methodology and Applications, Polkowski and Skowron (editors), Physica-Verlag, Heidelberg, Germany, Chapter 17, pp. 321-365, 1998.

[9]. A. Chouchoulas and Q. Shen. Rough set-aided keyword reduction for text categorisation. Applied Artificial Intelligence, Vol. 15, No. 9, pp. 843-873, 2001.

[10]. R. Jensen and Q. Shen, "A Rough Set-Aided System for Sorting WWW Bookmarks", In Proceedings of the First Asia-Pacific Conference on Web Intelligence: Research and Development (WI'2001), pp. 95 – 105, 2001.