RESEARCHERID
THOMSON REUTERS

[Surekha S, 6(4): April, 2017]
IC™ Value: 3.00

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A COMPARATIVE STUDY OF VARIOUS ROUGH SET THEORY ALGORITHMS FOR FEATURE SELECTION

S.Surekha *

* Department of CSE, JNTUK-University College of Engineering Vizianagaram
Andhra Pradesh, India

DOI: 10.5281/zenodo.495147

ABSTRACT

Machine Learning techniques can be used to improve the performance of intelligent software systems. The performance of any Machine Learning algorithm mainly depends on the quality and relevance of the training data. But, in real world the data is noisy, uncertain and often characterized by a number of features. Existence of uncertainties and the presence of irrelevant features in the high dimensional datasets often degrade the performance of the machine learning algorithms in all aspects. In this paper, the concepts of Rough Set Theory(RST) are applied to remove inconsistencies in data and various RST based algorithms are applied to identify the most prominent features in the data. The performance of the RST based reduct generation algorithms are compared by submitting the feature subsets to the RST based decision tree classification. Experiments were conducted on some of the medical datasets taken from Irvine UCI machine learning repository.

KEYWORDS: Machine Learning, Uncertain Data, Rough Set Theory, Reduct Generation Algorithms.

INTRODUCTION

Now a days, Machine Learning[1] are being applied successfully to improve the performance of many intelligent systems like Weather forecasting, Face detection, Image Classification, Disease diagnosis, Speech recognition, Signal denoising e.t.c., Machine Learning techniques help in developing an efficient intelligent system without much intervention of humans. Decision Tree Classification[2] is widely used in machine learning for classification. The primary factor that affects the performance of any machine learning algorithm is the quality of training data. Before modeling a classification technique, the quality of the data must be checked. Secondly, the dimensionality of the training data also affects the computational complexity of the machine learning algorithm. As data is characterized by many features and not all these features contribute for a particular task and hence there is a great demand to identify the features or attributes that are relevant for a particular task to reduce the feature space so as to reduce the computational complexity of the learning model. In this paper, the concepts of the most popular Rough Set Theory(RST)[3] is applied for inconsistent removal, feature subset selection and also to induce decision tree. At first, the basic concepts of RST are used to identify and eliminate inconsistencies in the data and then various versions of RST based Quick Reduct algorithm[4] are applied to know the most relevant set of features in the training data. And to compare the effectiveness of the RST based reduct generation algorithms, the generated feature subsets are submitted to the RST based decision tree classification and the obtained prediction accuracies are compared.

BASIC CONCEPTS OF ROUGH SET THEORY

Rough Set Theory[3] is mathematical theory developed by Z.Pawlak. The detailed explanation of the RST concepts can be found in the literature[3-7] and are given below.

Information System

The Universal facts are represented as an Information System(IS)[6] and is denoted as IS = (U,A), where U is the universe of facts and A is the set of features used to characterize the facts represented by U.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

Equivalence Classes

For any set of attributes B⊆ A, the set of equivalence classes[6] generated by B is denoted as U/IND(B) or U/B or $[X]_B$ and is defined as,

$$U/IND(B) = \otimes\{U/IND(\{b\}) \mid b \in B\} \quad (1)$$

Let R and S be two non empty finite sets then operation ⊗ is defined as,

$$R \otimes S = \{X \cap Y \mid X \in R, Y \in S, X \cap Y \neq \emptyset\} \quad (2)$$

Set Approximations

For a target set of objects O⊆U and for any set of attributes B⊆ A, the B-Lower approximation[6] is the set of objects that are unavoidably belongs to the target set of interest and the equation for B-Lower approximation is given by,

$$\underline{B}\,O = \{o \mid [o]_B \subseteq O\} \quad (3)$$

The B-Upper approximation[6] of set O is the set objects that possibly belong to the target set O. The equations for B-Upper approximations for B is given by,

$$\bar{B}O = \{o \mid [o]_B \cap O \neq \Phi\} \quad (4)$$

Boundary Region

The B-Boundary region of set O is the set of objects that can neither belongs to O nor does not ruled out from O and can be obtained by,

$$BND_B(O) = \bar{B}O - \underline{B}O \quad (5)$$

The objects that falls in boundary region are the inconsistent objects.

Explicit Region

The Explicit Region of an attribute set P with respect to a set of attribute Q is defined as,

$$Exp - Region(P, Q) = \cup_{p_i \in [Q]_B} \underline{P}(Q) \quad (6)$$

Where, $\underline{P}(Q)$ is the P-lower approximation with respect to Q.

Dependency of Attributes

Let P and Q be two subsets of the attribute set A of the Information System. Now, the dependency of attribute P on Q is defined by,

$$\gamma(P, Q) = \frac{|Exp-Region(P,Q)|}{|U|} \quad (7)$$

Where, |U| is the cardinality of the set U.

Worked Out Example

The sample dataset in Table 1 is consisting of four conditional attributes {P, Q, R, and S} and a decision attribute {D}.

Table 1. Sample DataSet with Inconsistenceis

U	P	Q	R	S	D
1	p1	q2	r1	s3	d2
2	p3	q3	r2	s3	***d2***
3	p3	q3	r2	s3	***d1***
4	p2	q1	r1	s2	d1
5	p2	q2	r1	s1	d2
6	p3	q2	r3	s1	d3

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

7	p2	q2	r3	s1	d3
8	p1	q1	r2	s2	d1
9	p3	q3	r3	s1	d3
10	p1	q1	r1	s1	d1

For the data represented in Table 1, the set of conditional attributes are C={P,Q,R,S} and the decision attribute is D with three decision classes d1,d2, and d3. The equivalence classes generated by Eq.(1) and Eq.(2) is given below,

U/IND(C)= (U/P) ⊗ (U/Q) ⊗ (U/R) ⊗ (U/S)
U/P = { {1,8,10}, {4,5,7}, {2,3,6,9} }
U/Q = { {4,8,10},{1,5,6,7},{2,3,9}}
U/R = { {1,4,5,10},{2,3,8},{6,7,9}}
U/S={{1,2,3},{4,8},{5,6,7,9,10}}
U/D={{3,4,8,10},{1,2,5},{6,7,9}}
Now, (U/P) ⊗ (U/Q) = { {1,8,10}, {4,5,7}, {2,3,6,9} }⊗ { {4,8,10},{1,5,6,7},{2,3,9}}
={{1,8,10} ∩ {4,8,10},{1,8,10} ∩ {1,5,6,7},{1,8,10} ∩ {2,3,9},{4,5,7}, ∩ {4,8,10}, {4,5,7}, ∩ {1,5,6,7},{4,5,7}, ∩ {2,3,9}, {2,3,6,9} ∩ {4,8,10},{2,3,6,9} ∩ {1,5,6,7}, {2,3,6,9} ∩ {2,3,9}}
= { {8,10}, {1},{4},{5,7},{6},{2,3,9} }

Similarly,
U/IND(C)= (U/P) ⊗ (U/Q) ⊗ (U/R) ⊗ (U/S) = {1},{2,3},{4},{5},{6},{7},{8},{9},{10}}
C-Lower Approximation of D can be calculated using Eq.(3) calculated as follows,
U/C= { {1},{2,3},{4},{5},{6},{7},{8},{9},{10}} and U/D={{1,2,5}, {3,4,8,10},{6,7,9}}
Let the target set D is consisting of 3 subsets
D_1= {1,2,5}, D_2={3,4,8,10} and D_3={6,7,9}
$\underline{C}D_1$ = { {1}, {5}}
$\underline{C}D_2$ = { {4}, {8}, {10} } and $\underline{C}D_3$ = { {6}, {7},{9} }
C-Upper Approximation of D is calculated using Eq.(4) and is obtained as,
$\overline{C}D_1$= { {1}, {2,3},{5} }, $\overline{C}D_2$= { {2,3},{4},{8},{10} }, $\overline{C}D_3$= { {6},{7},{9} }
The existence of inconsistencies in the dataset can be known by the Boundary Region, which can be calculated using Eq.(5).
A-Boundary Region of D = { {1},{2,3},{4},{5},{6},{7},{8},{9},{10} }-{ {1}{4},{5},{6},{7},{8},{ 9},{10}}
= **{2,3}**

After removing the inconsistencies in the data of Table 1, the number of instances will be reduced to eight and the consistent data is given in Table 2.

Table 2. Consistent Sample DataSet

U	P	Q	R	S	D
1	p1	q2	r1	s3	d2
2	p2	q1	r1	s2	d1
3	p2	q2	r1	s1	d2
4	p3	q2	r3	s1	d3
5	p2	q2	r3	s1	d3
6	p1	q1	r2	s2	d1
7	p3	q3	r3	s1	d3
8	p1	q1	r1	s1	d1

ROUGH SET THEORY BASED REDUCT GENERATION ALGORITHMS

Reduct[4] is a minimal subset of features that are essential and sufficient for categorization of objects in the universe. The popular RST based Reduct generation algorithm is the QuickReduct algorithm. It starts the reduct computation process with an empty reduct set and recursively adds attributes one after one that result in the

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

greatest increase in the rough set dependency, until a maximum possible value has been produced. Based on the criteria of adding features, there are three variants of Quick reduct algorithm; they are QuickReduct-Forward, QuickReduct-Backward and Improved QuickReduct generation.

In QuickReduct-Forward algorithm, the reduct generation algorithm starts from the first feature and then successively selects the next feature in order and checks for any improvement in the metric i.e., degree of dependency. In QuickReduct-Backward algorithm, the algorithm starts from the last feature in the feature set and successively adds the features from the last one and similarly observes for improvement and iteratively repeats the process and terminates when the degree of dependency of the reduct set is equals to the degree of dependency of the full conditional attribute set. The drawback with these two algorithms is, that whenever the stopping criteria mets the algorithm terminates by not examining all features. So, the improved version of the basic QuickReduct algorithm is the Improved QuickReduct algorithm, which examines the remaining attributes and from this it selects the one with maximum improvement in the metric.

The following example clearly explains the variants of RST based Reduct generation algorithms for the sample dataset represented in Table 2.

The equivalence classes for the given set of conditional and decision attributes are obtained as follows,

U/IND(C) = (U/P) ⊗ (U/Q) ⊗ (U/R) ⊗ (U/S)

U/P = { {1,6,8}, {2,3,5}, {4,7} }

U/Q = { {2,6,8},{1,3,4,5},{7}}

U/R = { {1,2,3,8},{6},{4,5,7}}

U/S = {{1},{2,6},{3,4,5,7,8}}

U/C = { {1},{2},{3},{4},{5},{6},{7},{8}}

U/D = {{2,6,8},{1,3},{4,5,7}}

The degree of dependency of the attributes can be calculated using Eq.(6) & (7) and the dependencies are obtained as follows,

$$\gamma(C,D) = \frac{|Exp-Region(C,D)|}{|U|} = \frac{|\underline{C}D_1 \cup \underline{C}D_2 \cup \underline{C}D_3|}{|U|} = \frac{|\{2,6,8\} \cup \{1,3\} \cup \{4,5,7\}|}{8} = 1$$

$$\gamma(P,D) = \frac{|Exp-Region(P,D)|}{|U|} = \frac{|\underline{P}D_1 \cup \underline{P}D_2 \cup \underline{P}D_3|}{|U|} = \frac{|\{4,7\}|}{8} = \frac{2}{8} = 0.25$$

$$\gamma(Q,D) = \frac{|Exp-Region(Q,D)|}{|U|} = \frac{|\underline{Q}D_1 \cup \underline{Q}D_2 \cup \underline{Q}D_3|}{|U|} = \frac{|\{2,6,8\} \cup \{7\}|}{8} = \frac{4}{8} = 0.5$$

$$\gamma(R,D) = \frac{|Exp-Region(R,D)|}{|U|} = \frac{|\underline{R}D_1 \cup \underline{R}D_2 \cup \underline{R}D_3|}{|U|} = \frac{|\{4,5,7\} \cup \{6\}|}{8} = \frac{4}{8} = 0.5$$

$$\gamma(S,D) = \frac{|Exp-Region(S,D)|}{|U|} = \frac{|\underline{S}D_1 \cup \underline{S}D_2 \cup \underline{S}D_3|}{|U|} = \frac{|\{2,6\} \cup \{1\}|}{8} = \frac{3}{8} = 0.375$$

QuickReduct-Forward algorithm

Initially, Reduct = Φ and $\gamma(Reduct, D) = 0$

Select the first feature i.e., P and add P to the Reduct, then check the degree of dependency of the reduct.

Reduct = Reduct ∪ {P} = {P} and $\gamma(Reduct, D) = 0.25$

There is an improvement in the dependency after adding the attribute P. Now, add the second attribute Q to the reduct and continue the process until the degree of dependency of the Reduct equals to the dependency of all conditional attributes.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

Reduct = Reduct ∪ {Q} = {P,Q} and $\gamma(Reduct, D) = 0.75$

Reduct = Reduct ∪ {R} = {P,Q,R} and $\gamma(Reduct, D) = 1$

The degree of dependency of the attributes {P,Q,R} is equals to the degree of dependency of the full set of attributes. So, terminate the Reduct is {P,Q,R}.

QuickReduct-Backward algorithm

Initially, Reduct = Φ and $\gamma(Reduct, D) = 0$

Select the first feature i.e., S and add S to the Reduct, then check the degree of dependency of the reduct.

Reduct = Reduct ∪ {S} = {S} and $\gamma(Reduct, D) = 0.375$

There is an improvement in the dependency after adding the attribute S. Now, add the second attribute R to the reduct and continue the process until the degree of dependency of the Reduct equals to the dependency of all conditional attributes.

Reduct = Reduct ∪ {R} = {R,S} and $\gamma(Reduct, D) = 0.75$

Reduct = Reduct ∪ {Q} = {Q,R,S} and $\gamma(Reduct, D) = 1$

The degree of dependency of the attributes {Q,R,S} is equals to the degree of dependency of the full set of attributes. So, terminate the Reduct is {Q,R,S}.

Improved-QuickReduct algorithm

Initially, Reduct=Φ and $\gamma(Reduct, D) = 0$

From the above, the attributes with maximum dependency degree are Q and R. So, first select Q then $\gamma(Reduct, D) = 0.5$.

Add the attribute R to the Reduct and $\gamma(Reduct, D) = 1$

The degree of dependency of the attributes {Q,R} is equals to the degree of dependency of the full set of attributes. So, terminate with the Reduct as {Q,R}.

RST BASED DECISION TREE CLASSIFICATION

In RST based decision tree classification [8], the attribute with highest size of Explicit Region is selected as the splitting attribute. The following steps illustrate the RST based decision tree induction process for the data in Table 2. The Explicit Regions for all the conditional attributes can be calculated using Eq.(6) and are obtained as follows,

$Exp - Region(P, D)$ = {4,7} and $|Exp - Region(P, D)| = 2$

$Exp - Region(Q, D)$ = {2,6,7,8} and $|Exp - Region(Q, D)| = 4$

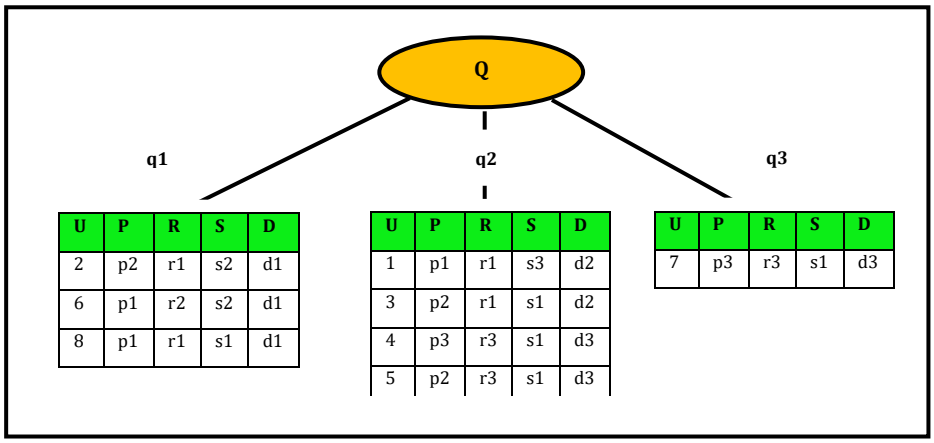
$Exp - Region(R, D)$ = {4,5,6,7} and $|Exp - Region(R, D)| = 4$

$Exp - Region(S, D)$ = {1,2,6} and $|Exp - Region(S, D)| = 3$

Two attributes Q and R are qualified with highest size of Explicit Region and hence, select any one randomly. Suppose, if Q is selected as the root node, and the tree at root level is shown in figure 1.

The partition induced by branch ‘q1’ belongs to class ‘d1’ and hence create a leaf node and label it as ‘d1’ and also the partition induced by branch ‘q3’ belongs to class ‘d3’. But, the samples of partition induced by branch ‘q2’ belong to classes ‘d2’ and ‘d3’. Hence, calculate the Explicit regions of the attributes on the data available at branch induced by ‘q2’ only i.e., find the explicit regions of the attributes P,R, and S. The final decision tree induced is shown in figure 2.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7



U	P	R	S	D
2	p2	r1	s2	d1
6	p1	r2	s2	d1
8	p1	r1	s1	d1

U	P	R	S	D
1	p1	r1	s3	d2
3	p2	r1	s1	d2
4	p3	r3	s1	d3
5	p2	r3	s1	d3

U	P	R	S	D
7	p3	r3	s1	d3

Fig. 1 Decision Tree induced by RST approach at Root level

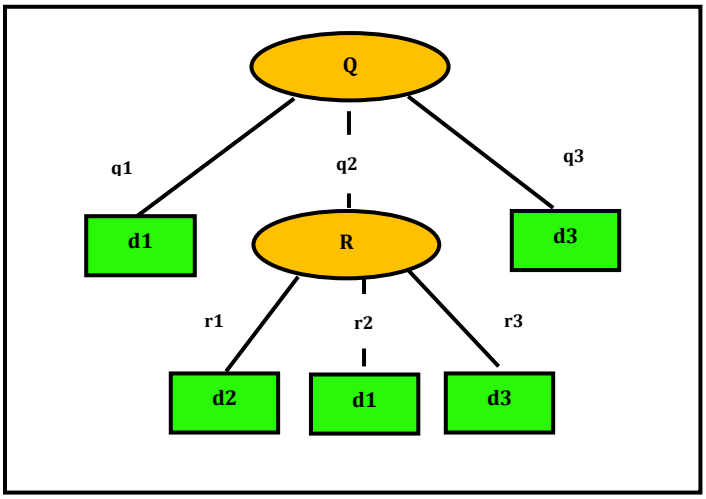


Fig. 2 Final Decision Tree induced by RST approach

The number of computations required is 4 at first level and at 3 at second level. So, total number of computations is 7. Now, the following steps illustrate the construction of decision tree for the reduced dataset given in Table 3. At first level, the number of computations required to select a root node have been reduced to 2 and then only one computation. Finally, the total number of computations required to induce the decision tree from the reduced dataset is reduced to 3 only.

Table 3. Reduced Sample DataSet

1	q2	r1	d2
2	q1	r1	d1
3	q2	r1	d2
4	q2	r3	d3
5	q1	r2	d1
6	q3	r3	d3

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

The decision tree induced by the RST approach on the reduced data in Table 3 is shown in figures 3 and 4.

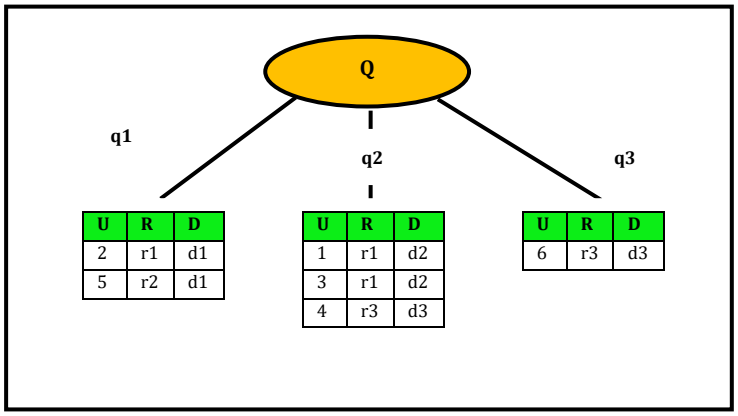


Fig. 3 Decision Tree induced by RST approach on reduced data at Root level

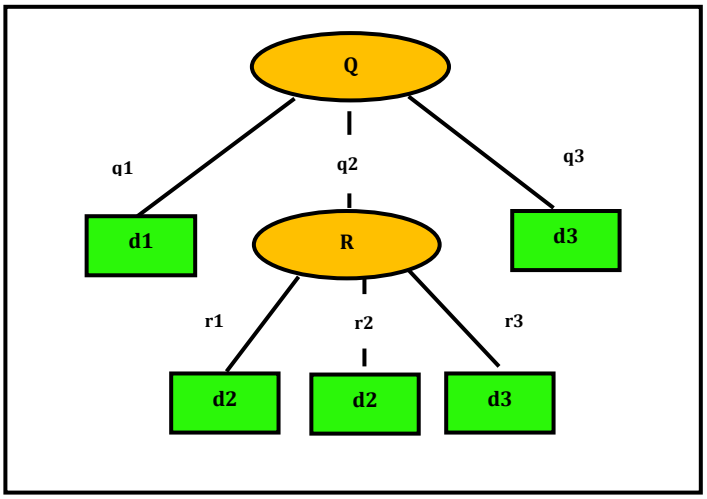


Fig. 4 Final Decision Tree induced by RST approach on reduced data

From the above example, it can be observed that the number of computations required to induce the decision tree using RST concepts for the consistent dataset represented in Table 2 is 7. But, applying the same concept on the reduced dataset the number of computations required to induce the same decision tree have been reduced from 7 to 3 i.e., the presence of irrelevant attributes in the data sometimes increases the complexity of the mining algorithm and hence, feature subset selection is becoming an important step in the data preprocessing.

RESULT ANALYSIS

For experimental analysis some standard medical datasets were taken from UCI repository [9]. Table 4 gives the description of the datasets. In experiments, 10-fold cross validation[10] examination is done to validate the results and the average of all 10 folds is taken as the average accuracy rate.

Table 4. Description of Datasets

DataSet	Instances	Attributes	Classes
Hepatitis	112	18	2
Diabetes	700	8	2
Thyroid	400	26	3
Thoraric Surgery	400	16	2

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

The classification accuracies of RST based decision tree on the raw dataset with inconsistencies and without inconsistencies is given in Table 5. Applying the RST concepts on the above mentioned five datasets to identify inconsistencies, no inconsistencies are observed in the Hepatitis dataset. But, in the remaining four datasets Diabetes, Thyroid, Thoraric surgery, and Liverdisorders some inconsistencies are identified and inconsistent objects are removed to make the dataset consistent.

Table 5. Classification Accuracies on InConsistent DataSet

Datasets	With Inconsistencies		Without Inconsistencies	
	Leaves	Accuracy (%)	Leaves	Accuracy (%)
Hepatitis	*No Inconsistencies*		83	***51.81***
Diabetes	393	58.70	264	***70.29***
Thyroid	51	94.44	11	***96.42***
Thoraric Surgery	161	73.86	87	***80.25***

Submitting the consistent dataset to the RST based reduct generation algorithm, the feature subset obtained for all the datasets is shown in Tables 6, 7 and 8.

Table 6. Feature Subset obtained by RST based Quick Reduct Forward Algorithm

DataSet	Quick Reduct-Forward			
	Reduct Size	Feature Subset	Leaves	Accuracy (%)
Hepatitis	15	1,2,3,4,5,6,7,8,9,10,11,15,16,17	85	51.81
Diabetes	5	1,5,6,7,8	140	62.19
Thyroid	9	1,4,5,8,12,18,20,22,26	15	96.35
Thoraric Surgery	13	1,3,4,5,6,7,8,9,10,11 13,14,16	86	80.70

Table 7. Feature Subset obtained by RST based Quick Reduct Backward Algorithm

DataSet	Quick Reduct-Backward			
	Reduct Size	Feature Subset	Leaves	Accuracy (%)
Hepatitis	11	6,7,9,10,11,12,15,16,17,18	82	60
Diabetes	6	1,2,3,4,5,8	202	68.19
Thyroid	9	3,6,10,11,16,18,20,22,26	11	96.35
Thoraric Surgery	14	1,2,3,4,5,6,7,8,9,10,12 13,15,16	93	79.61

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

Table 8. Feature Subset obtained by RST based Improved Quick Reduct Algorithm

DataSet	Quick Reduct-Improved			
	Reduct Size	Feature Subset	Leaves	Accuracy (%)
Hepatitis	7	1,2,4,15,16,17,18	83	51.81
Diabetes	5	1,5,6,7,8	140	62.19
Thyroid	8	2,3,4,14,18,20,22,26	11	96.57
Thoraric Surgery	2	12,15	4	86.06

For some datasets, the observed accuracies for the feature subset generated by the RST based reduct generation algorithms is same as that of the accuracies obtained for the raw dataset and for some of the datasets the accuracies have been increased at an acceptable rate.

The comparison of the performance of the RST based reduct generation algorithms is depicted in figures 5& 6.

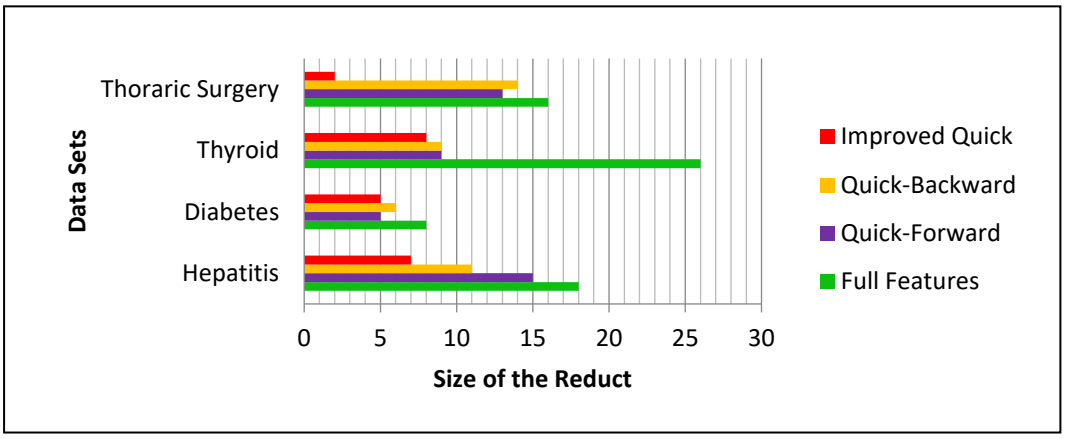


Fig. 5 Comparison of the feature subsets obtained by RST based reducts generation algorithms

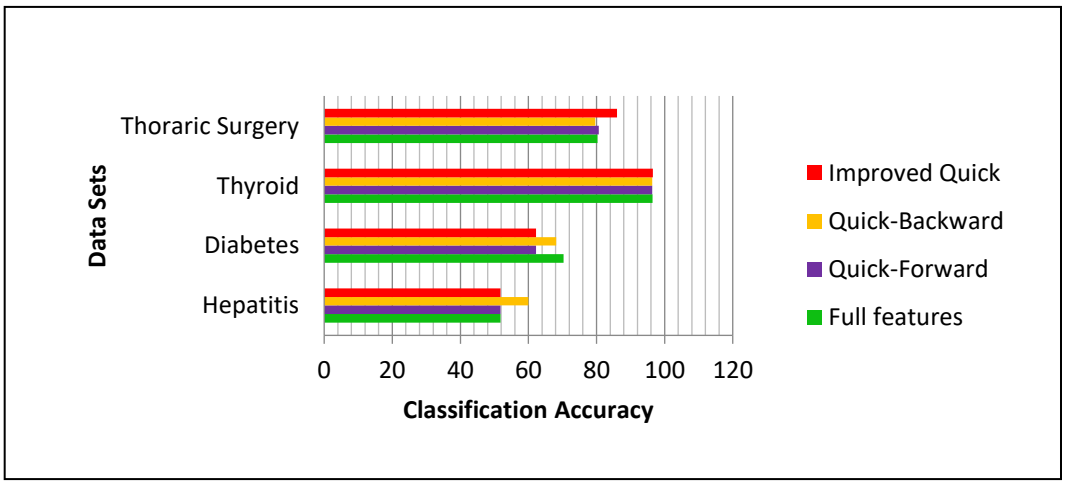


Fig. 6 Comparison of the performance of RST based reduct generation algorithms

From figure 3, the size of the reduct set generated by the Improved Quick Reduct generation algorithm is less when compared with the Quick Reduct Forward and Backward algorithms.

CONCLUSION

In this paper, RST concepts are applied for preprocessing of data to remove inconsistencies and various RST based reduct generation algorithms are studied. The efficiency of various RST algorithms for feature selection is tested by submitting the reduced feature set to the RST based decision tree classification. Experiments on UCI ML data reveal that the Improved Quick Reduct generation algorithm generated optimal feature subset and also observed improvement in the classification accuracy.

REFERENCES

[1] Mitchell, T.M., "*Machine Learning"*. McGraw Hill, New York, 1997.
[2] Rokach, L., & Maimon, O, "*Datamining with decision trees: theory and applications",* World Scientific Publishing Co. Pte. Ltd., Singapore. 2008.
[3] Pawlak, Z., "Rough Sets", *International Journal of Computer and Information Science*, 1982, 11(5), pp.341- 356.
[4] Thangavel, K., & Pethalakshmi, A., "Dimensionality Reduction based on Rough Set Theory", *Applied Soft Computing, Elsevier*, 2009,9(1), pp. 1-12.
[5] Pawlak, Z, "Rough set approach to knowledge-based decision support", *European Journal of Operational Research*, 1997,99(1), pp.48-57.
[6] Surekha S, "An RST based Efficient Preprocessing Technique for Handling Inconsistent Data", *Proceedings of 2016 IEEE International Conference on Computational Intelligence and Computing Research*, 2016, pp. 298-305.
[7] Jaganathan, P., Thangavel, K., Pethalakshmi, A., Karnan, M. "Classification Rule Discovery with Ant Colony Optimization and Improved Quick Reduct Algorith",. *IAENG International Journal of Computer Science*, 2007, 33(1), pp. 50-55.
[8] Jin Mao Wei, "Rough Set Based Approach to Selection of Node", *International Journal of Computational Cognition*, 2003, 1(2), pp. 25–40.
[9] Irvine UCI Machine Learning Repository. [http://archive.ics.uci.edu/ml/] A University of California, Center for Machine Learning and Intelligent Systems.
[10]Bengio, Y., & Grandvalet, Y., "No Unbiased Estimator of the Variance of K-Fold Cross-Validation", *Journal of Machine Learning Research*, 2004,5, pp.1089–1105 .