ScienSage

Journal of Advanced Scientific Research

Available online through http://www.sciensage.info/jasr

ISSN
0976-9595
Research Article

ROUGH SET APPROACH FOR FEATURE SELECTION AND GENERATION OF CLASSIFICATION RULES OF HYPOTHYROID DATA

Kavitha C.R [1*], Mahalekshmi T[2]

[1*] *Research Scholar, R&D, Bharathiar University, Coimbatore, Tamil Nadu*

[2] *Principal, Sree Narayana Institute of Technology, Kollam, Tamil Nadu*

**Corresponding author: kavithacr@gmail.com*

ABSTRACT

Rough Set theory (RST) has been successfully employed in several researches for feature selection, classification, dependencies among the attributes and rule induction. This paper presents two applications using rough sets. The first application deals with feature selection which uses RST for classifying hypothyroidism. The second application deals with generating classification rules from 498 samples of hypothyroid data.

Keywords: Classification, hypothyroid, reducts, rough set theory, rule generation

1. INTRODUCTION

Hypothyroidism, [1] also called underactive thyroid disease, is a common worldwide disorder. With hypothyroidism, the thyroid gland cannot make enough thyroid hormone. The thyroid gland is an endocrine gland. It is located in the front lower part of the neck. The thyroid gland releases thyroxine (T4) and trio dothyronine (T3) into the blood stream as the principal hormones. Hormones released by the gland travel through the bloodstream and affect nearly every part of the body, from heart and brain, to your muscles and skin. The thyroid controls how the body's cells use energy from food, a process called metabolism. The metabolism affects body's temperature, heartbeat, and to burn calories. If one doesn't have enough thyroid hormone, the body processes slow down.

Women, especially those older than age 60, are more likely to have hypothyroidism. Hypothyroidism upsets the normal balance of chemical reactions in your body. It seldom causes symptoms in the early stages, but, over time, untreated hypothyroidism can cause a number of health problems, such as obesity, joint pain, infertility and heart disease [2].

The most important and challenging task is to predict and diagnose hypothyroidism at the right time so that proper care and treatment can be provided to the patients by the doctors. Data mining techniques can be used for the prediction of hypothyroidism which would be helpful for the doctors to initiate treatments for the patients suffering from it.

This paper is organized as follows. The hypothyroid dataset, Basic concepts of rough set theory that are relevant to the work are introduced in section 2. Experimental Analysis, Results, discussion and a comparison study that shows feature selection can improve the classification accuracy and also leads to the generation of decision rules with better classification performance is discussed in section 3 followed by the conclusion in section 4.

2. MATERIAL AND METHODS

2.1. Data Set

The data set used for the experiment is the hypothyroid data which was downloaded from UCI repository ("https://archive.ics.uci.edu/ml/machine-learning-databases/thyroid-disease/") [3]. The characteristics of the data set are summarized in the table 1. The aim of the data set was to determine whether a patient referred to the clinic has hypothyroid on the basis of several input attributes such as are age, sex, on_thyroxine, query_on_thyroxine, onanti thyroid medication, sick, pregnant, thyroid_ surgery, I131_treatment, query hyperthyroid, query hypothyroid, lithium, goitre, tumour, hypopitutory, psych, TSH_ measured, T3_measured, TT4_measured, FTI_ measured, TBG_measured, TSH, T3, TT4, FTI, TBG, referral source, T4U_measured and T4U. The data set contains 500 observations and 29 variables with missing values reported.

Table 1: Hypothyroid Data Set Characteristics

Data Set	Hypothyroid
No of Example	500
Input Attributes	28
Output Classes	3
Total No. of Attributes	29
Missing Attributes status	Yes
Noisy Attributes status	Yes

2.2. Rough Set Theory-Basic Concepts

Rough set theory (RST) developed by Pawlak in the1980s, is used for data analysis, for discovering inter data

relationships, for finding interesting patterns and for decision making [4]. The basic concepts of RST that are relevant to this work are described below.

- Information System

A triple (U, A, F) is called an information system (IS), if U and A are sets and F is a set of relation between U and A. U= $\{x_1, \ldots, x_n\}$, each x_i (i <= n) is called an object. A = $\{a_1,\ldots,a_m\}$, each a_j(j<=m) is called an attribute. F = $\{f_j \backslash j<=m\}$, where $f_j = fa_j : U \rightarrow V_j$(j<=m) and V_j is the domain of the attribute a_j [5]. The set of attributes can be divided into two subsets, conditional set of attributes, C and decision set of attributes D, where C and D are subsets of A. Conditional set of attributes represent all the features or attributes of objects and decision set of attributes represent the classification of objects [6].

- Indiscernibility relation

For every set of attributes B $\subseteq$ A, an indiscernibility relation IND (B) is defined in the following: two objects, x_i and x_j, are indiscernible by the set of attributes B $\subseteq$ A, if b (x_i) = b(x_j) for every b $\subseteq$ B. The equivalence class of IND (B) is called elementary set in B because it represents the smallest discernible groups of objects. For any element xi of U, the equivalence class of xi, in relation IND (B) is represented as $[x_i]_{IND(B)}$. [7].

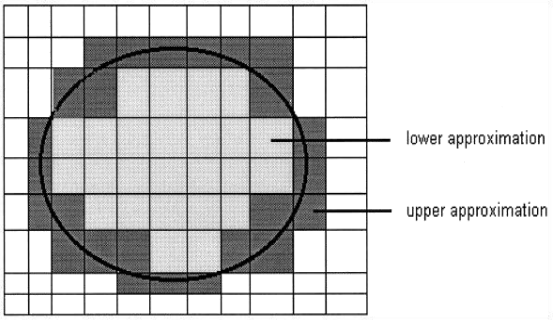

Fig.1. Schematic representation of the upper and lower approximation of set X [8].

- Lower and Upper Approximation

For any concept X$\subseteq U$ the attribute subset P $\subseteq A$, X could be approximated by the P- Upper and Lower approximation using the knowledge of P. The lower approximation of X is the set of objects of U that are surely in X ,where as the upper approximation of X is the set of objects of U that are possibly in X. The lower and upper approximations are defined as follows $P_*(X) = x\epsilon U \{P(x) : P(x) \subseteq X\}$ & $P^*(X) = x\epsilon U \{P(x) : P(x) \cap X \neq \emptyset\}$. The boundary region is defined as: $BN_P(X) = P^*(X) - P_*(X)$. If the boundary region is empty that is upper approximation is equals to lower approximation, concept X is said to be P definable or else X is a Rough Set with respect to P. The Positive Region of decision class U/IND(P) with respect to conditional attribute C is denoted by $POS_c(D) = P_*(X)$ [9]. The positive region POS_C (D) is a set of objects of U that can be classified with certainty to classes U/IND(D) employing attribute of C.

- Core and Reduct Attributes

The concepts of core and reduct are two fundamental concepts of the rough sets theory. The reduct is the essential part of an IS, which can discern all objects discernible by the original IS. The core is the common part of all reducts. If the set of attributes is dependent, one can be interested in finding all possible minimal subsets of attributes, which lead to the same number of elementary sets as the whole set of attributes (reducts) and in finding the set of all indispensable attributes (core). To compute reducts and core, the discernibility matrix is used.

Reduct is a minimal subset R of initial attribute set C(conditional) such that for a given set of decision attribute D$\gamma R\ D = \gamma R(C)$. Reduct is the minimal set of attributes preserving positive region. There may exist many reducts for an Information System. The Core is the set of attributes that are contained by all Reducts i.e. CORED(C) = $REDD(C)$. CORE is the set of attribute that cannot be removed without changing the positive region.

3. RESULTS AND DISCUSSION

3.1. Steps of the Experiment

In this paper, rough set theory is used for feature selection. The general process of feature subset selection is shown in the fig 2. The different steps that were carried out in the experiment and the result analysis are described here.

A. Preprocessing

The downloaded hypothyroid dataset contains missing and redundant attributes. The missing values of the dataset are completed by common or mean value imputation.

B. *Calculating Reducts*

After missing value imputation, reducts are generated using rough set reduction technique, by which 36 reducts were obtained. Reduction technique is used to remove the redundant attributes from the dataset initially.

C. Feature Selection

Feature selection refers to the process of choosing the most relevant features for a given task, while discarding the noisy, irrelevant and redundant features of the dataset. It refers to the selection of a subset of M features from a set of N features, M< N, such that the value of a criterion function is optimized over all subsets of size M [10]. The main aim of feature selection is to select a subset of features such that the classification accuracy does not significantly decrease; and the resulting class distribution, given only the values for the selected features, is as close as possible to the original class distribution, given all features [11].

The steps that were carried out for conducting feature selection are described below.

1. First the reducts were calculated using rough set reduction technique and 36 reducts were obtained. The attribute that occurred in the reducts were found out and were ranked

based on the decreasing order of their significance as shown in the table 2. Proportional rough set (PRS) [12] relevance method is used for the ranking the attributes that occurred in the reducts. To explain the feature ranking process, consider a decision table $T = \{U, A, d\}$, where U is the non-empty finite set of objects called the Universe, $A = \{a_1, a_2, \ldots, a_n\}$ be the finite set of conditional attributes and d is the decision attribute. Let $\{r_1, r_2, \ldots, r_p\}$ be the set of reducts generated from T. Then, for each conditional attribute $a_i \varepsilon A$, reduct based attribute significance $\beta(a_i)$ is defined as [13]:

$$\beta(a_i) = \frac{\left|\{r_j | a_i \in r_j, j = 1,2,3,\ldots, p\}\right|}{p}, i = 1,2,3,\ldots,n$$

where the numerator represents the occurrence frequency of the attribute a_i in various reducts.

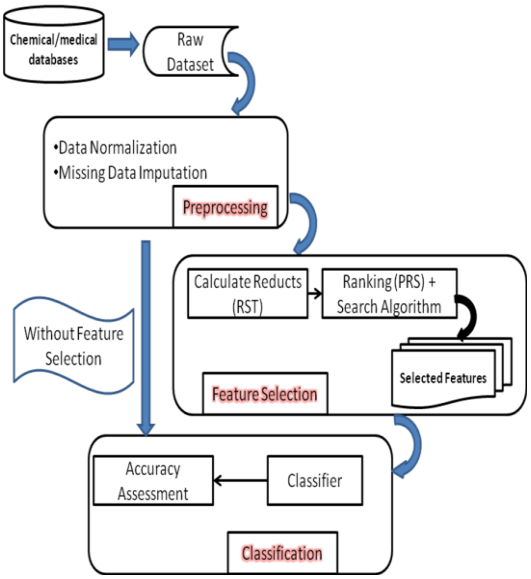

Fig.2: Process of Feature Subset Selection

2. After the first step, the classification accuracy of the whole hypothyroid dataset with all features is found out, which will be used for the construction of the best feature subset. Then, the set of ranked features is given to the feature selection algorithm FeaSel [14].
3. The algorithm starts with all features of hypothyroid and in the first iteration the algorithm selects lowest ranked feature Psych and on antithyroid (these 2 attributes have same rank 17) and query hypothyroid (rank 16) as the test feature. Since there is no change occurs in the original classification accuracy while eliminating this feature, it is designated as redundant and hence it is removed from the feature set. Since the features are ranked in decreasing order of significance, features with lower ranks gets eliminated during initial stages. The features are retained if there is a reduction in the classification accuracy which signifies the influence of this feature in determining the classification accuracy. The process is continued until all features are evaluated. The performance of various classifiers of different iteration is shown in the table 3.
4. After evaluating all features of the Hypothyroid dataset, the algorithm retains the set of features {age, tumor, TSH,T3, TT4, T4U, FTI}. These features are considered significant and important because though all other features are removed from the hypothyroid dataset, it doesn't affect the classification performance.

Table 2: Attributes with priority values in a decreasing order

Rank	Attribute	Frequency	Significance
1	TSH	18	1
2	TT4	17	0.9444
3	Age	12	0.6667
3	T3	12	0.6667
5	Tumor	9	0.5
6	FTI	8	0.4444
6	Referral source	8	0.4444
8	T4U	6	0.3333
9	Sex	4	0.2222
9	On thyroxine	4	0.2222
9	Thyroid surgery	4	0.2222
9	TT4 measured	4	0.2222
9	T3 measured	4	0.2222
9	T4U measured	4	0.2222
9	FTI measured	4	0.2222
16	Query hypothyroid	3	0.1667
17	Psych	1	0.0556
17	On antithyroid medication	1	0.0556

Table 3: Classifiers performance

Hypothyroid Dataset	Classifiers	Accuracy (%)	Time (Sec)
Complete Attributes (28 attributes)	J48	98	0.05
	Naïve Bayes	95	0.04
Occurrence of attributes in reducts	J48	99	0.01
	Naïve Bayes	96	0.01
Elimination of 3 lowest ranked attributes	J48	99	0.01
	Naïve Bayes	96	0.01
Elimination of next 7 lowest ranked attributes	J48	98	0
	Naïve Bayes	96	0.01

3.2 Classification Rule Generation

Decision rule generation is a very important task in any learning system. The aim of data reduction is to find a minimal subset of relevant attributes that have all the essential information of the data set, thus the minimal subset of the attributes can be used instead of the entire

attributes set for rule discovery [15]. Classification rule generation process is shown in the Fig. 3.

Decision rules are generated based on different cases: (i) with reducts (ii) without reducts (iii) with feature selection and (iv) without feature selection.

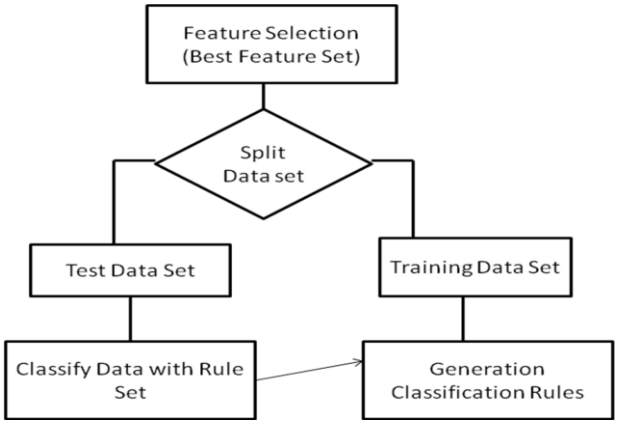

Fig.3. Classification Rule Generation process

Experimental results show that rough set theory in feature selection does improves the classification accuracy and efficiency of the algorithm. The selected feature subset is used to generate decision rules for the classification task. Rule generation is considered as very important and crucial task in any learning system [16].

Table 4: Accuracy Assessment at different iterations

	Data with all attributes	Data with attribute occurred in reducts	Data with best subset Attributes
Total Number of tested Objects	200	200	200
Total Accuracy	0.87	0.88	0.89
Total Coverage	1	1	1

3.3. A comparison study

In this section, a comparative study is made to find whether the feature selection improves the performance of the classification algorithm. The experiment was conducted using classifiers such as J48 and Naïve Bayes. From the table 3, we can understand that the accuracy of the classifier J48 is the same before and after feature selection and for the classifier Naïve Bayes, the classification accuracy after feature selection has been increased by 1% than without feature selection. Table 4 shows the classification accuracy assessment of the data set at different iterations.

From table 5, we can note that the number of reducts generated using feature selection is very less than compared to the reducts generated without feature selection. When rules are generated through reducts, we can note from table 6 and table 7 that the number of rules and filtered rules generated without feature selection are more than the number of rules and filtered rules generated with feature selection.

Table 5: Rules through Reducts

Feature Selection	Algorithm	No of reducts	Lengths of Reduct		
			Min	Max	Mean
No	Genetic Algorithm	46	2	5	3.4
Yes	Genetic Algorithm	7	2	3	2.1

Table 6: Rule Generation

Feature Selection	Algorithm	No of rules	Lengths of Reduct		
			Min	Max	Mean
No	Genetic Algorithm	11892	2	5	3.3
Yes	Genetic Algorithm	1925	2	3	2.1

Table 7: Filtered Rules

Feature Selection	Algorithm	Filtered rules	Lengths of Reduct		
			Min	Max	Mean
No	Genetic Algorithm	706	2	5	3.6
Yes	Genetic Algorithm	68	2	3	2.3

From table 8, we can understand that the accuracy is more for the classification done after feature selection.

Table 8: Accuracy Assessment

Feature Selection	Algorithm	Accuracy	Coverage
No	Genetic Algorithm	0.966	0.88
Yes	Genetic Algorithm	0.988	0.425

Table 9: Rule generation without reducts

Feature Selection	Algorithm	No of rules	Lengths of Reduct		
			Min	Max	Mean
No	Genetic Algorithm	825	1	5	1.5
Yes	Genetic Algorithm	649	1	3	1.3

When rules are generated without reducts, the number of rules and filtered rules generated without feature selection is higher than the number of rules and filtered rules generated with feature selection is shown in the table 9 and table 10.

From table 11, we can find that the accuracy of the classification algorithm with feature selection gives the high accuracy (89%) than the classification done without feature selection.

Table 10: Filtered rules without reducts

Feature Selection	Algorithm	Filtered rules	Lengths of Reduct		
			Min	Max	Mean
No	Genetic Algorithm	387	1	5	1.4
Yes	Genetic Algorithm	262	1	2	1

Table 11: Accuracy Assessment without reducts

Feature Selection	Algorithm	Accuracy	Coverage
No	Genetic Algorithm	0.875	1
Yes	Genetic Algorithm	0.885	1

4. CONCLUSION

In this paper, feature selection was performed using rough set theory to the hypothyroid dataset for classification that improved the performance and accuracy of the classifiers. The rough set rule induction algorithms are capable of generating decision rules which can potentially provide new insight and profound knowledge. The decision rules are generated for the cases (i) with reducts (ii) without reducts (iii) with feature selection and (ii) without feature selection by using the Genetic algorithm. It is evident from the experimental study and analysis that feature selection indeed leads to generation of decision rules with better classification accuracy. And finally, it is also found that the reducts certainly helps in improving the rule generation. The experiment proves that the rules generated through reducts showed better classification accuracy than rules generated without reducts.

5. REFERENCES

1. Webmd.com: the web's trusted source on Hypothyroidism [Internet], ©2005-2016 [cited 2016 Feb 2]. Available from: http://www.webmd.com/.
2. Banu RG, Baviya M. *International Journal of Modern Trends in Engineering and Research*, 2015; **2:** 666-700.
3. Lichma M. *UCI Machine Learning Repository* [Internet], 2013; [cited 2016 Feb 3] Available from: http://archive.ics.uci.edu/ml.
4. Pawlak JZ, Busse G, Slowinski R, Ziarko w. *Communications of the ACM*, 1995; **38**: 88-95.
5. JunQi J, Wei L, Li Z. *Springer*, 2005; **3641**: 74-83.
6. Sengupta N, Sil J. *International Journal of Information and Electronics Engineering*. 2012; **2**: 464-466.
7. Kavita S, Mukesh Z, Mukesh R. *Rough Sets And Knowledge Technology*. 2012; **7414**:91-100.
8. Anitha K, Venkatesan P. *International Journal of Innovative Research in Science, Engineering and Technology*, 2013; **2**:3989-3993.
9. Walczak B, Massart DL. *Chemometrics and Intelligent Laboratory Systems*, 1999; **47**: 1-16.
10. Narendra G, Fukunaga K. *IEEE Transactions on Computers*. 1977; **C-26**:917-922.
11. Dash M, Liu H. *Intelligent Data Analysis Elsevier Science*. 1997; **1088-467X**: 131-156.
12. Tang J, Alelyani S, Liu H. *Intelligent Data Analysis*, 1997; **116**: 1-29.
13. Salamo MM, Sanchez LM. *Pattern Recognition Letters*. 2011; **32**:280-292.
14. Sabu MK. *International Journal of Computer Science & Information Technology*, 2015; **50113**: 127-137.
15. Li J, Cercone N. *IEEE International Conference on Granular Computing*, 2005; **2**: 506-511.
16. Hassanien AE, Ali MH. *Informatica*, 2004; **15**:23-38.