Rough Set Approach for Generation of Classification Rules of Breast Cancer Data

Aboul Ella HASSANIEN, Jafar M.H. ALI

Quantitative Methods and Information Systems Department
College of Business Adminstration, Kuwait University
P.O. Box 5969 Safat, Code no. 13060, Kuwait
e-mail: abo@cba.edu.kw

Abstract. Extensive amounts of knowledge and data stored in medical databases require the development of specialized tools for storing, accessing, analysis, and effectiveness usage of stored knowledge and data. Intelligent methods such as neural networks, fuzzy sets, decision trees, and expert systems are, slowly but steadily, applied in the medical fields. Recently, rough set theory is a new intelligent technique was used for the discovery of data dependencies, data reduction, approximate set classification, and rule induction from databases.

In this paper, we present a rough set method for generating classification rules from a set of observed 360 samples of the breast cancer data. The attributes are selected, normalized and then the rough set dependency rules are generated directly from the real value attribute vector. Then the rough set reduction technique is applied to find all reducts of the data which contains the minimal subset of attributes that are associated with a class label for classification. Experimental results from applying the rough set analysis to the set of data samples are given and evaluated. In addition, the generated rules are also compared to the well-known IDS classifier algorithm. The study showed that the theory of rough sets seems to be a useful tool for inductive learning and a valuable aid for building expert systems.

Key words: knowledge discovery, data mining, rough sets, breast cancer analysis, rule generation and reduction, decision trees, intelligent data analysis, rule classification, and computational intelligence.

1. Introduction

The growth of the size of data and number of existing databases far exceeds the ability of humans to analyze this data, which creates both a need and an opportunity to extract knowledge from databases (Cios *et al.*, 1998). Medical databases have accumulated large quantities of information about patients and their medical conditions. Relationships and patterns within this data could provide new medical knowledge. Analysis of medical data is often concerned with treatment of incomplete knowledge, with management of inconsistent pieces of information and with manipulation of various levels of representation of data. Existing intelligent techniques (Lavrajc *et al.*, 1997; Wolf *et al.*, 2000) of data analysis are mainly based on quite strong assumptions (some knowledge about

dependencies, probability distributions, large number of experiments), are unable to derive conclusions from incomplete knowledge, or can not manage inconsistent pieces of information. The most commonly intelligent techniques used in medical data analysis are neural network (Choi and Rockett, 2002; Setiono, 2000), Bayesian classifier (Cheeseman and Stutz, 1996), genetic algorithms (Grzymala-Busse *et al.*, 1999), decision trees (Hassanien, 2003), fuzzy theory (Parido and Bonelli, 1993). In this contribution the rough set theory is introduced. Rough set concept was introduced by Polish logician, Professor Z. Pawlak in early eighties (Pawlak, 1982; Pawlak, 1991; Pawlak *et al.*, 1995).

The theory of rough sets is a mathematical tool for extracting knowledge from uncertain and incomplete data based information. The theory assumes that we first have necessary information or knowledge of all the objects in the universe with which the objects can be divided into different groups. If we have exactly the same information of two objects then we say that they are indiscernible (similar), i.e., we cannot distinguish them with known knowledge. The theory of RS can be used to find dependence relationship among data, evaluate the importance of attributes, discover the patterns of data, learn common decision-making rules, reduce all redundant objects and attributes and seek the minimum subset of attributes so as to attain satisfying classification. Moreover, the rough set reduction algorithms enable to approximate the decision classes using possibly large and simplified patterns (Grzymala-Busse *et al.*, 1999; Kent, 1994; Lin and Cercone, 1997; Ning *et al.*, 1995; Zhong and Skowron, 2000; Polkowski and Skowron, 1998a; Polkowski and Skowron, 1998b).

This theory become very popular among scientists around the world and the rough set is now one of the most developing intelligent data analysis. Unlike other intelligent methods such as fuzzy set theory, Dempster–Shafer theory or statistical methods, rough set analysis requires no external parameters and uses only the information presented in the given data (Ning *et al.*, 1995; Slowinski, 1995; Ziarko, 1999).

This paper discusses how rough set theory can be used to analysis medical data, and for generating classification rules from a set of observed samples of the breast cancer data. The rough set reduction technique is applied to find all reducts of the data which contains the minimal subset of attributes that are associated with a class label for classification.

This paper is organized as follows. The characteristics of breast cancer data and knowledge representation are discussed in Section 2. Theoretical aspects of rough set data analysis, which are relevant to the work and rule generation algorithm are introduced in Sections 3 and 4, respectively. Experimental results and discussion are reported in Section 5. Comparison with decision trees classifier algorithm is given in Section 6. Finally, conclusion is discussed in Section 7.

2. Data Collection and Knowledge Representation

2.1. *Characteristics of Breast Cancer Data*

The data sets used in our experiments consists of 360 samples taken from fine needle aspirates from human breast tissue. They have been collected by Dr. W. Wolberg at the

Table 1

Condition and decision attributes of breast cancer dataset

Label	Attribute	Domain
A_1	Clump thickness	1–10
A_2	Uniformity of cell size	1–10
A_3	Uniformity of cell shape	1–10
A_4	Marginal adhesion	1–10
A_5	Single epithelial cell size	1–10
A_6	Bare nuclei	1–10
A_7	Bland_Chromatin	1–10
A_8	Normal Nucleoli	1–10
A_9	Mitoses	1–10
$d = A_{10}$	Class	Benigm–Malignant

university of Wisconsim. Each sample consists of nine measurement or features along with a label that denotes its class. Each instance has one of two possible classes: benign or malignant. These features has integer values in the range 1 to 10 as shown in Table 1.

2.2. *Information Systems*

Knowledge representation in rough sets is done via information systems, which are a tabular form of an OBJECT → ATTRIBUTE VALUE relationship. More precisely, an information system, $\Gamma = <U, \Omega, V_q, f_q>_{q \varepsilon \Omega}$, where

U is a finite set of objects, $U = \{x_1, x_2, x_3, \ldots, x_n\}$;

Ω is a finite set of attributes (features), the attributes in Ω are further classified into disjoint condition attributes A and decision attributes D, $\Omega = A \cup D$;

For each $q \in \Omega$,
- V_q is a set of attribute values for q,
- Each f_q: $U \rightarrow V_q$ is an information function which assigns particular values from domains of attributes to objects such that $f_q(x_i) \in V_q$ for all $x_i \in U$ and $q \in \Omega$.

An example of a breast cancer information system is presented in Table 2. Each sample is a patient described in terms of the attributes $\{A_1, A_2, \ldots, A_9\}$.

3. Rough Set Theory: Basic Concepts

The theory of rough sets (Pawlak, 1982; Pawlak *et al.*, 1995; Polkowski and Skowron, 1998b) has emerged as a major mathematical tool for managing uncertainty that arises from granularity in the domain of discourse. For example, if objects are patients suffering from a certain disease, symptoms of the disease form information about patients.

Table 2

Information system for breast cancer dataset

Object	Condition attributes									Decision Class $A_{10} = d$
	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	
x_1	5	3	3	3	2	3	4	4	1	Malignant
x_2	1	1	1	1	2	3	3	1	1	Benign
x_3	8	7	5	10	7	9	5	5	4	Malignant
x_4	7	4	6	4	6	1	4	3	1	Malignant
x_5	4	1	1	1	2	1	2	1	!	Benign
x_6	4	1	1	1	2	1	3	1	1	Benign
x_7	10	7	7	6	4	10	4	1	2	Malignant
x_8	6	1	1	1	2	1	3	1	1	Benign
x_9	7	3	2	10	5	10	5	4	4	Malignant
x_{10}	10	5	5	3	6	7	7	10	1	Malignant
x_{11}	3	1	1	1	2	1	2	1	1	Benign
x_{12}	8	4	5	1	2	4	7	1	1	Malignant
x_{13}	1	1	1	1	2	1	3	1	1	Benign
x_{14}	5	2	3	4	2	7	3	6	1	Malignant
x_{15}	3	2	1	1	1	1	9	1	1	Benign

360 sample of breast cancer data

Objects characterized by the same information are indiscernible (similar) in view of the available information about them. The indiscernibility relation generated in this way is the mathematical basis of rough set theory. A fundamental principle of a rough set based learning system is to discover redundancies and dependencies between the given features of a problem to be classified. It approximates a given concept below and from above, using lower and upper approximations. Consequently, a rough set learning algorithm can be used to obtain a set of rules in IF-THEN form, from a decision table.

In this section, let us present some preliminaries of rough set theory, which are relevant to this work. For details one may refer to (Pawlak, 1982; Pawlak *et al.*, 1995; Polkowski and Skowron, 1998b; Stefanowski, 1993).

3.1. *Indiscernibility Relation*

The mathematical machinery of rough sets is derived from the assumption that granularity can be expressed by partitions and their associated equivalence relations on the set of objects, it is called indiscernibility relations.

With respect to a given $q \in \Omega$, the functions partitions the universe into a set of pairwise disjoints subsets of U:

$$R_q = \{x: \ x \in U \wedge f(x, q) = f(x_0, q) \ \forall x_0 \in U\}. \tag{1}$$

Assume a subset of the set of attributes, $P \subseteq A$. Two samples x and y in U are indiscernible with respect to P if and only if $f(x, q) = f(y, q) \; \forall q \in P$. The indiscernibility relation for all $P \subseteq A$ is written as $IND(P)$. $U/IND(P)$ is used to denote the partition of U given $IND(P)$ and is calculated as follows:

$$U/IND(P) = \otimes \{q \in P: U/IND(P)(\{q\})\}, \tag{2}$$

where

$$A \otimes B = \{X \cap Y: \forall q \in A, \; \forall Y \in B, X \cap Y \neq \{\}\}. \tag{3}$$

3.2. *Lower and Upper Approximations*

A rough set approximates traditional sets using a pair of sets named the lower and upper approximation of the set. The lower and upper approximations of a set $P \subseteq U$, are defined by equations (4) and (5), respectively.

$$\underline{P}Y = \bigcup \{X: X \in U/IND(P), X \subseteq Y\}, \tag{4}$$

$$\overline{P}Y = \bigcup \{X: X \in U/IND(P), X \cup Y \neq \{\}\}. \tag{5}$$

Assuming P and Q are equivalence relations in U, the important concept positive region $POS_P(Q)$ is defined as:

$$POS_P(Q) = \bigcup_{X \in Q} \underline{P}X. \tag{6}$$

A positive region contains all patterns in U that can be classified in attribute set Q using the information in attribute set P.

DEFINITION 1 (Degree of dependency). The degree of dependency $\gamma(P, Q)$ of a set P of attributes with respect to a set Q of class labeling is defined as:

$$\gamma(P, Q) = \frac{|POS_P(Q)|}{|U|}, \tag{7}$$

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

The degree of dependency provides a measure of how important P is in mapping the dataset examples into Q. If $\gamma(P, Q) = 0$, then classification Q is independent of the attributes in P, hence the decision attributes are of no use to this classification. If $\gamma(P, Q) = 1$, then Q is completely dependent on P, hence the attributes are indispensable. Values $0 < \gamma(P, Q) < 1$ denote partial dependency, which shows that only some of the attributes in P may be useful, or that the dataset was flawed to begin with. In addition, the complement of $\gamma(P, Q)$ gives a measure of the contradictions in the selected subset of the dataset.

3.3. *Rough Set Attribute Reduction*

In an information system there often exist some condition attributes that don not provide any additional information about the objects in U. So, we should remove those attributes since the complexity and cost of decision process can be reduced if those condition attributes are eliminated (Bazan *et al.*, 1994; Kryszkiewicz and Rybinski, 1996a; Kryszkiewicz and Rybinski, 1996b).

DEFINITION 2 (Reduct). Given a classification task mapping a set of variables C to a set of labeling D, a reduct is defined as any $R \subseteq C$, such that $\gamma(C, D) = \gamma(R, D)$.

DEFINITION 3 (Reduct Set). Given a classification task mapping a set of variables C to a set of labeling D, a reduct set is defined with respect to the power set $\mathbf{P}(C)$ as the set $R \subseteq \mathbf{P}(C)$ such that $R = \{A \in \mathbf{P}(C) \colon \gamma(A, D) = \gamma(C, D)\}$. That is, the reduct set is the set of all possible reducts of the equivalence relation denoted by C and D.

DEFINITION 4 (Significance). Given P, Q and an object $x \in P$, the significant $\sigma_x(P, Q)$ of x in the equivalence relation denoted by P and Q is $\sigma_x(P, Q) = \gamma(P, Q) - \gamma(P - \{x\}, Q)$.

DEFINITION 5 (Minimal Reduct). Given a classification task mapping a set of variables C to a set of labeling D, and R, the reduct set for this problem space, a minimal reduct is defined as any reduct R such that $|R| \leqslant |A|, \forall A \in R$. That is, the minimal reduct is the reduct of least cardinality for the equivalence relation denoted by C and D.

4. Generating Decision Rule Algorithm

Rule generation is a crucial task in any learning system (Kryszkiewicz and Rybinski, 1996b; Lavrajc *et al.*, 1997; Lin and Cercone, 1997). In this section, we describe how decision rules are generated based on the reduct system obtained from subsection 3.3. The relation among the equivalence relations associated with attribute sets can be used to generate decision rules. Suppose that we have a set $Q = \{qi, qi, \ldots, q_n\}$ of independent attributes and a single dependent attribute d. There is no restriction of generality, since we are using only the partition information of θ_d, and thus d can be a composite attribute obtained from some $P \subseteq \Omega$.

Suppose that the partition induced by θ_Q is $\{X_1, X_2, \ldots, X_s\}$, and the one induced by θ_d is $\{Y_1, Y_2, \ldots, Y_t\}$. With each X_i we associate the set $M_i = \{Y_j \colon X_i \cap Y_j \neq \Phi\}$. Since the sets $Y_1, Y2, \ldots, Y_t$ partition U, we get:

$$\text{If } x \in X_i, \text{ then } x \in Y_{j_1} \text{ or } \ldots \text{ or } x \in Y_{j_{i(j)}}. \tag{8}$$

Each class X_i of θ_Q corresponds to a feature vector $(a_i)_{1 \leqslant i \leqslant n}$, where $x \in X_i$ if and only if $f_{q_1} = a_1$ and $\ldots$ and $f_{q_n}(x) = a_n$, similarly, $x \in Y_j$ if and only if $f_d(x) = b_j$ for

some $b_j \in V_d$. Eq. 6 leads to a rule of the form:

$$
\begin{aligned}
&\text{If } \ f_{q_1}(x) = a_1 \ \text{ and } \ \dots \ \text{ and } \ f_{q_n}(x) = a_n \\
&\qquad \text{then } \ f_d(x) = b_{j_1} \ \text{ or } \ \dots \ \text{ or } \ f_d(x) = b_{j_{i(j)}}.
\end{aligned}
\tag{9}
$$

In rough set theory, we distinguish between two kinds of rules: determinisic rules and non-determinisic rules: If some class X_i of θ_Q intersects exactly one Y_j, then $X_i \subseteq Y_j$, and the value of d of any $x \in X_i$ is uniquely determined. Otherwise $f_d(x)$ may be in any class contained in M_i and we have a proper disjunction on the right hand side of Eq. 9. A class X_i is called determinisic if it is contained in some Y_j, otherwise we call it indeterminisic. If all classes X_i are determinisic then $\theta_Q \subseteq \theta_d$ and d is dependent on Q.

EXAMPLE 4.1. Table 3 represents a sample of the data set as an example in order to extract the rules. Let $A = \{A_1, A_2\}$ be the set of condition attributes and d is the decision attribute. Let $X_1 = \{x_1, x_3, x_9\}$, $X_2 = \{x_2, x_4\}$, $X_3 = \{x_5, x_6\}$, $X_4 = \{x_7, x_8\}$, $Y_1 = \{x_1, x_2, x_3, x_4, x_7, x_9\}$, and $Y_2 = \{x_5, x_6, x_8\}$. From the previous section we get that X_1, X_2, X_3 are deterministic classes.

The following deterministic deicision rules are extracted:

$$
\left.
\begin{aligned}
&\text{If } \ A_1 = 1 \text{ and } \ A_2 = 1 \ => \ d = \text{Benign} \\
&\text{If } \ A_1 = 1 \text{ and } \ A_2 = 2 \ => \ d = \text{Benign} \\
&\text{If } \ A_1 = 2 \text{ and } \ A_2 = 1 \ => \ d = \text{Malignant}
\end{aligned}
\right\}
\tag{10}
$$

4.1. *Simplification Rule Algorithm*

The process by which the maximum number of condition attribute values are removed without loosing essential information is called value reduction (Kryszkiewicz and Rybinski, 1994; Ning *et al.*, 1994; Starzyk, 2000) and the resulting rule is called maximally general or minimal length. Computing maximally general rules is of particular importance in knowledge discovery since they represent general patterns existing in the data.

Table 3

Sample data sets

U	A_1	A_2	D
x_1	1	1	Benign
x_2	1	2	Benign
x_3	1	1	Benign
x_4	1	2	Benign
x_5	2	1	Malignant
x_6	2	1	Malignant
x_7	2	2	Benign
x_8	2	2	Malignant
x_9	1	1	Benign

Input: A set of specific decision rules RULE
Output: A set of general rules GRULE
GRULE $\leftarrow \Phi$
$N \leftarrow |\text{RULE}|$
For $i = O$ to $N - 1$ do
      $r \leftarrow r_i$
      $M \leftarrow |r|$
      For $j = O$ to $M - 1$ do
          Remove the jth condition attribute a_j in rule r
          If r inconsistant with any rule $r_n \in \text{RULE}$ then
             Restor the dropped condition a_j
          end if
      end for
      Remove any rule $r^` \in \text{GRULE}$ that is logically included in rule r
      If rule r is not logically included in a rule
          $r^`$ GRULE then
          GRULE $\leftarrow r \cup$ GRULE
      end if
end for

Fig. 1. Simplification rule algorithm.

In this subsection, we discuss a method to simplify the generated decision rules by dropping some condition attributes. The proposed simplification rule generation algorithm is presented in Fig. 1.

The simplification rule algorithm initialize general rules GRULE to empty set and copies one rule $r_1 \in \text{RULE}$ to rule r. A condition is dropped from rule r, and then rule r is checked for decision consistency with every rule $r_j \in \text{RULE}$. If rule r is inconsistent, then the dropped condition is restored. This step is repeated until every condition of the rule has been dropped once. The resulting rule is the generalized rule. Before rule r is added to GRULE, it is checked for rule redundancy. If rule r is logically included in any rule $r_a \in \text{GRULE}$, rule r is discarded. If any rules in GRULE are logically included in rule r, these rules are removed from GRULE. After all rules in RULE have been processed, GRULE contains a set of general rules.

EXAMPLE 4.2. By applying the generalization rule algorithm on rules given in Eq. 8 associated with the attribute set $A = \{A_1, A_2\}$. We get the rules given in Table 4.

Table 4

Example of generated rules

A_1	A_2	D
1	–	Benign
–	2	Benign
2	–	Malignant

Table 4 represent the following rules

$$
\left.
\begin{aligned}
&\text{If } A_1 = 1 \implies d = \text{Benign} \\
&\text{If } A_2 = 2 \implies d = \text{Benign} \\
&\text{If } A_1 = 2 \implies d = \text{Malignant}
\end{aligned}
\right\}
\tag{11}
$$

4.2. *Significance Testing*

The rough set methods developed so far are not always sufficient for extracting rules from decision tables. The set of all decision rules generated from all conditional attributes can be too large and can contain many chaotic rules not appropriate for unseen object classification (Gunther and Ivo, 2000; Ivo and Gunther, 1997). A high approximation quality is not grantee the validity of the rule, so the approximation quality is not a suitable tool to measures the classification success.

For example, the rough set method discovers a rule $Q \to P$ which is based on only a small number of observations which one might call a casual rule. The approximation quality of the rule may be due to chance. Thus the validity of inference rules for classification may be due to chance validated by statistical techniques. To evaluate the validity of a rule, in this paper, we apply a statistical significance test of attributes.

Suppose we want to test the statistical significance of the rule $Q \to P$. Let Σ be the set of all permutations on U, $\sigma \in \Sigma$, we define an information function f_r by

$$
f_r^{\sigma(P)} =
\begin{cases}
f_r \sigma(x) & \text{if } r = P, \\
f_r(x) & \text{otherwise.}
\end{cases}
\tag{12}
$$

The resulting information system Γ_σ permutes the P-columns according to σ, while leaving the Q-columns constant. We let $\gamma(Q \to \sigma(P))$ be the approximation quality of the prediction of $\sigma(P)$ by Q in Γ_σ. In order to evaluate the strength of the prediction $Q \to P$, we use the permutation distribution $\{\gamma(Q \to \sigma(P)): \sigma \in \Sigma\}$. The value $\varrho(\gamma(Q \to P) \backslash H_0)$ measures the extremeness of the observed approximation quality and it is defined by:

$$
\varrho\left(\gamma(Q - P)\backslash H_0\right) := \frac{\left|\left\{\sigma \in \Sigma: \gamma(Q \to \sigma(P)) \geqslant \gamma(Q \to P)\right\}\right|}{|U|!}.
\tag{13}
$$

If $\alpha = \varrho(\gamma(Q \to P)\backslash H_0)$ is less than 5%, then the rule $Q \to P$ is deemed significant, and the null hypothesis $Q \to P$ is due to chance can be rejected. Otherwise, we call it casual.

EXAMPLE 4.3. Table 5 shows the approximation qualities and the significance testing of the sets $\{A_1\}, \{A_2\}, \{A_1, A_2\}$ for the prediction of (d) for the information system given in Table 3.

From Table 5 we see that the best approximation quality is attained by the combination of both predicting attributes $\{A_1, A_2\}$. However, in terms of statistical significance

Table 5

Approximation quality and significance of predicting attributes

Attribute set	γ	Significance	Interpretation
$\{A_1\}$	0.556	0.047	Not casual
$\{A_2\}$	0.000	1.000	casual
$\{A_1, A_2\}$	0.778	0.144	casual

the set $\{A_1, A_2\}$ is not significant predictor for the outcome of d, because there is no evidence that the prediction success is not due to chance. Therefore, the set $\{A_1\}$ should be preferred to predict d, because it is unlikely that the prediction success is due to chance, although the approximation quality of $\{A_1\}$ is smaller than the approximation quality of $\{A_1, A_2\}$.

5. Experimental Results and Discussion

The aim of this experiment is to identify the essential subset of non-redundant attributes, which is relevant to determine the discovery task for a specific decision attribute. We try to extract relevant features from a given set of features with respect to the decision attribute class (d) as given in Table 2. The experiments come with some steps.

- Check the relevant and irrelevant attributes.
- Check dependencies between single attribute and a set of attributes.
- Choose reducts of the system.
- Extract the hidden rules.
- Generalize the rules in order to make it more understandable.
- Testing the significance of the rules generated.

• *Relevant and Irrelevant Attributes*

Table 6 represents the essential subset of non-redundant attributes, which is relevant to determine the discovery task for a specific decision attribute.

Table 6

Revealed attribute result

Case	Available attributes	Relevant attributes
1	A_4, A_5, A_7, A_9	A_4, A_7, A_9
2	$A_2, A_5, A_7, A_9, A_{10}$	A_2, A_5, A_7, A_9
3	A_2, A_3, A_5, A_6, A_7	A_2, A_3, A_5

Table 7

Data dependency

Case	Condition Attributes	Dependency Coefficient	Type of Dependency
1	A_2, A_3, A_4, A_{10}	1	Total
2	A_3, A_4, A_5, A_7	1	Total
3	A_2, A_6, A_7	0.992	Partial
4	A_5, A_6, A_8, A_9	0.754	Partial
5	A_2, A_{10}	0.308	Partial

● *Dependencies between Single Attribute and a Set of Attributes*

Table 7 represents the degree of dependency of a specific decision attribute on a given set of condition attributes in the target dataset. The degree of dependency ranges from the value zero to the value one. Zero value indicating that no dependency, one value for a total dependency and in between for a partial dependency. From Table 7, we see that the degree of dependency $\gamma(Q, d)$ of case (1) and case (2) is one. It means that this dependency is functional. Whereas, the degree of dependency of cases (3, 4 and 5) is less than one, it is only partially functional.

● *Reducts of the System*

Rough set data analysis was applied to the breast cancer data to find the reducts and core of the data. We have found that the full breast cancer data set of 360 patients have twenty seven reducts with respect to d. Table 8 represented a sample of result reducts which concentrate with no more than three elements. For each reduct we list the number of rules. Note that, the core of the system is empty. This signifies a huge inhomogeneity among the attributes, in other words the dependency among the attributes is high, and there are many possibilities for substitution.

Table 8

The set of reducts

Reduct	Rules	Reduct	Rules
A_2, A_3, A_5	25	A_2, A_5, A_8	36
A_3, A_7, A_9	18	A_2, A_6, A_8	32
A_2, A_3, A_8	24	A_2, A_8, A_9	33
A_2, A_4, A_6	33	A_3, A_4, A_7	20
A_2, A_4, A_8	33	A_3, A_5, A_7	L2J
A_4, A_7, A_{10}	21	A_4, A_7, A_9	21
A_4, A_6, A_7	23	A_7, A_8	23
A_3, A_6, A_7	22		

Table 9

Approximation qualities

Attributes	γ	Attributes	γ
A_2, A_3, A_4, A_{10}	1.00	A_2, A_6, A_{10}	0.93
A_2, A_3, A_5	1.00	A_2, A_3, A_8	1.00
A_3, A_7, A_9	1.00	A_2, A_4, A_5, A_9	1.00
A_2, A_6, A_7	0.992	A_5, A_6, A_{10}	0.94
A_5, A_6, A_8, A_9	0.754	A_2, A_{10}	0.308
A_4	0.785	A_3	0.879
A_2	0.266	A_5	0.085
A_{10}	0.055	A_7	0.145

● *The Approximation Qualities*

The approximation qualities of the nonempty attribute sets are given in Table 9 which represent part of the approximation qualities of subsets of attributes. We see that, uniformity of cell size as attribute has a high classification quality followed by uniformity of cell shape attribute A_4. Uniformity of cell size and uniformity of cell shape attributes together, account for 95% of all cases. Also, uniformity of cell size, uniformity of cell shape and bare nuclei can account attributes account for 100% of all cases. Also according to the size, we see that the attributes bare nuclei and bland_chromatin can account for 100% of all cases. The result clearly indicate that the combination of A_3 and any other attribute describing the breast cancer data have approximation quality greater than 90%. Also, according to γ and the size of attributes we see that A_7 and A_8 attributes are the best choice for classification.

● *Generalize and Extract a Simplified Rules*

Table 10 represent a part of the rules extracted from the breast cancer data. The total number of rules is equal to 428 rule and after simplification the number of rules is reduced to 30 rules.

Table 10

Breast cancer rules

Rule	Instances	Rule	Instances
$A_7 = 2 \wedge A_8 = 1 \Rightarrow d = 2$	2	$A_7 = 1 \wedge A_8 = 2 \Rightarrow d = 2$	69
$A_7 = 1 \wedge A_8 = 2 \Rightarrow d = 2$	69	$A_7 = 1 \wedge A_8 = 1 \Rightarrow d = 2$	87
$A_7 = 1O \Rightarrow d = 4$	33	$A_8 = 8 \Rightarrow d = 4$	9
$A_8 = 10 \Rightarrow d = 4$	14	$A_7 = 6 \Rightarrow d = 4$	2
$A_7 = 5 \wedge A_8 = 1 \Rightarrow d = 2$	2	$A_8 = 7 \Rightarrow d = 4$	18
$A_8 = 9 \Rightarrow d = 4$	5	$A_7 = 3 \wedge A_8 = 1 \Rightarrow d = 2$	1
$A_7 = 3 \wedge A_8 = 3 \Rightarrow d = 4$	1	$A_7 = 4 \wedge A_8 = 4 \Rightarrow d = 4$	1
$A_7 = 5 \wedge A_8 = 2 \Rightarrow d = 2$	2	$A_7 = 1 \wedge A_8 = 3 \Rightarrow d = 2$	23
$A_7 = 5 \wedge A_8 = 4 \Rightarrow d = 4$	2	$A_7 = 4 \wedge A_8 = 1 \Rightarrow d = 2$	4

We have test the significance of the rule in the full data. We found that no attribute set of reducts was casual with respect to (d), so that the rule derived from the data can be considered significance.

6. Comparson with Decision Tree Classifier

Intelligent methods such as neural networks, fuzzy sets, baysiean classifier, decision trees, and genetic algorithm has been successfully applied to analysis medical data. A comparison between the obtained results of applying rough set and IDS decision trees (Hassanien, 2003) has been made. There are several reasons to choose decision tree for such comparison. First, compared to other classifire, a decision tree is easily interpreted, comprehended by human. Second, while training neural networks can take large amounts of time and thousands of iterations, decision trees is efficient and is thus suitable for large training sets. Also, decision tree generation algorithms do not require additional information besides that already contained in the training data (e.g., domain knowledge or prior knowledge of distributions on the data or classes). The same as neural networks, decision trees display good classification accuracy.

Table 11 shows the comparison of rough sets and decision trees results. Rough set results were much better in terms of the number of rules and the classification accuracy. Moreover, in the decision tree classifier, more robust features are required to improve the performance of the decision tree classifier.

7. Conclusions and Future Work

In this paper, we present an intelligent data analysis approach for generating classification rules from a set of observed 360 samples of the breast cancer data. Statistical information is used to enhance the generated rules. In this paper, we show that the total number of generated rules is reduced from 428 to 30 rules after applying the proposed simplification algorithm. Consequently, more accurate rule sets can be extracted from the pruning simplification algorithm that achieves high accuracy rates. A comparison between the obtained results of rough set with the well-known ID3 decision tree classifier algorithm has been made. Rough sets showed a higher with higher overall accuracy rates and it generates more compact rules.

Table 11

Number of generated rules and accuracy results

Algorithm	Number of rules		Classification Accuracy
	Before pruning	After pruning (Simplification)	
Decision tree	1022	76	85.25%
Rough Set	472	30	98%

The integration of rough sets with other intelligent tools such as fuzzy sets and neural network for classification and rule generation in soft computing paradigm is the aim of our future work.

Acknowledgment

References

Bazan, J., A. Skowron and P. Synak (1994). Dynamic reducts as a tool for extracting laws from decision tables. In *Proc. of the Symp. on Methodologies for Intelligent Systems*, Charlotte, NC, October 16–19. *Lecture Notes in Artificial Intelligence*, vol. 869. Springer–Verlag, Berlin. pp. 346–355.

Cheeseman, P., and J. Stutz (1996). Bayesian classification (AutoClass): theory and results. In U.M. Fayyad, G. Piatetsky–Shapiro, P. Smyth and R. Uthurusamy (Eds.), *Advances in Knowledge Discovery and Data Mining*. AAAI Press/MIT Press.

Cios, K., W. Pedrycz and R. Swiniarski (1998). *Data Mining Methods for Knowledge Discovery*. Kluwer Academic Publishers.

Grzymala–Busse, J., Z. Pawlak, R. Slowinski and W. Ziarko (1999). Rough sets. *Communications of the ACM*, **38**(11).

Gunther, G., and D. Ivo (2000). *Statistical Techniques for Rough Set Data Analysis in Rough Sets: New Developments*. Physica–Verlag.

Hassanien, A.E. (2003). Classification and feature selection of breast cancer data based on decision tree algorithm. *International Journal of Studies in Informatics and Control Journal*, **12**(1), 33–39.

Ivo, D., and G. Gunther (1997). Statistical evaluation of rough set dependency analysis. *International Journal of Human-Computer Studies*, **46**, 589–604.

Kent, R.E. (1994). Rough concept analysis, rough sets, fuzzy sets knowledge discovery. In W.P. Ziarko (Ed.), *Proceedings of the International Workshop on Rough, Sets, Knowledge, Discovery*. Banff, Alta., Canada. Springer–Verlag. pp. 248–255.

Kryszkiewicz, M., and H. Rybinski (1994). Finding reducts in composed information systems, rough sets, fuzzy sets knowledge discovery. In W.P. Ziarko (Ed.), *Proceedings of the International Workshop on Rough Sets, Knowledge, Discovery*. Banff, Alta., Canada. Springer–Verlag. pp. 261–273.

Kryszkiewicz, M., and H. Rybinski (1996a). Attribute reduction versus property reduction. In *Fourth European Congress on Intelligent Techniques and Soft Computing*, *Proceedings*, (EUFIT96), Vol. I. September 2–5, Aachen, Germany. pp. 204–208.

Kryszkiewicz, M., and H. Rybinski (1996b). Reducing information systems with uncertain real value attributes. In *Sixth International Conferences, Informatio Processing and Management of Uncertainty in Knowledge-Based Systems*, *Proceedings* (IPMU'96), Vol. II. July 1–5, Grenada. pp. 1165.

Lavrajc, N., E. Keravnou and B. Zupan (1997). *Intelligent Data Analysis in Medicine and Pharmacology*. Kluwer Academic Publishers.

Lin, T.Y., and N. Cercone (1997). *Rough Sets and Data Mining*. Kluwer Academic Publishers.

Ning, S., H. Xiaohua, W. Ziarko and N. Cercone (1994). A generalized rough sets model. In *Proceedings of the 3rd Pacific Rim International Conference on Artificial Intelligence*, Vol. 431. Beijing, China. Int. Acad. Publishers. pp. 437–443.

Ning, S., W. Ziarko, J. Hamilton and N. Cercone (1995). Using rough sets as tools for knowledge discovery. In U.M. Fayyad, R. Uthurusamy (Eds.), *KDD'95 Proceedings First International Conference on Knowledge Discovery Data Mining*. Montreal, Que., Canada, AAAI. pp. 263–268.

Parido, A., and P. Bonelli (1993). A new approach to fuzzy classifier systems. In *Proceedings of the Fifth International Conference on Genetic Algorithms*. pp. 223–230.

Pawlak, Z. (1982). Rough sets. *Int. J. Computer and Information Sci.*, **11**, 341–356.

Pawlak, Z. (1991). *Rough Sets-Theoretical Aspect of Reasoning about Data*. Kluwer Academic Publishers.

Pawlak, Z., J. Grzymala–Busse, R. Slowinski, W. Ziarko (1995). Rough sets. *Communications of the ACM*, **38**(11), 89–95.

Polkowski, L., and A. Skowron (1998a). *Rough Sets in Knowledge Discovery*, Vol. 1/2. Studies in Fuzziness and Soft Computing series, Physica–Verlag.

Polkowski, L., and A. Skowron (1998b). *Rough Sets and Current Trends in Computing*, LNAI 1424, Springer.

Se-Ho, Ch., and P. Rockett (2002). The training of neural classifiers with condensed datasets. *SMCB*, **32**(2), 202–206.

Setiono, R. (2000). Generating concise and accurate classification rules for breast cancer diagnosis. *Artificial Intelligence in Medicine*, **18**(3), 205–219.

Starzyk, J.A., N. Dale and K. Sturtz (2000). A mathematical foundation for improved reduct generation in information systems. *Knoweldge and Information Systems Journal*, 131–147.

Slowinski, R. (1995). Rough set approach to decision analysis. *AI Expert*, March, 19–25.

Stefanowski, J. (1993). Classification support based on the rough sets. *Foundations of Computing and Decision Sciences*, **18**(3–4), 371–380.

Wolf, S., H. Oliver, S. Herbert and M. Michael (2000). Intelligent data mining for medical quality management. In *Proceedings of the Fifth International Workshop on Intelligent Data Analysis in Medicine and Pharmacology* (IDAMAP-2000). Berlin, Germany.

Zhong, N., and A. Skowron (2000). Rough sets in KDD: tutorial notes. *Bulletin of International Rough Set Society*, **4**(1/2).

Ziarko, W. (1999). Discovery through rough set theory. Knowledge Discovery: viewing wisdom from all perspectives. *Communications of the ACM*, **42**(11).

A.E. Hassanien received his BSc with honors in 1986 and MSc degree in 1993, both from Ain Shams University, Faculty of Science, Pure Mathematics and Computer Science Department, Cairo, Egypt. On September 1998, he received his doctoral degree from the Department of Computer Science, Graduate School of Science & Engineering, Tokyo Institute of Technology, Japan. He is an assistant professor at Cairo University, Faculty of Computer and Information Technology Department. Currently, he is a visiting professor at Kuwait University, College of Business Administration, Quantitative and Information System Department. His research interests include visualization, rough set, wavelet, mathematical morphology, fractal theory, computer animation, medical image analysis, multimedia data mining.

J.M.H. Ali received his BSc degree in 1989 and MSc degree in 1991 from Bradley University, Peoria, Illinois, United States, majoring in computer science. In 1995, he received his doctoral degree from the Department of Computer Science jointly with the Business Graduate School at Illinois Institute of Technology, Chicago, Illinois, United States. Currently, he is an associate professor and department chair in the Quantitative and Information System Department at College of Business, Kuwait University, Kuwait. His research interest includes neural networks, genetic algorithms, visualization, rough set theory, information systems, image analysis, data mining, e-commerce and e-government.

Grubių aibių analizė krūties vėžio duomenų sumažinto kiekio generavime

Aboul Ella HASSANIEN, Jafar M.H. ALI

Grubių aibių teorija panaudojama atskleidžiant duomenų priklausomybę, jų kiekio sumažinimo kriterijus, sprendžiant klasifikacijos uždavinius. Išanalizuoti krūties vėžio duomenys, panaudojant grubių aibių analizės metodus ir siekiant rasti duomenis su minimaliu kiekiu požymių. Atlikti skaitiniai eksperimentai parodė grubių aibių pranašumą.