A New Approach of Rough Set Theory for Feature Selection and Bayes Net Classifier Applied on Heart Disease Dataset

Eman S.Al-Shamery **Ali A.Rahoomi Al-Obaidi**
Department of software ,College of Information Technology ,University of Babylon.
emanalshamery@yahoo.com a_li05@yahoo.com

Abstracts

In this paper a new approach of rough set features selection has been proposed. Feature selection has been used for several reasons a) decrease time of prediction b) feature possibly is not found c) present of feature case bad prediction. Rough set has been used to select most significant features. The proposed rough set has been applied on heart diseases data sets. The main problem is how to predict patient has heart disease or not depend on given features. The problem is challenge, because it cannot determine decision directly .Rough set has been modified to get attributes for prediction by ignored unnecessary and bad features. Bayes net has been used for classified method. 10-fold cross validation is used for evaluation. The Correct Classified Instances were 82.17, 83.49, and 74.58 when use full, 12, 7 length of attributes respectively. Traditional rough set has been applied, the minimum Correct Classified Instances were 58.41 and 81.51 when use 2 length of attributes respectively.

Keywords heart disease, rough set, Bayes net, feature selection.

الخلاصة

درسنا في هذا البحث اختيار الصفات بالاعتماد على نهج جديد من خوارزمية مجموعة التقريب حيث تعتمد هذه الطريقة على اختيار الصفات الأكثر تاثيرا. لجئنا الى انتقاء الصفات اختصارا للوقت , وجود الصفة تؤثر على دقة النتائج او قد تكون الصفة غير متوفرة . تم تطبيق الخوارزمية على بيانات امراض القلب لاختيار افضل الصفات المؤثرة. ان المشكلة الرئيسية هو كيفية تشخيص الإصابة فيما لو كان مصاب بمرض القلب من عدمه.هذه المشكلة تمثل تحدي لان لا نسطيع اتخاذ القرار بصورة مباشرة. تعتمد الطريقة المقترحة على ترميز البيانات الاصلية .ان الناتج من هذه الخوارزميه هي الصفات الأكثر أهمية حيث تهمل الصفات السيئة والغير ضرورية.وتم تطبيق النتائج على خوارزمية شكبة بيزينت كخوارزمية للتنبؤ بالمرض وقد حصلنا على النتائج 82.17 , 83.49 , 74.58 عند استخدام جميع الصفات, 12 , 7 طول الصفات على التوالي.وتم تطبيق نتائج خوارزمية مجموعة التقريب الاصلية على خوارزمية البيزين وحصلنا على النتائج 81.51, 58.41 عند استخدام 2 , 12 طول الصفات على التوالي.

الكلمات المفتاحية : امراض القلب,خوارزمية مجموعة التقريب,خوارزمية البيزين نت, اختيار الصفات.

1. Introductions

Rough set is proposed by Pawlak in1982 to deal with uncertainty and incompleteness. It is a new intelligent mathematical tool based on approximation space [Chen (2011)]. Rough set has been applied in many fields such as data mining, pattern recognition, machine learning, and signal processing, datasets containing huge numbers of features, etc [Wang (2006)].

In machine learning data contain vague and incomplete information. In [Vluymans (2015)] they highlight the interaction between practical machine learning tools and theoretical advances on fuzzy rough sets that take advantage of them. In [Mahajan (2012)] they include various machine learning techniques such as clustering, rule induction and feature selection by using rough set theory.

In Data mining rough set theory proposals extract decision rules of the data sets. The rough set theory offers approach for extraction meaningful knowledge and then making predictions for an individual data object. A rule extracted from a data set is one of many models that describing data set [Kusiak (2001)].

In pattern recognition rough set methods present applications for feature selection. The role of rough set feature selection, namely reducts and their approximations, including dynamic reducts [Swiniarski (2003)].

In signal processing, a rule-based rough-set decision system has been used for the development of a disease inference engine. Using image-processing techniques an offline-data-acquisition system is developed for electrocardiogram (ECG) records [Mitra (2006)].

Rough-Set Feature Selection has been applied on different fields such as:

In [Qamar (2012)] Rough-Set Feature Selection used for Clustering to knowledge discovery and the conclusion indicated a very significant result that removal of individual numeric attributes, it analyses the effects of rough sets on clustering using 10 datasets.

In [Gupta (2006)] Textual Case-Based Classification using Rough Set Feature Selection Algorithms has been used to increase case-based task performance and reduce the high dimensionality of textual cases.

Unfortunately, examining exhaustively all subsets of features for selecting the optimal one is NP-hard because the number of possible subsets is always very large when N is large features so the number of possible subsets are 2^N subsets [Chen (2010)]. Feature selection using rough set theory has been widely used in data analysis, it used to select subset of attributes has the same equivalence relation of entire attribute, is referred to as reduct [Chen (2011)] .

2. Rough set theory

2.1 Traditional rough set

In [Nahato(2015)] traditional rough steps .

2.2. Modified rough set

Input: full features of heart diseases data set

Output: most significant features

----

1. coding data
2. Find lower approximation:

   Data value must give same class (normal or abnormal)
3. Find upper approximation:

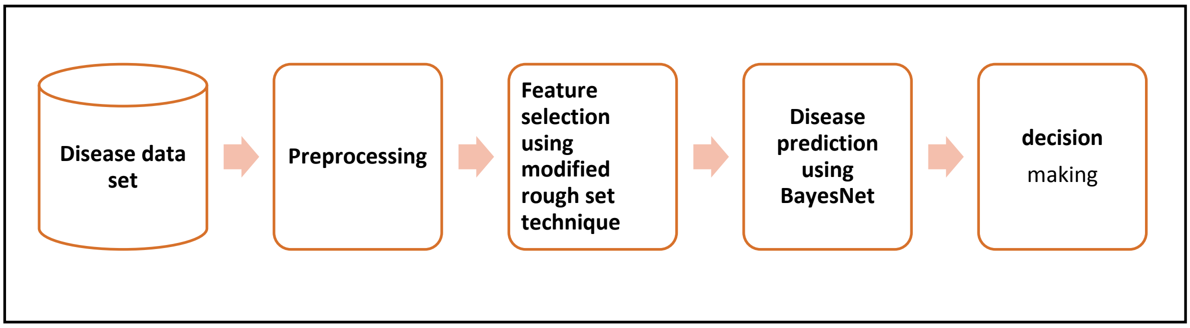
   Data value possible give same class (normal or abnormal)
4. Find boundary

   Data value has no decision
5. Find positive reign $POS = \cup \ \underline{AX}$

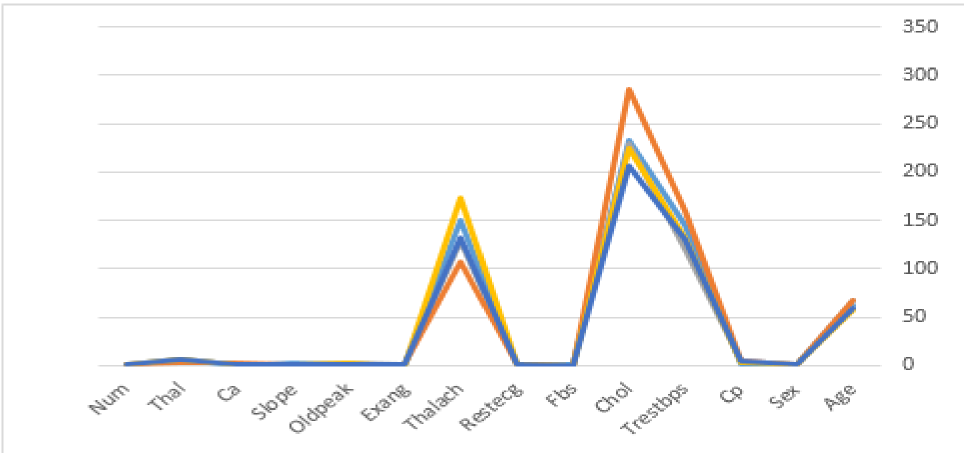
   Union of lower reign
6. Find indiscernibility for positive reign
7. Find reducts
   $$RED = \min IND$$
8. determine sub set of attribute including core attributes
   $$\text{Core} = \cap \ RED$$
9. Determine number of desired features
10. Select appropriate features from reducts has same length for number of desired features

3. Proposed system

In this paper, Cleveland heart diseases dataset has been used from machine learning repository [Asuncion (2007)]. Dataset contains 303 instances of which 139 instances belonged to the heart disease and 164 instances belonged to the healthy.



No	Attribute name	Min	Max	Description
1	age	29	77	years
2	sex	0	1	sex
3	cp	1	4	chest pain 4 types
4	Trestbps	94	200	resting blood pressure
5	Chol	126	564	cholesterol
6	Fbs	0	1	resting blood sugar 0=f , 1=t if fbs >120
7	Restecg	0	2	ECG
8	Thalach	71	202	max heart rate
9	Exang	0	1	exercise induced angina
10	Old peak	0	6.2	ST depression induced by exercise relative to rest
11	Slope	1	3	slope of peak exercise ST segment
12	Ca	0	3	major vessels 0-3
13	Thal	3	7	Thalium scan



Five Patient data

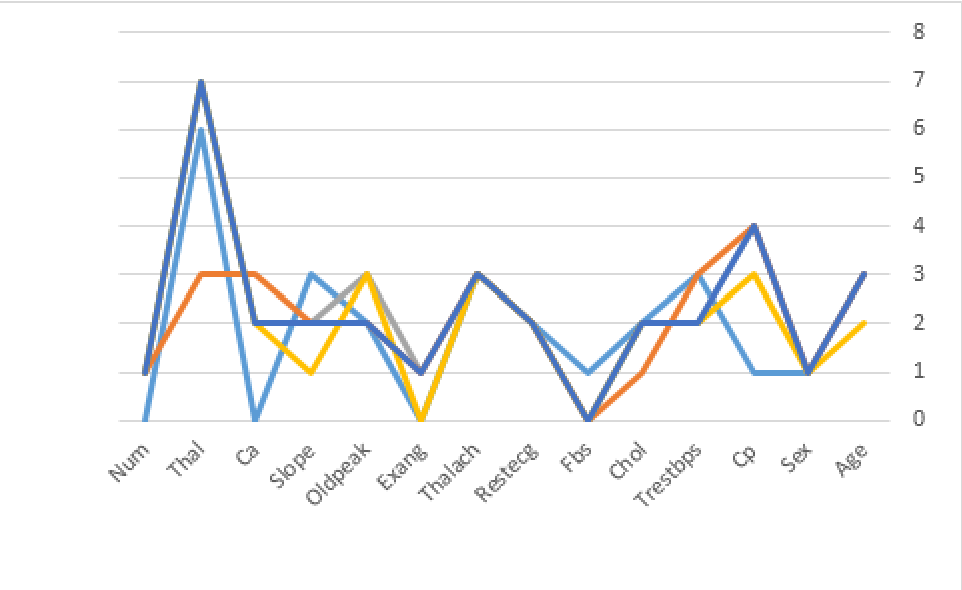
4. Experiments results and discussion

This research applied on heart disease data set. Rough set used two classes named as normal and abnormal, when class attributes equals zero and more than zero respectively.

4.2. Coding attributes

Some attributes have been coding to make general procedure when use Indiscernibility relation of rough set.

Input Field	Range	Linguistic Representation
Age	[20 – 40) [40-60) [60-80)	Young Mid Old
(Blood Pressure)	[90 – 139) >140	Normal Abnormal
Cholestoral	[150-250] <150 or >250	Normal Abnormal
Maximum heart rate	<60 [60-100] >100	Low Normal Hight



Five Patient data after coding

4.3. Lower approximation

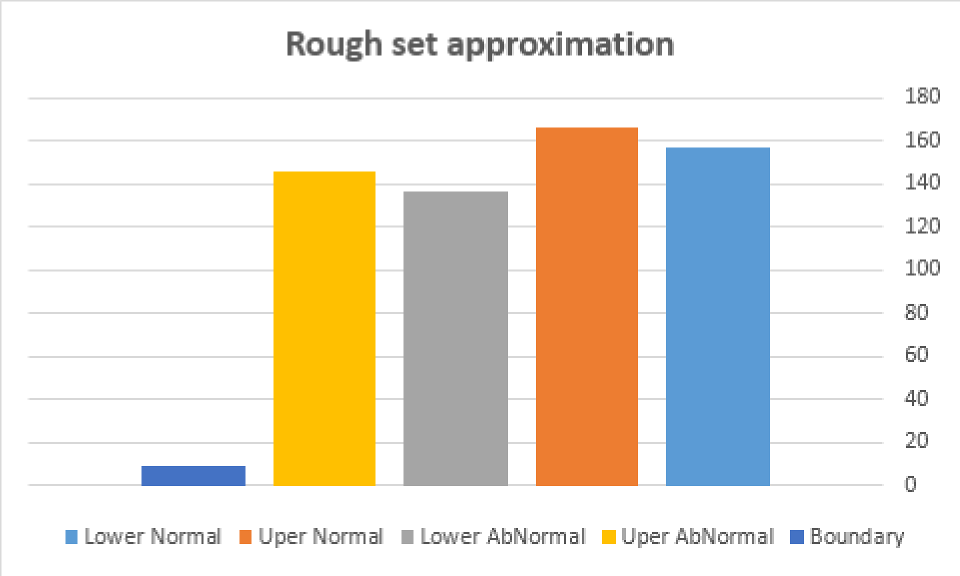
In this step assign normal and abnormal data in data set.

4.4. Boundary region

In boundary no decision making, people may be normal or abnormal.

4.5. Upper approximation

In this step assign people may be normal or abnormal, In other words union lower approximation and Boundary region.



4.6. Positive region

In this step union lower approximation, i.e. union of lower normal and lower abnormal is called positive region.

4.7. Indiscernibility

Ind (M) is defined as: if $m(x_i) = m(x_j)$ for every $M \subset N$, two objects, x_i and x_j, are indiscernible by the set of attributes M in Y, for every set of attributes $M \subset N$.

4.8. Reducts

It is minimal subsets of attributes can be interested in finding all possible. If indiscernibility relation of **set of attributes** and **its superset** is same then any attribute that found in superset and not found in the set is redundant and called reduct on subset.

4.8.1. Reducts of traditional rough set

Reducts	
Length of attributes	counts
2	1
3	17
5	244
6	138
7	39
8	6
9	1
Total	446

4.8.2. Reducts of proposed rough set

Reducts	
Length of attributes	**counts**
7	1
8	15
9	41
10	34
11	10
12	1
Total of Reducts	102

4.9. Core attribute

The intersection of the elements of reducts is called core.Trestbps and Restecg are core attributes of proposed rough set while traditional rough set doesn't has reduct.

5. Bayes net

Bayesian network is a method have been used to represent dependencies in a probability distribution graphically by using structure of a directed acyclic graph. Each feature represent as node and dependencies between these features represent as arcs [Salami (1996)]. A Bayesian network it is a statistical model compute for any subset of unobserved stochastic variables find the joint probability distribution given that the variables in the complementary subset are observed. Bayesian network it is a statistical classifier by using winner-takes-all rule to the posterior probability distribution for the unobserved class node.

1. Compute the conditional mutual information given the class variable C, $I(X_i; X_j | C)$ between each pair of features, $i \neq j$.
   $I(X_i; X_j | C)$ is defend as follows
   $$I(X_i; X_j | C) = \sum_{X_i; X_j} P(X_i = x_i, X_j = x_j, C = c_1) \times \log \frac{P(X_i = x_i, X_j = x_j / C = c_1)}{P(X_i = x_i | C = c_1) P(X_j = x_j | C = c_1)}$$
2. Build a complete undirected graph in which the nodes are the variables. Assign to each arc connecting X_i to X_j the weight $I(X_i; X_j | C)$.
3. Build a maximum weighted spanning tree.
4. Transform the resulting undirected tree to a directed one by choosing a root variable and setting the direction of all arcs to be outward from it.
5. Add the classification node C and draw an arc from C to each X_i. [Verstraeten (2004)]

6. Experimental results

Symbols	Meaning
CCI	Correctly Classified Instances
ICI	Incorrectly Classified Instances
MAE	Mean absolute error
RMSE	Root mean squared error
RAE	Relative absolute error
RRSE	Root relative squared error
TNOI	Total Number of Instances
TM	Time taken to build model

6.1 Evaluate on training data using full attributes

CCI	277	91.4191 %
ICI	26	8.5809 %
MAE		0.1009
RMSE		0.2491
RAE		20.3094 %
RRSE		49.9837 %
TNOI	303	
TM	0.03 seconds	

Confusion Matrix		
a	b	classified as
154	10	a = 0
16	123	b = 1

Table (1) confusion matrix when using full attributes of training data

6.2 10-fold cross-validation using full attributes

CCI	249	82.1782 %
ICI	54	17.8218 %
MAE		0.2041
RMSE		0.3772
RAE		41.0891 %
RRSE		75.7033 %
TNOI	303	
TM	0.01 seconds	

Confusion Matrix		
a	**b**	classified as
140	24	a = 0
24	109	b = 1

Table (2) confusion matrix when using full attributes of 10-fold cross-validation

6.3 Evaluate on training data using twelve length of attributes

Attributes		
Age	ex	Cp
Trestbps	Chol	Fbs
Restecg	Exang	Oldpeak
Slope	Ca	Thal

CCI	275	90.7591 %
ICI	28	9.2409 %
MAE		0.1138
RMSE		0.26
RAE		22.9208 %
RRSE		52.1768 %
TNOI	303	
TM	0.03 seconds	

Confusion Matrix		
a	**b**	classified as
154	10	a = 0
16	121	b = 1

Table (3) confusion matrix when using twelve length of attributes of training data

6.4 10-fold cross-validation using twelve length of attributes

Attributes		
Age	ex	Cp
Trestbps	Chol	Fbs
Restecg	Exang	Oldpeak
Slope	Ca	Thal

CCI	253	83.4983 %
ICI	50	16.5017 %
MAE		0.2006
RMSE		0.3689
RAE		40.3978 %
RRSE		74.0209 %
TNOI	303	
TM	0 seconds	

Confusion Matrix		
a	b	classified as
142	22	a = 0
28	111	b = 1

Table (4) confusion matrix when using twelve length of attributes of 10-fold cross-validation

6.5 Evaluate on training data using seven length of attributes

Attributes		
Age	Sex	Cp
Trestbps	Restecg	Oldpeak
Ca		

CCI	255	84.1584 %
ICI	48	15.8416 %
MAE		0.1969
RMSE		0.3162
RAE		39.6474 %
RRSE		63.4494 %
TNOI	303	
TM	0.03 seconds	

Confusion Matrix		
a	b	classified as
143	21	a = 0
27	112	b = 1

Table (5) confusion matrix when using seven length of attributes of training data

6.6 10-fold cross-validation using seven length of attributes

Attributes		
Age	Sex	Cp
Trestbps	Restecg	Oldpeak
Ca		

CCI	226	74.5875 %
ICI	77	25.4125 %
MAE		0.288
RMSE		0.4227
RAE		57.9803 %
RRSE		84.8232 %
TNOI	303	
TM	0.01 seconds	

Confusion Matrix		
a	b	classified as
131	33	a = 0
44	95	b = 1

Table (6) confusion matrix when using seven length of attributes of 10-fold cross-validation

7. Conclusions and future work

- Modified rough set is better than traditional rough set
- decrease time of prediction when modified rough set has been used
- present of some features case bad prediction

In future work making the algorithm operates on multilateral and bilateral and give the degree of importance for each attributes.

References

Andrew Kusiak,(2001),Rough Set Theory: A Data Mining Tool for Semiconductor Manufacturing,IEEE TRANSACTIONS ON ELECTRONICS PACKAGING MANUFACTURING, VOL. 24, NO. 1,1-12

Asuncion, A.; D.J. Newman, UCI Machine Learning Repository [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA: University of California, Department of Information and Computer Science, 2007.

Geert Verstraeten, Dirk Van den Poel, Michael Egmont-Petersen, Patrick Van Kenhove and Jan Vanthienen, (2004),Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers,Elsevier0377-2217,Computing, Artificial Intelligence and Information Technology,European Journal of Operational Research,508–523

Hui-Ling Chen, Bo Yang, Jie Liu and Da-You Liu ,(2011),A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis,Elsevier 0957-4174,Expert Systems with Applications,9014–9022

Kalyan Moy Gupta, David W. Aha and and Philip Moore,(2006),Rough Set Feature Selection Algorithms for Textual Case-Based Classification,Proceedings of Eigth European Conference on Case-Based Reasoning, ECCBR-06, Ölündeniz, Turkey: Springer,1-15

Kindie Biredagn Nahato, Khanna Nehemiah Harichandran, and Kannan Arputharaj,(2015), Knowledge Mining from Clinical Datasets Using Rough Sets and Backpropagation Neural Network, Hindawi Publishing Corporation Computational and Mathematical Methods in Medicine Volume 2015, Article ID 460189, 1-13

Mehran Sahami,Gates Building,1996, Stanford University, Computer Science Department, Learning Limited Dependence Bayesian Classifiers, Rule Induction 6r Decision Tree Induction,335-338

Prerna Mahajan,Rekha Kandwal and Ritu Vijay,(2012),Rough Set Approach in Machine Learning: A Review,International Journal of Computer Applications,Volume 56– No.10,1-12

Roman W. Swiniarski and Andrzej Skowron ,(2003),Rough set methods in feature selection and recognition,Elsevier Science,0167-8655/03,Pattern Recognition Letters 24, 833–849

Sarah Vluymans and Lynn Deer,(2015), Applications of Fuzzy Rough Set Theory in Machine Learning a Survey, Fundamenta Informaticae, DOI 10.3233/FI-2012-0000, IOS Press, 1–34

Sucharita Mitra,(2006),A Rough-Set-Based Inference Engine for ECG Classification,IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, VOL. 55, NO. 6,2199-2203

Usman Qamar, John A. Keane ,(2012),Clustering Using Rough-Set Feature Selection,Journal of Basic and Applied Scientific Research,ISSN 2090-4304, 5915-5920

Xiangyang Wang , Jie Yang, Xiaolong Teng, Weijun Xia and Richard Jensen, (2006), Feature selection based on rough sets and particle swarm optimization, Elsevier 0167-8655, 459–471

Yumin Chen, Duoqian Miao and Ruizhi Wang ,(2010),A rough set approach to feature selection based on ant colony optimization,Elsevier 0167-8655,Pattern Recognition Letters,226–233