IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-3, Issue-6 E-ISSN: 2347-2693

Flowgraph Representation of Information System Using Rough Set Theory

Niharika Upadhyay[1*] and Pragati Jain[2]

[1*] Department of Science, Pacific University, India
[2] Department of Science, St. Paul Institute of Professional Studies, India

www.ijcseonline.org

Received: June/02/2015 Revised: June/08/2015 Accepted: June/24/2015 Published: June/30/ 2015

***Abstract*—** Lot of work has been done to classify raw data, for e.g. Data Mining, Machine Learning etc. But in these approaches uncertainty is not focused. Thus in this we are trying to find accuracy when uncertainty is in raw data. The starting point of rough set theory is an information system In this paper, Rough Set Theory is applied on 9 different mobile brands. Survey is conducted where 334 participants were asked to evaluate these 9 brands on different parameters. Using this information, Decision table and Flow Graph will be created and will found accuracy and strength of same.

***Keywords*—** *rough sets, information system and Flow Graph.*

I. INTRODUCTION

In day to day life, we come across with the incomplete or imprecise information or knowledge to understand our surroundings, to learn new things, and to make plans for the future. Pawlak [1,2,3,4] first proposed rough sets theory in the early 1980s. Rough Set Theory has been employed in various domains, including the information, electrical, environmental, engineering, medical, economics, finance, social science, chemical science, and decision analysis domains. In addition, RST can be used in the classification analysis of data tables and removes redundant conditional attributes according to two approximation concepts (lower and upper approximations). This Theory is based on the original data only and it does not require external information [5, 6, 7].

II. INFORMATION SYSTEM AND APPROXIMATION OF SETS

In this section we review some notions of Rough Set which we will use throughout this paper.

Defination1 [8]: Any 4-tuples S=<U, A_t, V, ρ> is called an Information System, where U is a finite set of objects. A_t is a finite set of attributes V= $\bigcup_{a\in A_t} v_a$, where v_a is a domain of the attributes, $\rho : U \times A_t \to V$ is a function called as Information function by $\rho(x,a) = a(x) \in v_a$ for every $a \in A_t$ and $x \in U$. An Information system is denoted by S=(U, A_t)

Defination2 [8]: Let S=(U, A_t) be an Information System. Every $B \subseteq A$, generates a binary relation on U, which is called an indiscernibility relation, and defined by $IND(B) = \{(x, y) \in U \times U : a(x) = a(y); \forall a \in B\}$. It is denoted by IND(B) or simply by B.

Defination3 [8]: Let S=(U, A_t) be an Information System and $B \subseteq A_t$ .The pair A=(U,IND(B)) is called an approximation space.

Defination4 [8]: Let A=(U,IND(B)) be an Information System and $B \subseteq A_t$. Then $[x]_B = \{ y \in U : (x, y) \in IND(B)\}$ is the equivalence class of $x \in U$ in IND(B). it is called atom set or elementary set. The set of atoms will be denoted by U/IND(B).

Defination5 [4]: Let A=(U,IND(B)) be an approximation space and $X \subseteq U$. The Upper approximation of X is A is defined as

$$\overline{A}(X) = \{ x \in U : [x]_B \cap X \neq \varphi \}$$

Defination6 [4]: Let A=(U,IND(B)) be an approximation space and $X \subseteq U$. The Lower approximation of X is A is defined as

$$\underline{A}(X) = \{ x \in U : [x]_B \subseteq X \}$$

III. DECISION TABLE

In this section we are going to review the concept of decision table and decision rules. Also we introduce definition related to decision table.

Definition [9]: let S=(U,A) be an Information system. If there are $C, D \subseteq A$, such that $C \cap D = \phi$ and $C \cup D = A$. Then S is called a decision table, which we denote by S=(U,C,D). we call them C,D condition and decision attributes, respectively.

IV. IN TABLE

The table contains data concerning 9 segment of mobile that have been tested for different parameters.

Table 1 is a decision table. It describes the serve result from group of 334* persons for 9 brands of mobile which are denoted by A, B, C, D, E, F, G, H. I. Option is given to each person to choose the best phone, average phone and worst phone among these nine brands of mobile. Column measuring factor represent values for best phone, average phone and worst phone as + , 0 and – respectively. Support column represents number of votes. For example row one says Brand A was selected as best phone (Measuring Factor = +) by 20 persons (Support = 20).

This decision table contains only one condition attribute that is Brand, whereas the decision attribute is measuring factor.

Sr. No.	Brand	Measuring Factor	Support
1	A	+	20
2	A	0	98
3	B	+	59
4	B	0	59
5	C	+	137
6	C	0	137
7	D	+	20
8	D	-	59
9	E	+	39
10	E	0	20
11	E	-	59
12	F	+	39
13	G	+	20
14	G	0	20
15	H	-	98
16	I	-	116

Table 1

[*Total of Support column should be 334 * 3 = 1002 but since two person has evaluated only two brand hence total is 332*3 + 2*2 = 1000.]

V. PROBABILISTIC PROPERTIES OF DECISION RULE

Let $C \xrightarrow[x]{} D$ be a decision rule, and let $\Gamma = C(x)$ and $\Delta = D(x)$. Then the following properties are

$$\sum_{y \in \Gamma} cer_y(C, D) = 1 \quad \ldots\ldots\ldots\ldots\ldots\ (1)$$

$$\sum_{y \in \Delta} \operatorname{cov}_y(C, D) = 1 \quad \ldots\ldots\ldots\ldots\ldots\ (2)$$

$$\Pi[D(x)] = \sum_{y \in \Gamma} cer_y(C, D).\Pi[C(y)] = \sum_{y \in \Gamma} \sigma_y(C, D) \quad \ldots\ldots\ldots\ldots\ldots\ (3)$$

$$\Pi[C(x)] = \sum_{y \in \Delta} \operatorname{cov}_y(C, D).\Pi[D(y)] = \sum_{y \in \Delta} \sigma_y(C, D) \quad \ldots\ldots\ldots\ldots\ldots\ (4)$$

$$cer_x(C, D) = \frac{\operatorname{cov}_x(C, D).\Pi[D(x)]}{\sum_{y \in \Delta} \operatorname{cov}_y(C, D).\Pi[D(y)]} = \frac{\sigma_x(C, D)}{\Pi[C(x)]} \quad \ldots\ldots\ldots\ldots\ldots\ (5)$$

$$\operatorname{cov}_x(C, D) = \frac{cer_x(C, D).\Pi[C(x)]}{\sum_{y \in \Gamma} cer_y(C, D).\Pi[C(y)]} = \frac{\sigma_x(C, D)}{\Pi[D(x)]} \quad \ldots\ldots\ldots\ldots\ldots\ (6)$$

This is decision table satisfy (1) to (6). To compute the certainty and coverage factors of decision rules according to formula (5) and (6), it is enough to know only the strength (support) of all decision rules.

These properties will be used as a basic for the rough set processor organization.

The certainty and coverage factor for the decision table presented in table 1 are shown in table 2.

Decision Rule	Strength	Coverage	Certainty
1	.02	0.0599	0.17
2	0.098	0.2934	0.83
3	0.059	0.1766	0.5
4	0.059	0.1766	0.5
5	0.137	0.4102	0.5
6	0.137	0.4101	0.5
7	0.02	0.0599	0.253
8	0.059	0.1777	0.747
9	0.039	0.1168	0.331
10	0.02	0.0599	0.169
11	0.059	0.1777	0.5
12	0.039	0.1168	1
13	0.02	0.0599	0.5
14	0.02	0.0599	0.5
15	0.098	0.2952	1
16	0.116	0.3494	1

TABLE 2

VI. DECISION TABLE AND FLOW GRAPH

With every decision table, a flow graph can be associated which is defined as follows:

To every decision rule $C \xrightarrow[x]{} D$, we assign a directed branch x- connecting the input node C(x) and the output node D(x).

The strength of the decision rule represents a through flow of the corresponding branch. The through flow of the graph is governed by formula (1) to (6).

Formula (1) and (2) say that the outflow of an input node or an output node is equal to their respective in flows. Formula (3) states that the outflow of the output node amount to the sum of its inflows: Whereas formula (4) says that the sum of the outflow of the input node equal its inflow.

Finally, formula (5), (6) reveal how through flow in the flow graph is distributed between its input and outputs.

The flow graph associated with the decision table presented in table 2 is shown in Fig 1

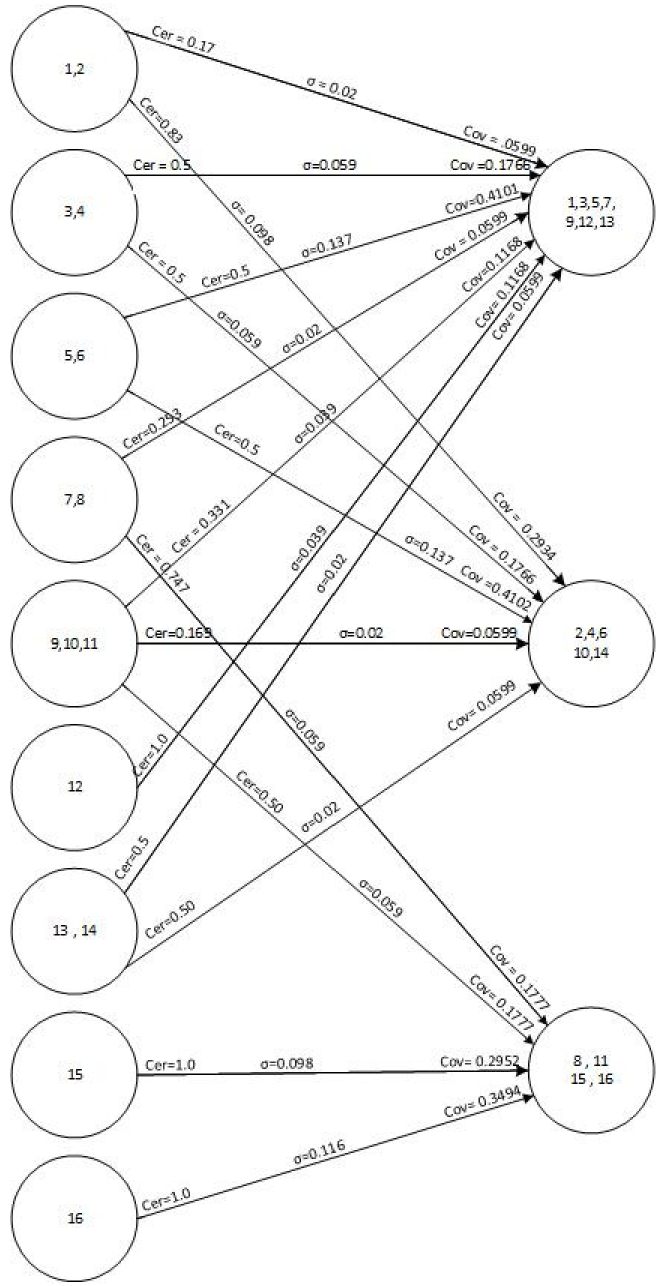

Fig. 1

The use of flow graphs to represent decision tables gives very clear insight into the decision process

VII. CONCLUSION

We have applied Rough Set Theory on data through selected to serve about some brands that should be preferred by new customer. Our method gives best and worst brand among all the available brands. So this method will helpful for business analysis. This theory can be improve by applying same theory on standard big data set.

REFERENCES

[1] Abraham A., Bello R. and Falcon R., "Rough Set Theory: A True Landmark in Data Analysis", Studies in Computational Intelligence, Vol. 174, 1-324.

[2] Rissino S. and Torres G. L., "Rough Set Theory – Fundamental Concepts, Principals, Data Extraction, and Applications", Julio Ponce and Adem Karahoca, ISBN, 35-58.

[3] Zbigniew S., "An Introduction to Rough Set Theory and Its Applications", ICENCO, 27- 30.

[4] Pawlak Z., "Rough sets", International Journal of Computer and Information Sciences, 11,341- 356.

[5] Biswas, R., and S. Nanda, "Rough groups and rough subgroups", Bulletin of the Polish Academy of Sciences Mathematics, vol.42, no.3, pp.251-254.

[6] Pawlak, Z., "Rough sets and fuzzy sets", Fuzzy Sests and Systems, vol.17, pp.99-102.

[7] Pawlak, Z., "Hard and soft sets", Fuzzy Sets and Knowledge-Discovery, pp.130-135.

[8] Pawlak, Z., and R. Slowinski, "Rough set approach to multi–attribute decision analysis", European Journal of Operational Research, vol.72, pp.443-449.

[9] Pawlak, Z, "Rough sets and intelligent data analysis", Information Sciences, vol.147, pp.1-12.