ISSN XXXX XXXX © 2017 IJESC

Research Article Volume 7 Issue No.7

Significance of Attributes using Rough Data Transition Probability Matrix

L.Jesmalar[1], S.Robinson Chellathurai[2]
Assistant Professor[1], Associate Professor[2]
Department of Mathematics
Holy cross college (Autonomous), Nagercoil, Tamil Nadu, India[1]
Scott Christian College, Nagercoil, Kanyakumari, Tamil Nadu, India[2]

Abstract:
The chief objective of this study is to show the usefulness of Rough set theory in Matrix theory. The aim of this paper is to significance of an attribute can be evaluated by lower rough data symmetric transition probability matrix from an information table.

Keywords: Lower rough data symmetric matrix, upper rough data symmetric matrix, lower rough data transition probability matrix, upper rough data transition probability matrix

1. INTRODUCTION

The theory of Rough set proposed by Polish computer scientist Zdzisław I. Pawlak [1,2,3]. There are two generalized method for Pawlak rough set model, the constructive and the algebraic methods. Some information systems may have no core attributes, in order to solve the problem to measure the importance of the degree of attribute, which can significantly decrease the ratio that the important attribute is taken as redundant attribute to remove. The idea of attribute reduction can be generalized by introducing a concept of *significance of attributes*.

1.1Algebraic Rough set method

Definition 1.1.1
A ***rough set*** is a formal approximation of a crisp set in terms of a pair of sets which the lower and upper approximation of the original set. Let U denote the set of objects called universe and let R be an equivalence relation on U. Then (U,R) is called an approximation space. For u,v ∈U & (u,v) ∈ R, u and v belong to the same equivalence class it is denoted by U/R and we say that they are indistinguishable. The relation R is called an indiscernibility relation. Let $[x]_R$ denote an equivalence class of R containing element x, then lower approximation$\underline{R}$(X) & upper approximation$\overline{R}$(X) for a subset X ⊆ U are defined by

$\underline{R}(X) = \{x \in U/[x]_R \subset X\}$, $\overline{R}(X) = \{x \in U/[x]_R \cap X \neq \phi\}$

Thus if an object $x \in \underline{R}(X)$ then "x surely belongs to X"

If $x \in \overline{R}(X)$ then "x possibly belong to X"

$R(X) = (\underline{R}(X), \overline{R}(X))$ is called a rough set with respect to R.

Definition 1.1.2.
The *membership value* of X is $\mu(X) = \frac{|\underline{R}(X)|}{|\overline{R}(X)|}$

The *membership value* of each element of X is $\mu_X(x) = \frac{|([x]_R \cap X)|}{|[x]_R|}$

Definition 1.1.3.

The ***rough membership*** can be interpreted as a degree that x belongs to X in view of information about x expressed by R. The rough membership function can be used to define approximations and the boundary region $BN_R(X)$ of a set:
$\underline{R}(X)=\{x \in U: \mu_X(x) = 1\}$
$\overline{R}(X)=\{x \in U: \mu_X(x) > 0\}$
$BN_R(X) = \{x \in U: 0 < \mu_X(x) < 1\}$

1.2 Constructive Rough set method
Approximations are fundamental concepts of rough set theory. Rough set based data analysis starts from a data table called a *decision table* or an *information system*, columns of which are labeled by *attributes*, rows – by *objects* of interest and entries of the table are *attribute values*. Attributes of the decision table are divided into two disjoint groups called *condition* and *decision* attributes, respectively. Each row of a decision table induces *decision rule*, which specifies decision (action, results, outcome, etc.) if some conditions are satisfied. If a decision rule uniquely determines decision in terms of conditions the decision rule is *certain*. Otherwise the decision rule is *uncertain*. *Lower approximation*– the set of items, which can be certainly classified as items of X. *Upper approximation*– the set of items, which can be possibly classified as items of X. *Boundary region*– the set of items, which can be classified either as items of X or not Set X is crisp with respect to R, if the boundary region of X is empty. Set X is rough with respect to R, if the boundary region of X is nonempty

Information System 1.2.1
Consider the simple information system (U,A) where $U=\{x_1, x_2,,, x_{10}\}$ set of objects, A={Age, I.Q, Eagerness to Learn, Communication Skill}with four conditional attributes and decision attribute D={ Performance} and a ∈A, the set of attributes a: U→V_a. The identity of the students.

Table 1

U	Age	I.Q	Eagerness to Learn	Communication Skill	Performance
x_1	16	90	Good	Oral, Written	Good
x_2	14	60	Good	Written	Good
x_3	15	90	Good	Oral, Written	Good
x_4	14	80	Fair	Written	Good
x_5	15	70	Fair	Written	Above average
x_6	16	60	fair	Oral	Above average
x_7	14	80	Bad	Oral, Written	Average
x_8	15	95	Bad	Oral	Average
x_9	16	90	Good	Oral, Written	Above average
x_{10}	15	95	Bad	Oral	Average

U = $\{x_1, x_2, x_3, x_4, x_5, x_6, x_8, x_9, x_{10}\}$
A = {Age, I.Q, Eagerness to Learn, Communication Skill}
V_{Age} = {14,15,16}={1,2,3}
$V_{I.Q}$ = {95, 90, 80, 70, 60}={1,2,3,4,5}
$V_{E.L}$ = {Good, Fair, Bad} = {1, 2, 3}
$V_{C.S}$ = {Oral Written, Written, Oral} ={1,2,3}
$V_{Performance}$= {Good, above average, Average} ={1,2,3}

A table may be redundant in two ways. The first form of redundancy is easy to observe. Some objects may have same features in all the attributes. This is true the case of objects x_1,x_9 and x_8,x_{10} in Table3.1.Here for reducing data it is enough if we store only one of the two. This has to be done for all the pairs. Such pairs are termed as *indiscernible objects*.

2. Significance of attributes using lower rough data transition probability matrix

Significance of an attribute can be evaluated by measuring effect of removing the attribute from an information table on classification defined by the table. Let us first start our consideration with decision tables.

Table.2. Value for reduced table1

U	Age (a_1)	I.Q (a_2)	Eagerness to Learn (a_3)	Communication Skill (a_4)	Performance (d)
x_1	3	2	1	1	1
x_2	1	5	1	2	1
x_3	2	2	1	1	1
x_4	1	3	2	2	2
x_5	2	4	2	2	2
x_6	3	5	3	3	2
x_7	1	3	3	1	3
x_8	2	1	3	3	3

Definition2.1: *Relation* between each object is defined by $d(x_i, x_j) = \sum_{k=1}^{n}(x_{ik} - x_{jk})$ where n is the number of attribute

Each element of the skew symmetric matrix is defined by $x_{ij} = d(x_i, x_j)$ where i, j=1,2,…m , where m is the number of object**.**

Table.3. Skew-Symmetric *matrix for Table2*.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0	-2	1	-2	-4	-8	-3	-4
x_2	2	0	3	0	-2	-7	-1	-2
x_3	-1	-3	0	-3	-5	-9	-4	-5
x_4	2	0	3	0	-2	-6	-1	-2
x_5	4	2	5	2	0	-4	-1	0
x_6	8	7	9	6	4	0	1	4
x_7	5	3	6	3	1	-3	0	-1
x_8	4	2	5	2	0	-4	1	0

Definition2.2

Relation between each attribute is defined by

$d(a_i, a_j) = \sum_{k=1}^{m} |x_{ki} - x_{kj}|$ where m is the number of object.

Each element of the symmetric matrix is defined by

$a_{ij} = d(a_i, a_j)$ where i, j=1, 2,…n , where n is the number of attribute

Table.4. Symmetric *matrix for Table2*

d	a_1	a_2	a_3	a_4
a_1	0	14	7	6
a_2	14	0	13	14
a_3	7	13	0	3
a_4	6	14	3	0

$R_{a_1} = \sum_{k=1}^{4} a_{k1} = 27, R_{a_2} = \sum_{k=1}^{4} a_{k2} = 41,$

$R_{a_3} = \sum_{k=1}^{4} a_{k3} = 23, \; R_{a_4} = \sum_{k=1}^{4} a_{k4} = 23$

$$R_{a_3} \le R_{a_2} \le R_{a_1} \le R_{a_2}$$

Definition2.3 : Let $\underline{R}_{ij} = \frac{\sum_{k=1}^{m} |x_{ki} - x_{kj}|}{R_{a_i}}$ for all $x_{ij} \in \underline{R}(X)$ where i, j=1, 2,…n, n is the number of attribute, $R_{a_i} = \sum_{k=1}^{n} a_{ki}$. The matrix $\underline{R} = (\underline{R}_{ij})$ is called a ***Lower rough data symmetric matrix*** if $\underline{R}(X)$ is the lower approximation of the information system under the conditional and decision attributes.

Definition2.4 : Let $\overline{R}_{ij} = \frac{\sum_{k=1}^{m} |x_{ki} - x_{kj}|}{R_{a_i}}$ for all $x_{ij} \in \overline{R}(X)$ where i, j=1,2,…n, n is the number of attribute, $R_{a_i} = \sum_{k=1}^{n} a_{ki}$. The matrix $\overline{R} = (\overline{R}_{ij})$ is called an ***upper rough data symmetric matrix*** if $\overline{R}(X)$ is the lower approximation of the information system under the conditional and decision attributes.

Definition2.5: The matrix $\underline{R} = (\underline{R}_{ij})$ is called a ***Lower rough data symmetric transition probability matrix*** satisfying the conditions

(i) $\underline{R}_{ij} \ge 0$ where $\underline{R}_{ij} = \frac{\sum_{k=1}^{m} |x_{ki} - x_{kj}|}{R_{a_i}}$, m is the number of object.

(ii) $\sum \underline{R}_{ij} = 1$ for all i.

Definition2.6: The matrix $\overline{R} = (\overline{R}_{ij})$ is called an upper ***rough data symmetric transition probability matrix*** satisfying the conditions

(i) $\overline{R}_{ij} \ge 0$ where $\overline{R}_{ij} = \frac{\sum_{k=1}^{m} |x_{ki} - x_{kj}|}{R_{a_i}}$, m is the number of object.

(ii) $\sum \overline{R}_{ij} = 1$ for all i.

Definition2.7

The matrix $\underline{R} = (\underline{R}_{ij})$ is said to be a *regular* matrix if all the entries of $(\underline{R}_{ij})^m$ are positive.

The matrix $\overline{R} = (\overline{R}_{ij})$ is said to be a *regular* matrix if all the entries of $(\overline{R}_{ij})^m$ are positive.

Definition2.8: If the lower rough t.p.m is regular, then every state value approaches a unique fixed value called the *steady state solution.* That is $\underline{R}^{(r)} \to \pi$ as $r \to \infty$ where $\underline{R}^{(r)} = \{\underline{R}_1^{(r)}, \underline{R}_2^{(r)}, \ldots \underline{R}_k^{(r)}\}$ and $\pi = (\pi_1, \pi_2, \ldots \pi_k)$

Definition2.9: If $\underline{R}$ is the regular lower rough t.p.m and $\pi = (\pi_1, \pi_2, \ldots \pi_k)$, then $\pi\underline{R} = \pi$ and $\pi_1 + \pi_2 + \cdots + \pi_k = 1$

Example 2.10

Lower Rough data symmetric transition probability matrix for Table 4 is

$$\begin{array}{c} \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{array} \begin{array}{c} \begin{array}{cccc} a_1 & a_2 & a_3 & a_4 \end{array} \\ \begin{bmatrix} 0 & \frac{14}{27} & \frac{7}{27} & \frac{6}{27} \\ \frac{14}{41} & 0 & \frac{13}{41} & \frac{14}{41} \\ \frac{7}{23} & \frac{13}{23} & 0 & \frac{3}{23} \\ \frac{6}{23} & \frac{14}{23} & \frac{3}{23} & 0 \end{bmatrix} \end{array}$$

We require the consistency of each attribute, so we can find the steady state solution of the given lower rough data t.p.m

$$(\pi_1, \pi_2, \pi_3, \pi_4) \begin{bmatrix} 0 & \frac{14}{27} & \frac{7}{27} & \frac{6}{27} \\ \frac{14}{41} & 0 & \frac{13}{41} & \frac{14}{41} \\ \frac{7}{23} & \frac{13}{23} & 0 & \frac{3}{23} \\ \frac{6}{23} & \frac{14}{23} & \frac{3}{23} & 0 \end{bmatrix} = (\pi_1, \pi_2, \pi_3, \pi_4)$$

$$\frac{14}{41}\pi_2 + \frac{7}{23}\pi_3 + \frac{6}{23}\pi_4 = \pi_1$$
$$\frac{14}{27}\pi_1 + \frac{13}{23}\pi_3 + \frac{6}{23}\pi_4 = \pi_2$$
$$\frac{7}{27}\pi_1 + \frac{13}{41}\pi_2 + \frac{3}{23}\pi_4 = \pi_3$$
$$\frac{6}{27}\pi_1 + \frac{14}{41}\pi_2 + \frac{3}{23}\pi_3 = \pi_4$$

Solving the above equations and using $\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$, we get $(\pi_1, \pi_2, \pi_3, \pi_4) = (0.22, 0.35, 0.18, 0.23)$ this is the consistency of each attribute. Thus the attribute a_1 and a_4 are equal importance in decision making. And the attribute a_2 is the most important attribute in decision making.

3. REFERENCES

[1]. Pawlak, Z. Rough set theory and its applications, Journal of tele communications and information technology, 3/2002.

[2]. Pawlak, Z. Rough sets, Theoretical aspects of reasoning about data, Kluwer Academic Publishers, Dordrecht,1991.

[3]. Pawlak, Z. Rough sets, International Journal of computer and information sciences 11:341-356, 198211:341-356,1982

[4]. S.Robinson Chellathurai & L.Jesmalar, Core in Rough Graph and Weighted Rough Graph, International Journal of Contemporary Mathematical Sciences ,Vol. 11, 2016, no. 6, 251 – 265,ISSN:1314-7544 HIKARI Ltd.

[5]. S.Robinson Chellathurai &L.Jesmalar, Rough sets on Clustering, International Journal of Engineering and Computing, Volume.7, Issue No.3, PP 5032-5035, ISSN (2321-3361), March-2017. Impact factor:5.611

[6]. S.Robinson Chellathurai&L.Jesmalar, Rough sets-Application in data mining, Journal of computing Technologies, Volume 3 Issue 1,p 1- 8,ISSN (2278-3814),June-2014. Impact factor:3.017

[7]. Venkatarama Krishnan ,Probability and Random Processes, Wiley publications ISBN: 978-0-471-99828-0