Rough set based decision rule generation to find behavioural patterns of customers

L SUMALATHA[1], P UMA SANKAR[1,*] and B SUJATHA[2]

[1]Department of Computer Science and Engineering, University College of Engineering Kakinada, JNTUK, Kakinada 533003, India
[2]Department of Computer Science and Engineering, Godavari Institute of Engineering and Technology, Rajahmundry 533296, India
e-mail: lsumalatha@jntucek.ac.in; umashankar.5544@gmail.com; birudusujatha@gmail.com

Abstract. Rough sets help in finding significant attributes of large data sets and generating decision rules for classifying new instances. Though multiple regression analysis, discriminant analysis, log-it analysis and several other techniques can be used for predicting results, they consider insignificant information also for processing which may lead to false positives and false negatives. In this study, we proposed rough set based decision rule generation framework to find reduct and to generate decision rules for predicting the Decision class. We conducted experiments over data of Portuguese Banking institution. From the proposed method, the dimensionality of data is reduced and decision rules are generated which predicts deposit nature of customers by 90% accuracy.

Keywords. Rough sets; decision rules; classification; reduct; dimensionality reduction.

1. Introduction

For any bank, obtaining deposits is a preliminary function. Obtaining the deposits at lower rate and investing that amount at higher rate helps the bank to survive. The increase of customer deposits and loans granted for the interest margins have a positive impact on bank profitability. When the economic activity decreases, the demand for loans and deposits decreases and negatively affects the profit margins leading to liquidity risk. The significance of deposits leads to banks profitability and research carried on Turkish banks explain the importance of deposits [1]. Efforts like campaigning are done by every bank to motivate customers for increasing their deposits. In this paper, we proposed a rough set based decision (RSD) rule generation framework for predicting deposit nature of customers which will intern help a bank in finding customers who are willing to make deposits. By this it drastically reduces time and cost of campaigning. A data mining approach to predict success of bank telemarketing was done by Sergio [2]. There exist several classification models such as classical Logistic Regression, Decision Trees and the more recent Neural Networks and Support Vector Machines for predicting classification tasks [3, 4]. These models perform well when the data is crisp but inefficient when data is inconsistent and ambiguous, Rough set approach provides an efficient

methodology to perform classification even on inconsistent data. Rough set theory (RST) was introduced by Pawlak [5]. According to Rough Set Theory every object in existence will be associated with a certain amount of information known as attributes and these attributes can be used for identifying an object from others. Objects which cannot be distinguished from one another with available information are said to be indiscernible. The discernibility relation (properties on which two objects differ from one another) generated over all available objects results in a discernible matrix which is base for the Rough set theory [6]. A reduct is the reduced information system consisting of set of attributes R which is independent of original information system with attribute set C such that no more attributes can be removed from R without any loss in information and $R \subseteq C$. With reduct, useful patterns residing in data can be discovered and decision rules can be generated. These rules are results of hidden relations between the objects. Decision rules can be defined as: IF (conditional attributes) THEN (Decision class). These rules are useful in predicting decision class label for future data processing. With fast growing rate of data produced by organizations finding useful pattern and maintaining decision rules as knowledge is very helpful to store valuable information. These rough sets can be used in wide range of applications in different fields like economic, banking [7, 8], pharmacology [9], and text mining [10]. In this paper we have conducted experiment on data of Portuguese banking institution provided by UCI machine

*For correspondence

learning repository, containing information about their customers, 80% data is considered for training as information system with set of eight different attributes, a reduct is calculated from which discernibility matrix and decision rules are generated. The decision rules are verified on test data set which predicts decision class proving that rough sets are very useful in finding hidden pattern in large data sets.

The cost for implementing RSD depends on finding reducts and decision rules. If M denote the total number of attributes, n denote the total number of instances of data and k is the length of the reduct, then the time and space complexity is $O(mn^2)$ and time for generating decision rules is $O(km^2)$.

2. Rough sets

Pawlak [5] introduced the concept of Rough Sets used to represent ambiguity, vagueness and general uncertainty. Rough sets are defined via equivalence relation based on lower and upper approximation. Rough sets can be used for feature selection, feature extraction, decision rule generation, and pattern extraction.

2.1 *Information system (IS) and decision system (DS)*

Any information system is denoted as pair (U, A) where

U is a non-empty finite set of objects.
A is a non-empty finite set of attributes such that, $a: U \rightarrow V_a$ for every $a \in A$, where V_a is called the value set of attribute 'a'.

Decision system is Information data along with a decision attribute d,
$T = (U, A \cup \{d\})$ where $d \notin A$, is the decision attribute which is not considered in A. The elements of A are called the condition attributes [11]. The conditional attributes gives related information of an object and decision attribute is the one we are interested to analyse.

2.2 *Indiscernibility relation*

For a set of attributes $B \subset A$, an indiscernibility relation IND(B) is defined in the following way:

Two objects X_i and X_j, indiscernible by the set of attributes B in A, if $b(X_i) = b(X_j)$ for every attribute $b \in B$.

Let $T = (U, A)$, $B \subseteq A$ and $X \sqsubseteq U$, we can approximate X using only the information contained in B by constructing the B-lower and B-upper approximations of X, denoted $\underline{B}\,X$ and $\bar{B}\,X$ respectively, where

$$\underline{B}X = \{X/[X]_B \subseteq X\}$$

$$\bar{B}X = \{X/[X]_B \cap X \neq \phi\}$$

B-boundary region of X denoted as $BN_B(X) = \bar{B}\,X - \underline{B}\,X$, consists of those objects that cannot be classified in B.

2.3 *Attributes dependency*

An information system represented by

$S = (U, Q, V, F)$, where U is the universal set of objects $\{X_1, X_2, X_3, \ldots X_n\}$
Q represents set of all attributes
$F: U * Q \rightarrow V$ is the decision function
'A' subset of attributes such that $A \subseteq Q$ defines an indiscernibility relation on U.
IND $(A) = \{X, Y \in U:$ for all $a \in A, f(X, a) = f(Y, a)\}$

Decision table DT $= (U, C \cup D, V, F)$, among all the attributes C represents set of conditional attributes and D represents decision attributes. A set of attributes D said to be depending on set of attributes C where all the attribute values of D can be determined from C. Dependency between D and C can be defined with a degree k as $C \rightarrow_k D$

$$\text{where } k = \gamma(C, D) = \frac{|POS_c(D)|}{|U|}.$$

The positive region with respect to set of attributes C defined as

$$POS_c(D) = \bigcup_{x \in U/D} \underline{C}X$$

The coefficient $k(0 \leq k \leq 1)$ shows the ratio of all elements classified into blocks of the partition U/D.

2.4 *Significance of attributes*

It will be useful to calculate relevant features by eliminating some of the attributes without losing information contained in the information system [12].

Significance of an attribute 'a' in a decision table $A = (U, C \cup D)$ is evaluated by measuring the effect of removing an attribute $a \in C$ from the attribute set C on the positive region defined by the set A. We consider the difference between $\gamma(C, D)$ and $\gamma(C \leftrightarrow \{a\}, D)$ as significance of an attribute 'a' as

$$\sigma(C, D)(a) = \frac{(\gamma(C, D) \leftrightarrow \gamma(C \leftrightarrow \{a\}, D))}{\gamma(C, D)} = 1$$

So the coefficient $\sigma(C, D)(a)$ can be represented as the error of classification when attribute 'a' is not taken. The significance coefficient can be extended to the set of attributes as follows:

$$\sigma(C,D)(B) = \frac{(\gamma(C,D) \leftrightarrow \gamma(C \leftrightarrow B, D))}{\gamma(C,D)} = 1$$

where B is subset of C.

If B is the reduct of C, then $\sigma(C \leftrightarrow B) = 0$, i.e., removing any reduct complement from the set of conditional attributes enables us to make decisions with certainty.

Any subset B of C can be treated as an approximate reduct of C, with error denoted as $\varepsilon(B)$, will be called an error of reduct approximation.

$$\varepsilon(C,D)(B) = \frac{(\gamma(C,D) \leftrightarrow \gamma(B,D))}{\gamma(C,D)}$$

It expresses how exactly the set of attributes B approximates the set of condition attributes C. The idea of an approximate reduct can be useful in those cases when a smaller number of condition attributes is preferred over the accuracy of classification on training data.

2.5 *Computation of reducts*

A reduct is defined as a minimal subset R of the attribute set C such that for given set of attributes D, $\gamma_R(D) = \gamma_C(D)$ where no more attributes can be removed from R. A given data set may have many attribute reduct sets [13, 14], known as actual reducts.

Unlike actual reducts, approximate reducts are also used in which reducts are obtained with a degree of approximation. The approaches to obtain approximate reducts are attribute weighting [15] and Hash based approximation [16]. In attribute weighting the attributes with high frequency in discernibility matrix are considered to be more significant. When two attributes have same number of occurrences then the value cardinality of attribute is used as supplementary weight where attribute with less cardinality is considered to be more significant. Hash based approximate reduct generation does not consider discernibility matrix it uses a *dHash* function to obtain inconsistent categories with a set of attributes. *dHash* takes 1st attribute and obtains inconsistent instances based on attribute values and passes to next attribute where new inconsistent categories are generated. These operations are executed iteratively until the degree of approximation is achieved by those set of attributes.

Computing approximate reducts takes $O(mn^2)$ and $O(m^2 n \log n)$ time to generate reduct, where m is total number of attributes and n is total number of instances. These approximate reducts are used when a degree of approximation can be allowed, resulting in approximate decision rules where actual reducts are used for generating strict or consistent rules.

The Reduct algorithm given below calculates the minimal reduct by generating all possible feature sets. The algorithm starts by considering an empty set R and an attribute is added to set R. Its significance is calculated if the dependency of the reduct set equals to the dependency of conditional attribute such that $R \subseteq C$, the reduct is returned otherwise another attribute is added iteratively until the dependency of the reduct set equals to the dependency of conditional attribute.

Procedure REDUCT (*C*, *D*)

INPUT: *C*, the set of all conditional features and *D*, the set of decision features.

OUTPUT: *R*, the reduct set of attributes.

(1)	Initialize R with a null set $R \leftarrow \{\}$
(2)	Repeat
(3)	Assign set R to T, $T \leftarrow R$
(4)	For each attribute from set C, $\forall X \in (C - R)$
(5)	Compare degree of dependency with assumed reduct set T. If $\gamma_{R \cup \{X\}}(D) > \gamma_T(D)$
(6)	If its degree of dependency is high then include that attribute in to set T, $T \leftarrow R \cup \{X\}$
(7)	Make set T as new reduct, $R \leftarrow T$
(8)	Until reduct R can define the decision class, $\gamma_R(D) = \gamma_C(D)$
(9)	Return R

From the above procedure the degree of dependency of each attribute is calculated and attributes with high degree is added to reduct R until the dependency of the reduct obtained equals the dependency of entire features in data set

2.6 *Discernibility matrix*

A discernibility matrix of decision table $(\cup, C \cup D)$ is a symmetric matrix of size $|U| \times |U|$ with entries defined by

$$S_{i,j} = \{a \in c | a(X_i) \neq a(X_j)\} \quad \text{where } i,j = 1, \ldots |U|.$$

where $S_{i,j}$ contains those attributes that differ between objects i and j.

For finding reduct, we have to find the discernibility matrix shown in table 1. For example, cell corresponding to objects 0 and 1 it shows on what attributes they differ like a, b, c and d. Discernibility matrix is symmetric i.e., although some attributes in objects 1 and 0 differ, their

Table 1. Shows sample discernibility matrix.

$X \in U$	0	1	2	3	4
0					
1	*a, b, d*				
2	*c, d*	*a, b, c*			
3	*b, c*	φ	*a, b, d*		
4	*d*	*a, c, d*	φ	*b, c, d*	

corresponding values are same as referred to position 0 and 1 so no need to specify value in discernibility matrix. By grouping all the entries forms the core for the reducts.

Here C represents disjunction of discernible attributes between every pair of objects.

$$C_{i,j} = \{a_1 \vee a_2 \vee a_3 \vee \ldots a_n\} \quad \text{where}$$
$$S_{i,j} = \{a_1, a_2, a_3 \ldots a_n\}.$$

A discernibility function f_D is a Boolean function of variables with conjunction over all elements $C_{i,j}$ where $1 \leq i < n$ and $j < i$.

$$f_D = \{C_{1,0} \wedge C_{2,0} \wedge C_{2,1} \wedge C_{3,0} \ldots C_{n,n-1}\}.$$

For above sample discernibility matrix,

$S_{1,0} = \{a,b,d\}, S_{2,0} = \{c,d\}, S_{2,1} = \{a,b,c\}, S_{3,0} = \{b,c\},$
$S_{3,1} = \{\},$
$S_{3,2} = \{a,b,d\}, S_{4,0} = \{d\}, S_{4,1} = \{a,c,d\}, S_{4,2} = \{\},$
$S_{4,3} = \{b,c,d\}.$
$C_{1,0} = \{a \vee b \vee d\}, C_{2,0} = \{c \vee d\}, C_{2,1} = \{a \vee b \vee c\},$
$C_{3,0} = \{b \vee c\},$
$C_{3,1} = \{\}, C_{3,2} = \{a \vee b \vee d\}, C_{4,0} = \{d\},$
$C_{4,1} = \{a \vee c \vee d\}, C_{4,2} = \{\}, C_{4,3} = \{b \vee c \vee d\}.$

The discernibility function,

$$\begin{aligned}
f_D &= \{C_{1,0} \wedge C_{2,0} \wedge C_{2,1} \wedge C_{3,0} \wedge C_{3,1} \wedge C_{3,2} \\
&\quad \wedge C_{4,0} \wedge C_{4,1} \wedge C_{4,2} \wedge C_{4,3}\} \\
&= (a \vee b \vee d) \wedge (c \vee d) \wedge (a \vee b \vee c) \wedge (b \vee c) \\
&\quad \wedge (d) \wedge (a \vee c \vee d) \wedge (b \vee c \vee d)
\end{aligned}$$

Further simplification can be performed by removing those clauses that are subsumed by others:

$$f_D(a,b,c,d) = (b \vee c) \wedge (d).$$

The reducts of the data set may be obtained by converting the above expression from conjunctive normal form to disjunctive normal form. Hence, the minimal reducts are $\{b, d\}$ and $\{c, d\}$.

3. Proposed rough set based decision rule generation (RSD) method

The procedure for computing reduct and decision rules is explained as follows:

1. A data set is considered which contains information about customers like age, balance, marital status, etc. This data is raw data of the customer.

$|D|$ defines the size of the data set

2. An Information System or Table is constructed by considering each customer as an object and their information as attributes Information system is represented as $\{U, A\}$ where, set of objects represented by $U = \{u_1, u_2, u_3, u_4 \ldots u_n\}$, say u_1 represents customer 1 considered for analysis and $A = \{r_1, r_2, r_3, r_4 \ldots r_n\}$, representing set of attributes for a customer.

3. A discernibility matrix among all customers is generated as $\text{IND}_{(n \times n)}$ where n represent the data size,

$$S = \text{DISCERNIBLE}(\text{IND}_{j,k})$$

$\text{IND}_{j,k}$ is a cell in $\text{IND}_{(n \times n)}$ representing relation between two particular objects j and k, where S represents attributes or properties on which they are discernible.

4. From the matrix, if A represent set of all attributes and sub_i represent possible subset of attributes then the Reduct R is found such that $R \, \varepsilon \, sub_i$ with $\gamma_R(D) = \gamma_A(D)$ where γ_R, γ_A represent degree of dependency of decision class D with respect to attribute set R, A respectively.

5. Final part in rough set analysis is generating decision rules; based on the reduct R. Decision rules represent the knowledge generated by the reduct.

Decision rules $(\text{RULES}) = \text{COMPARE} \, (\bigvee_n^{i=1} D_i)$ where I represents set of decision classes in data. $D_{i=\text{yes}}$ and $D_{i=\text{no}}$ represents sets of data with decision class yes or no respectively. COMPARE function is used to find discernible relation between set of objects belonging to different decision class.

$$\text{COMPARE}\left(\bigvee_{i=1}^n D_i\right) = \text{Discernible}(V_j D_j, V_k D_k)$$
$$\forall j, k \in i$$

The Discernible defines set of attributes on which V_j and V_k differ where V_j represents set of objects with decision class 'j' and V_k represents set of objects with decision class 'k'.

A decision rule is logical statement of type: "IF (conditions) THEN element belongs to DECISION CLASS". Using these decision rules we can define decision attribute of unclassified data.

For example consider the following table containing information of bank account holders as objects in decision system to generate decision rules. Here Id is used to identify objects uniquely. Reduct set has Age, Income and Student as the conditional attributes and Credit_rating as decision attribute with class value 'fair' and 'excellent'.

Id	Age (A)	Income (I)	Student (S)	Credit_rating (decision class)
a_1	<30	High	No	Fair
a_2	30–40	High	No	Excellent
a_3	30–40	Medium	Yes	Fair
a_4	<30	High	Yes	Excellent
a_5	30–40	Medium	No	Excellent

The following table shows for each object what are the objects with different decision class. Consider a_1 and we want to find out with what objects of other decision class a_1 is discernible. Here the decision class of a_1 is 'fair' comparing it with objects of decision class 'excellent' a_2, a_4 and a_5 they differ by attributes {Age}, {student} and {Age, Income} respectively.

Id	Discernible decision
a_1	a_2, a_4, a_5
a_2	a_1, a_3
a_3	a_2, a_4, a_5
a_4	a_1, a_3
a_5	a_1, a_3

The Decision rules can be formed from objects of different decision class with respect to the attributes on which they differ as

$$f(p_1, \text{fair}) = \{(A) \wedge (S) \wedge (A \vee I)\}$$
$$= \{(A) \wedge (S)\}$$
$$= \{(\text{Age} < 30) \text{ and } (\text{student} = \text{no})\}$$
$$f(p_2, \text{excellent}) = \{(A) \wedge (I \vee S)\}$$
$$= (\text{Age} = (30-40)) \text{ and}$$
$$(\text{Income} = \text{high or Student} = \text{no})\}$$
$$f(p_3, \text{fair}) = \{(I \vee s) \wedge (A \vee I) \wedge S\}$$
$$= \{(S) \wedge (A \vee I)\}$$
$$= \{(\text{Student} = \text{yes}) \text{ and}$$
$$(\text{Age} = 30-40 \text{ or Income} = \text{medium})\}$$
$$f(p_4, \text{excellent}) = \{(S) \wedge (A \vee I)\}$$
$$= \{(\text{Student} = \text{yes}) \text{ and}$$
$$(\text{Age} = 30 \text{ or Income} = \text{high})\}$$
$$f(p_5, \text{excellent}) = \{(A \vee I) \wedge s\}$$
$$= \{(\text{student} = \text{no}) \text{ and}$$
$$(\text{Age} = (30-40) \text{ or Income} = \text{medium})\}$$

Rules obtained for two decision class are:
credit_rating = fair implies that

$$\{(\text{Age} < 30) \text{ and } (\text{student} = \text{no})\}$$
$$\{(\text{Student} = \text{yes}) \text{ and } (\text{Age} = 30-40 \text{ or Income} = \text{medium})\}$$

credit_rating = excellent implies that

$$\{(\text{Age} = (30-40)) \text{ and } (\text{Income} = \text{high or Student} = \text{no})\}$$
$$\{(\text{student} = \text{yes}) \text{ and } (\text{Age} < 30 \text{ or Income} = \text{high})\}$$
$$\{(\text{student} = \text{no}) \text{ and } (\text{Age} = (30-40) \text{ or Income} = \text{medium})\}$$

4. System architecture

The system architecture is given in figure 1. At first, the training data is taken from the data set and given as input to information system to get discernibility matrix. The discernibility matrix is of order $m \times n$, where m denotes number of instances and n denotes number of attributes, with each cell representing set of attributes on which they differ.

The reducts are formed from discernibility matrix a subset of conditional attributes where dependency of decision class with given conditional attributes is equal to the dependency with attributes of reduct, based on the reduct set attributes the decision rules are generated.

5. Experimental results

The main objective of this study is to find out bank customers who are willing to make deposits, which reduces overhead for bank and helps to do campaigning on selected people to obtain better results. The data is provided by UCI Machine Learning Repository [17] obtained by conducting the marketing campaigns over Portuguese banking institution. The data set consists of 45,211 instances. The motive of the campaign is to find out if the bank term deposit would be subscribed or not. The data set is ordered by date (from May 2008 to November 2010). The data is defined with 16 attributes containing one decision attribute. Table 2 summarizes the variables and their data types that are self-explanatory to the customer.

These variables were chosen to create condition attributes of the information table used in this study. Information about each customer is described with these variables. In table 3 the C_1, C_2, C_3 and C_4 represent four customers and the value in the discernibility matrix represents on what properties they differ. Here in cell (C_2, C_1) defines customers C_1 and C_2 are distinguished with properties job and marital status.

Using the discernibility matrix reducts are obtained. By using reduct generation procedure two reducts are generated as
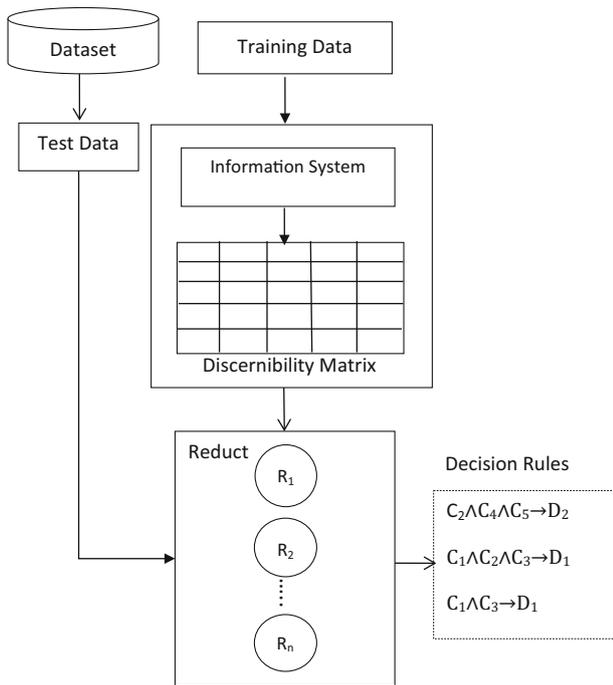
Figure 1. Architectural diagram of decision rule generation by RSD.

Table 2. Attributes and their data types.

Attribute	Data type
Age	Numeric
Job	Categorical: admin, unemployed, management, housemaid, entrepreneur, student, retired, technician, services
Marital status	Categorical: married, divorced, single
Education	Categorical: unknown, secondary, primary, tertiary
Credit	Has credit in default (binary 'yes' or 'no')
Balance	Average yearly balance (numeric)
Housing	Has housing loan (binary 'yes' or 'no')
Loan	Has personal loan (binary 'yes' or 'no')
Deposit	Has client subscribed a term deposit (binary 'yes' or 'no')

Table 3. Discernibility matrix of four customers.

Customers	C_1	C_2	C_3	C_4
C_1				
C_2	{job, marital status}			
C_3	{age, job}	{education, loan}		
C_4	{balance, job, age}	{loan, job}	{job, credit}	

Table 4. Decision rule set with decision class based on reduct R_1.

Decision rule	Decision class
If age = {20,21} AND job = student	Yes
(If age = 68 OR job = retired) AND (If age = 68 OR marital status = divorced)	Yes
(If age = 78 OR job = retired) AND (If age = 78 OR marital status = divorced)	Yes
If job = {management, services, retired, admin, self-employed, retired} AND age = {32,52,56,52,35,32}	Yes
IF age = {34,35,37} AND (job = {manager, technician} OR marital status = married)	Yes
IF age = {38,42,48} AND marital status = {married, divorced}	Yes
IF age = 29 OR marital status = single	Yes
IF marital status = single OR loan = yes	Yes

Table 5. Results obtained after validating on test data.

	Correct classification	
Reduct	Yes (%)	No (%)
R_1 = {age, job, marital status, education, balance, housing}	90.21	85.34
R_2 = {age, job, marital status, education, balance, loan}	83.33	80.33

R_1 = {age, job, marital status, education, balance, housing}

R_2 = {age, job, marital status, education, balance, loan}

In table 4, based on reduct R_1 a set of decision rules are generated and are listed.

Table 4 explains about the sample decision class for sample set of decision properties. To verify the decision rules we performed a cross validation using test data. The results are presented in table 5.

The proposed RSD method is trained with 45,200 instances producing two reducts, above results are obtained after verifying with test data of 4500 instances where 90% of the test data is classified accurately. These results prove that rough set methodology gives efficient results in the classification of data sets.

6. Conclusion

In this study we presented a Rough Set methodology determining decision rules helpful for predicting client's behaviour in taking a term deposit. It also describes a model for determining the variables that serve to explain the

decision class. In order to obtain acceptable percentages of correct classification using rough sets, there is no need to assume any type of prior statistical behaviour of the variables involved in the classification. Among the more significant advantages of this methodology is that, it considers only the significant attributes as reduct to generate decision rules. The decision rules are of more importance because they represent a language that is much closer to the one that is normally utilized by experts.

References

[1] Deger A and Adem A 2011 Bank specific and macroeconomic determinants of commercial bank profitability—empirical evidence from Turkey. *Bus. Econ. Res. J.* 2: 139–152

[2] Sergio M and Paulo C 2014 A data-driven approach to predict the success of bank telemarketing. *Decis. Support Syst.* 62(2014): 22–31

[3] Joaquin A, Rebecca M B and Frank P A C 2014 Classification with decision trees from a nonparametric predictive inference perspective. *Comput. Stat. Data Anal.* 71(2014): 789–802

[4] Mu-Yen C 2011 Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Syst. Appl.* 38(2011): 11261–11272

[5] Pawlak Z 2002 Rough set theory and its applications. *J. Telecommun. Inf. Technol.* 3: 7–10

[6] Eric C C T, Chen D and Daniel S Y 2008 Approximations and reducts with covering generalized rough sets. *Comput. Math. Appl.* 56(2008): 279–289

[7] Francis E H T and Lixiang S 2002 Economic and financial prediction using rough sets model. *Eur. J. Oper. Res.* 141(2002): 641–659

[8] You-Shyang C 2012 Classifying credit ratings for Asian banks using integrating feature selection and the CPDA-based rough sets approach. *Knowl. Based Syst.* 26(2012): 259–270

[9] Hsu-Hao Y and Chang-Lun W 2009 Rough sets to help medical diagnosis—Evidence from a Taiwan's clinic. *Expert Syst. Appl.* 36(2009): 9293–9298

[10] Duoqian M, Qiguo D and Hongyun Z 2009 Rough set based hybrid algorithm for text classification. *Expert Syst. Appl.* 36(2009): 9168–9174

[11] Zuqiang M and Zhongzhi S 2012 Extended rough set-based attribute reduction in inconsistent incomplete decision systems. *Inf. Sci.* 204(2012): 44–69

[12] Changzhong W, Mingwen S, Baiqing S and Qinghua H 2015 An improved attribute reduction scheme with covering based rough sets. *Appl. Soft Comput.* 26: 235–243

[13] Jia-yang W and Jie Z 2009 Research of reduct features in the variable precision rough set model. *Neurocomputing* 72: 2643–2648

[14] Malcolm B 2001 Reducts within the variable precision rough sets model A further investigation. *Eur. J. Oper. Res.* 134(2001): 592–605

[15] Al-Radaideh Q A, Sulaiman M N, Selamat M H and Ibrahim H 2005 Approximate reduct computation by rough sets based attribute weighting. *IEEE International Conference on Granular Computing*, vol 2, pp 383–386

[16] Wang P-C 2011 Efficient hash-based approximate reduct generation. *IEEE International Conference on Granular Computing*, Kaohsiung, pp 703–707

[17] Asuncion A and Newman D 2012 *University of California, School of Information and Computer Science, UCI machine learning repository*. Irvine