Rough Set Theory Approach to Generating Classification Rules

K Anitha

Department of Mathematics
S A Engineering College
Chennai, India
subramanianitha@gmail.com

P Venkatesan

Department of Statistics.
Indian Council for Medical Research
Chennai, India

Abstract- **Data mining is the process of extracting hidden information from large databases. Rule induction is the most common data mining technique. Dimensionality reduction gives an optimal subset according to an objective function. Rule generation is one of the important processes in the knowledge discovery system. Rough set approach of generating rules can be used to increase the correct predictions by identifying and removing redundant variables. In this paper we emphasize the role of Reducts, Core and their approximations. Data from UCI repository have been taken to exhibit rules for soybean data set by using ROSETTA.**

Keywords- Rough Sets, Reducts, Core, Upper and Lower Approximations, Dimensionality Reduction

I. INTRODUCTION

Feature Selection is the process of identifying the most important/useful feature from a given data set which is encountered in many fields such as Machine Learning, Image and Signal Processing and Pattern recognition. Feature Selection Process should preserve the exact content after reduction. Rough Set theory plays vital role in dimensionality reduction. The main advantage of the Rough set theory is it requires no additional information other than the supplied data. This paper briefly explains the terminologies of Rough sets and importance of rules in section II. Application of rules have been implemented for soybean data set and their results are exhibited in section III. Section IV describes conclusion of this paper.

II. ROUGH SET TERMINOLOGIES

A. Information System

An Information System is a pair $IS = (U, A)$ where U is a non-empty finite set of objects called Universal set and A is a non-empty finite set of attributes such that $a: U \rightarrow V_a$ for every $a \in A$. The set V_a is called the value set of a.

B. Indiscernibility

For any $P \subseteq A$ there is an equivalence relation $IND(P)$ which is defined as

$$IND(P) = \{(x,y) \in U^2 / \forall \, a\epsilon P \, , \, a(x) = a(y)\} \tag{1}$$

Which represents two objects are equivalent if and only if their attribute values are same in P.

C. Upper Approximation And Lower Approximation

Let $X \subseteq U$, X can be approximated using only the information contained within P by constructing the Lower and Upper approximation of P. Equivalence classes contained within X belong to the Lower Approximation ($\underline{P}X$). Equivalence classes within X and along its border form the Upper Approximation ($\overline{P}X$). They are expressed as

$$\underline{P}X = \{ \, x / [x]_P \subseteq X\} \tag{2}$$

$$\overline{P}X = \{ \, x / [x]_P \cap X \neq \emptyset\} \tag{3}$$

The Upper and Lower Approximations have the following properties:

(i) $\underline{P}X \subseteq \overline{P}X$
(ii) $\underline{P}(\emptyset) = \overline{P}(\emptyset) = \emptyset$, $\underline{P}(U) = \overline{P}(U) = U$
(iii) $\overline{\overline{P}}(XUY) = \overline{P}(X)U\overline{P}(Y)$, $\underline{P}(X \cap Y) = \underline{P}(X) \cap \underline{P}(Y)$

Positive , Negative , Boundary Regions And Feature Dependency

Let P and Q be sets of attributes inducing equivalence relations over U, then the positive, negative and boundary regions are defined as

$$POSP(Q) = \bigcup_{X \epsilon U/Q} \underline{P}\ X \tag{4}$$

$$NEGP(Q)\ =\ U - \bigcup_{X \epsilon U/Q} \overline{P}\ X \tag{5}$$

$$BNDP(Q)\ =\ \bigcup_{X \epsilon U/Q} \overline{P}\ X - \bigcup_{X \epsilon U/Q} \underline{P}\ X \tag{6}$$

The important step in data analysis is finding dependencies between attributes. Rough set dependency is defined by,

For $P, Q \subset A$, it is said that Q depends on P in a degree k where $0 \le k \le 1$ denoted by $P \Rightarrow_k Q$ if

$$k = \gamma P(Q)\ = \frac{\left|POS_{P(Q)}\right|}{|U|} \tag{7}$$

where $|S|$ is the cardinality of set S.

If k = 1, Q depends totally on P, and if $0 < k < 1$, Q depends partially on P , and if k $= 0$ then Q does not depend on P. For calculating the change of dependency it is necessary to estimate the significance value of particular feature that can be removed. The higher the change in dependency the more significant is the feature. If the significance value is zero then the feature is dispensible [12]. The formula for finding the significance value is given by the following relation,

$$\sigma_P(Q,a) = \gamma_P(Q)\ - \gamma_{P-\{a\}}(Q) \tag{8}$$

D. Reduct

Reduct is the minimal representation of original data set and defined as a minimal subset R of the initial attribute set C such that for a given set of attributes D,

$$\gamma_R(D)\ = \gamma_c(D) \tag{9}$$

That is R is a minimal subset if

$$\gamma_{R-\{a\}}(D)\ \ne \gamma_c(D)\ \text{ for all } a \epsilon R \tag{10}$$

Which means that no attributes can be removed from the subset without affecting dependency degree. A given data set may have many reduct sets, and the collection of all reducts is denoted by

$$R_{ALL} = \{\ X/X \subseteq C\ , \gamma_R(D) = \gamma_c(D)\ \text{ and }\ \gamma_{R-\{a\}}(D)\ \ne \gamma_c(D)\ \text{ for all } a \epsilon X\} \tag{11}$$

The intersection of all the sets in R_{ALL} is called Core. The condition for locate a single element of the reduct set is $R_{MIN} \subseteq R_{ALL}$. A Reduct of a decision table is a set of condition attributes that is sufficient to define the decision attributes. A reduct does not contain redundant attributes towards a classification task. It is often used in the attribute selection process to reduce the redundant attributes, and to reduce the computation cost for rule generations [1].

E. Generation Of Rules On Rough Set Thoery

In Knowledge Discovery process various rules such as Association Rules, Sequential Patterns, Classification Rules and so on are representing different forms.

An Association rule is a rule of the form $\alpha \to \beta$, where $\alpha\ and\ \beta$ represent item sets which do not share common items [5]. The Association rule $\alpha \to \beta$ holds in the transaction set D with $Confidence\ C$ if $C\%$ of transactions in D that contain α and also contain β. $Confidence\ C\ \&\ Support\ S$ can be defined by the following

$$C = \frac{|\alpha \cup \beta|}{|\alpha|}\ \&\ \ S = \frac{|\alpha \cup \beta|}{|D|} \tag{12}$$

F. Generation Of Rules From Reducts

Reducts are framed from original data set can be used to represent original knowledge. After finding the reduct rules are generated in the form such that the antecedents of a rule are from the value of condition attributes in the reduct set and the consequents of a rule are from the value of decision attributes from the original data set. A reduct contains the most representative and important condition attributes of a decision table, based on this assumption each of the rules are generated .

Let $T = (U, C, D)$ be the Decision Table, where $U = \{a_0, a_1, \dots a_{m-1}\}$ is a set of data and $C = \{c_0, c_1, \dots c_{p-1}\}$ is the conditional attributes and D is a set of Decision attributes. Let us assume that the decision table contains one decision attributes.

Let R be the set of rulels generated from $T = \{Rule_0, Rule_1, \dots Rule_{n-1}\}$. Then the new decision table $A_{m \times (n+1)}$ where each data from original decision table $a_0, a_1, \dots a_{m-1}$ is the row, $Rule_0, Rule_1, \dots Rule_{n-1}$ and decision attribute are columns of new decision table.

For each $Rule_j (j \in [0, \dots n-1])$, we assign 1 to the cell $A[i, j](i \in [0, \dots m-1])$, if the $Rule_j$ can be applied to the record a_i, otherwise set Zero. The Decision Attribute $A[i, j](i \in [0, \dots m-1])$, remains the same as the original values of the decision attribute in the original decision table. The following equation   defines the condition for the value of the new decision table [2].

$$A[i, j] = \begin{cases} 1 & if\ j < n\ and\ Rule_j\ can\ be\ applied\ to\ a_i \\ 0 & if\ j < n\ and\ Rule_j\ can\ not\ be\ applied\ to\ a_i \\ d_i & if\ j = n\ and\ d_i\ is\ the\ corresponding \\ & decision\ attrubute\ for\ a_i \end{cases} \qquad (13)$$

Where $i \in [0, \dots m-1], j \in [0, \dots n-1]$.

G. Tool Kit For Rough Set Knowledge Discovery Process

Rosetta (A Rough Set Toolkit for Analysis of Data) : Which supports the complete data mining process from data preprocessing, including handling incomplete data, data discretization, generating reducts which contain essential attributes for the given data set to classification , rule generation and cross validation evaluation.

III. APPLICATION

A. Rule generation for soybean data set.

In this experiment Soybean (Large) data set which is taken from UCI- Repository to demonstrate Rule Generation. There are 19 classes (diaporthe-stem-canker , charcoal-rot, rhizoctonia-root-rot, phytophthora-rot, brown-stem-rot, powdery-mildew… herbicide-injury) with 36 attributes (date, temp , plant-growth, leaves, stem-cankers,…,mycelium,..). ROSETTA tool kit is used for finding rules and Reducts. The decision table consists of 305 objects.

By Applying Johnson's Reduction Algorithm 126 deterministic rules (Partition-I = 48, Partition-II= 78) are generated which are given in TABLE I and TABLE II. First the objects are splitted into two partitions with 153 objects as PARTITION –I and 152 as PARTITION-II. For PARTITION –I, 48 rules and for PARTITION-II, 78 rules are generated.

TABLE I.   PARTITION –I

RULES	LHS SUPPORT	RHS SUPPORT	RHS ACCUR.	LHS COVER.	RHS COVER.	RHS STABLITY	LHS LENG.	RHS LENG.
diaporthe-stem-canker(diaporthe-stem-canker) AND 0(0) AND 1(2) => 0(0)	2	2	1	0.013072	0.015038	1	3	1
diaporthe-stem-canker(diaporthe-stem-canker) AND 0(0) AND 1(1) => 0(0)	4	4	1	0.026144	0.030075	1	3	1
diaporthe-stem-canker(charcoal-rot) AND 0(0) AND 1(1) => 0(0)	6	6	1	0.039216	0.045113	1	3	1

diaporthe-stem-canker(rhizoctonia-root-rot) AND 0(1) AND 1(2) => 0(0)	5	5	1	0.03268	0.037594	1	3	1
diaporthe-stem-canker(rhizoctonia-root-rot) AND 0(0) AND 1(2) => 0(0)	1	1	1	0.006536	0.007519	1	3	1
diaporthe-stem-canker(rhizoctonia-root-rot) AND 0(1) AND 1(1) => 0(0)	1	1	1	0.006536	0.007519	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 0(1) AND 1(1) => 0(0)	2	2	1	0.013072	0.015038	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 0(1) AND 1(?) => 0(1)	10	10	1	0.065359	0.714286	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 0(1) AND 1(2) => 0(0)	4	4	1	0.026144	0.030075	1	3	1
diaporthe-stem-canker(brown-stem-rot) AND 0(0) AND 1(1) => 0(0)	6	6	1	0.039216	0.045113	1	3	1
diaporthe-stem-canker(brown-stem-rot) AND 0(1) AND 1(1) => 0(0)	3	3	1	0.019608	0.022556	1	3	1
diaporthe-stem-canker(powdery-mildew) AND 0(0) AND 1(1) => 0(0)	1	1	1	0.006536	0.007519	1	3	1
diaporthe-stem-canker(powdery-mildew) AND 0(0) AND 1(0) => 0(0)	1	1	1	0.006536	0.007519	1	3	1
diaporthe-stem-canker(powdery-mildew) AND 0(1) AND 1(0) => 0(0)	2	2	1	0.013072	0.015038	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 6(1) AND 0(0) => 0(0)	1	1	1	0.006579	0.008	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 6(2) AND 0(?) => 0(1)	6	6	1	0.039474	0.315789	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 6(3) AND 0(0) => 0(0)	1	1	1	0.006579	0.008	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 6(2) AND 0(0) => 0(0)	2	2	1	0.013158	0.016	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 6(1) AND 0(?) => 0(1)	3	3	1	0.019737	0.157895	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 6(0) AND 0(0) => 0(0)	3	3	1	0.019737	0.024	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 6(0) AND 0(1) => 0(0)	1	1	1	0.006579	0.008	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 6(2) AND 0(1) => 0(0)	1	1	1	0.006579	0.008	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 6(1) AND 0(0) => 0(0)	1	1	1	0.006579	0.008	1	3	1
……………………………	…	….	….	…….	…….	…..	…	….
…………..	….	…..	….	…..	…….	…….	…..	…..
……………………………	…	….	….	…….	…….	…..	…	….
diaporthe-stem-canker(phytophthora-rot) AND 6(3) AND 0(?) => 0(1)	5	5	1	0.032895	0.263158	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 6(1) AND 0(1) => 0(0)	1	1	1	0.006579	0.008	1	3	1
diaporthe-stem-canker(brown-stem-rot) AND 6(4) AND 0(0) => 0(0)	5	5	1	0.032895	0.04	1	3	1

TABLE-II Partition -II

RULES	LHS SUPPORT	RHS SUPPORT	RHS ACCUR.	LHS COVER.	RHS COVER.	RHS STABLITY	LHS LENG.	RHS LENG.
diaporthe-stem-canker(diaporthe-stem-canker) AND 0(0) AND 1(2) => 0(0)	2	2	1	0.013072	0.015038	1	3	1
diaporthe-stem-canker(diaporthe-stem-canker) AND 0(0) AND 1(1) => 0(0)	4	4	1	0.026144	0.030075	1	3	1
diaporthe-stem-canker(charcoal-rot) AND 0(0) AND 1(1) => 0(0)	6	6	1	0.039216	0.045113	1	3	1
diaporthe-stem-canker(rhizoctonia-root-rot) AND 0(1) AND 1(2) => 0(0)	5	5	1	0.03268	0.037594	1	3	1
diaporthe-stem-canker(rhizoctonia-root-rot) AND 0(0) AND 1(2) => 0(0)	1	1	1	0.006536	0.007519	1	3	1
diaporthe-stem-canker(rhizoctonia-root-rot) AND 0(1) AND 1(1) => 0(0)	1	1	1	0.006536	0.007519	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 0(1) AND 1(1) => 0(0)	2	2	1	0.013072	0.015038	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 0(1) AND 1(?) => 0(1)	10	10	1	0.065359	0.714286	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 0(1) AND 1(2) => 0(0)	4	4	1	0.026144	0.030075	1	3	1
diaporthe-stem-canker(brown-stem-rot) AND 0(0) AND 1(1) => 0(0)	6	6	1	0.039216	0.045113	1	3	1
diaporthe-stem-canker(brown-stem-rot) AND 0(1) AND 1(1) => 0(0)	3	3	1	0.019608	0.022556	1	3	1
diaporthe-stem-canker(powdery-mildew) AND 0(0) AND 1(1) => 0(0)	1	1	1	0.006536	0.007519	1	3	1
diaporthe-stem-canker(powdery-mildew) AND 0(0) AND 1(0) => 0(0)	1	1	1	0.006536	0.007519	1	3	1
diaporthe-stem-canker(powdery-mildew) AND 0(1) AND 1(0) => 0(0)	2	2	1	0.013072	0.015038	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 6(1) AND 0(0) => 0(0)	1	1	1	0.006579	0.008	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 6(2) AND 0(?) => 0(1)	6	6	1	0.039474	0.315789	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 6(3) AND 0(0) => 0(0)	1	1	1	0.006579	0.008	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 6(2) AND 0(0) => 0(0)	2	2	1	0.013158	0.016	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 6(1) AND 0(?) => 0(1)	3	3	1	0.019737	0.157895	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 6(0) AND 0(0) => 0(0)	3	3	1	0.019737	0.024	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 6(0) AND 0(1) => 0(0)	1	1	1	0.006579	0.008	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 6(2) AND 0(1) => 0(0)	1	1	1	0.006579	0.008	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 6(1) AND 0(0) => 0(0)	1	1	1	0.006579	0.008	1	3	1
………………………………	...	....	....	......	......	.....	...	....
………….	....	......	....	.....	......	......	.....	.....
………………………………	...	....	....	......	......	.....	...	....
diaporthe-stem-canker(phytophthora-rot) AND 6(3) AND 0(?) => 0(1)	5	5	1	0.032895	0.263158	1	3	1
diaporthe-stem-canker(phytophthora-rot) AND 6(1) AND 0(1) => 0(0)	1	1	1	0.006579	0.008	1	3	1
diaporthe-stem-canker(brown-stem-rot) AND 6(4) AND 0(0) => 0(0)	5	5	1	0.032895	0.04	1	3	1

Totally 2 Reducts are generated for the entire data set Their Support, Accuracy and Mean values are clearly explained in TABLE-III.

TABLE-III. CENTRAL MEASURES FOR SOYBEAN DATA

Partition	Support	Mean	Std.Deviation	Median	Min	Max
153	RHS (THEN)	3.1875	2.590131	2.0	1	10
	LHS(IF)	3.00	0.00	3.00	3	3
152	RHS (THEN)	2.02667	1.739628	1.0	1	10
	LHS(IF)	3.00	0.00	3.00	3	3

From the above 126 deterministic rules finally two Reducts are obtained for the entire data set and the corresponding Core are listed in the Tab.IV.

TABLE.IV- REDUCT AND CORE

S.NO	REDUCT	SUPPORT	LENGTH
1.(153)	{diaporthe-stem-canker, 0, 1}	100	3
2.(152)	{diaporthe-stem-canker, 6, 1, 0}	100	4
Core : {diaporthe-stem-canker, 0, 1}			

Upper and Lower Approximation : All objects which surely belong to the set characterize the lower approximation and all objects which possibly belong to the set characterize the upper approximation. The Upper and Lower Approximations with boundary region are given in TABLE.V and TABLE.VI with 94% accuracy for PARTITION-I , 97% accuracy for PARTITION -II

TABLE V. UPPER AND LOWER APPROXIMATIONS FOR PARTITION-I (153 Objects)

	Number of equivalence classes.	Total cardinality	CARDINALITY RATIO
Universe	19	153	
Upp.App	14	130	UPPER APPROXIMATION=1.1769
Low.APP	14	130	LOWER APPROXIMATION=1.1769
Boundary region	0	0	**PERFORMANCE ACCURACY=0.94**
Outside region	5	23	

TABLE VI. UPPER AND LOWER APPROXIMATIONS FOR PARTITION-II (152 OBJECTS)

	Number of equivalence classes.	Total cardinality	CARDINALITY RATIO
Universe	18	152	
Upp.App	13	114	UPPER APPROXIMATION=1.1333
Low.APP	13	114	LOWER APPROXIMATION=1.1333
Boundary region	0	0	**PERFORMANCE ACCURACY=0.972**
Outside region	5	38	

IV. CONCLUSION

In this paper rule generation using Rough Set is proposed and which is implemented for Soybean (Large) data set using ROSETTA software. From these set of rules Reducts, Upper and Lower Approximations have been evaluated. Reducts are minimal representation of the original dataset, and this minimal representation is useful in classification analysis for the entire dataset. Computation of Reducts and Core using Rough set theory in data mining minimizes the computational cost in mining process, reduce the noisy information and extract decision rules which will contribute to select relevant and consistent features of the large dataset and hence this is useful to the decision makers in taking necessary steps in all fields like Bio-Informatics and financial sectors.

REFERENCE

[1] Arauzo-Azofra, A.,Aznarte,J.L. and Benitez,J.M., Empirical Study of Feature Selection Methods based on Individual Feature evaluation for classification problems, Expert systems with applications, 38: 8170-8177, 2011.

[2] Dash,M. and Liu,H. Feature Selection for Classification, International Journal of Data Analysis,1: 131-156, 1997.

[3] Eschrich,S., Ke,J., Hall,L. and Goldgof,D. Fast accurate fuzzy clustering through data Reduction, Transactions on Fuzzy systems,11: 262-270, 2003.

[4] Fazayeli,F., Wang,L. and Mandziuk,J., Feature selection based on Rough set theory and Expectation-Maximization clustering algorithm,Springer, LNAI 5306:272-282, 2008.

[5] Gray,B., and Orlowska,M.E., Clustering categorical attributes into interesting association rules, Proceedings of the Second Pacific-Asia Conference on Knowledge Discovery and Data Mining, 132-143, 1998.

[6] Hu,X., Lin,T.Y. and Han,J., A new Rough sets model based on Data Base Systems, Fundam. Inf., 59:135–152, 2004.

[7] Kaneiwa,K. and Kudo,Y. ,X A sequential pattern mining algorithm using Rough set theory,52:881-893, 2004.

[8] Komorowski,J., Pawlak,Z., Polkowski,L. and Skowron,A., RoughSets : A Tutorial, 1999.

[9] Mak,B. and Munakata,T. , Rule extraction from expert heuristics: A comparative study of rough sets with neural networks and ID3. Artificial Intelligence and Information. 136:212-229, 2002.

[10] Rissino,S. and Torres,G.L., Rough Set Theory- Fundamental Concepts, Principals, Data Extraction, and Applications. Data Mining and Knowledge Discovery in real life applications,10:1-58, 2009.

[11] Thangavel,K., Peethalakshmi,A. and Jaganathan,P. , A Novel Reduct Algorithm for Dimensionality Reduction with Missing Values Based on Rough Set Theory., International Journal of Soft Computing,1:111-117, 2006.

[12] Vashist,R. and Garg,M.L., Rule generation based on Reduct and Core, A Rough Set Approach, 29:3595-4989, 2011.