Alexandria University

Faculty Of Science

The First One Day Workshop on Rough Sets and Their Applications

13, August, 2006, Alexandria, Egypt

ERS

Egyptian Rough Sets Group

WRSTA 2006

Proceeding of the First One Day Workshop on Rough Sets and Their Applications

Alexandria, EGYPT

August 13, 2006

Organizing and Steering Committees

Workshop Chairs

- Professor Mohamed Hassan Gabr, Alexandria University
- Professor Abdel-Badeeh M. Salem, Ain Shams University

Workshop co-Chairs

- A.M Kozae, Tanta University
- Yasser Fouad Mahmoud Hassan, Alexandria University
- Wael Abd El-Kader Awad Suez Canal University

Workshop Coordinator

- Aboul Ella Hassanien, Cairo University

Workshop Program Committee

- Professor Mahmoud Mohamed Hassan Gabr, Alexandria University
- Professor Aly Fahmy, Cairo University
- Professor A.M Kozae, Tanta University,
- Prof. Abdel-Badeeh M. Salem, Ain Shams University
- Dr. Yasser Fouad Mahmoud Hassan, Alexandria University
- Dr. Hala Shawky Own, NRIAG, Helwan
- Dr. Nahla El-haggar, NRIAG, Helwan
- Dr. Wael Abd El-Kader Awad Suez Canal University
- Dr. Amgad Salama Salem, Tanta University
- Prof.Dr. Farahat Farag Farhat,Sadat Academy
- Dr Hussam Elbehiery,Egyptian Armed Forces Research Center
- Dr. Tarek Gharib Fouad, Ain Shams University.
- Dr. Mohamed Abdel-Monsef Mohamed Ezzat, Tanta University

Content

Invited Speaker

Regular Papers

Session-I Theoretical Aspects

Session-II Models and Applications

Egyptian Rough Sets Working Group Members

- Professor Mahmoud Mohamed Hassan Gabr, Alexandria University
- Professor A.M Kozae, Tanta University,
- Prof. Abdel-Badeeh M. Salem, Ain Shams University
- Prof.essor Farahat Farag Farhat,Sadat Academy
- Dr. Aboul Ella Hassanien, Cairo University
- Dr. Yasser Fouad Mahmoud Hassan, Alexandria University
- Dr. Hala Shawky Own, NRIAG, Helwan
- Dr. Nahla El-haggar, NRIAG, Helwan
- Dr. Wael Abd El-Kader Awad Suez Canal University
- Dr. Amgad Salama Salem, Tanta University
- Dr Hussam Elbehiery,Egyptian Armed Forces Research Center
- Dr. Tarek Gharib Fouad, Ain Shams University.
- Dr. Mohamed Abdel-Monsef Mohamed Ezzat, Tanta University
- Mr. Mohamed Tahoun Mohamed (Student)
- **Mrs Wesam Fathy Jasser** (Student)
- **Mrs. Wedad Abdalla Sagar** (Student)
- Sanaa Abdallah Rashed (Student)
- Ahmed Abd el rahman eltahawy (Student)
- Soaad Abd El-Badie (Student)
- Eng. Tamer Medhat Mohamed "T. Medhat (Student)
- Wafaa Hassanein Anwar (Student)
- Neveen Kilany (Student)
- Omnia Gamal El-Dean Anwer El-Barbary (Student)
- Mostafa El-Gyar Atia (Student)

Egyptian Rough Computing Journal

The Egyptian Rough Computing Journal ***(ERCJ)*** is biannual and refereed journal published by the Egyptian rough sets working group. This journal provides a forum for the research and development in the fields of rough sets and their applications. It will publish in electroinc form (on line) and hard form. The ***ERCJ*** has as its principal aim the fostering of professional exchanges between Egyptian scientists and practitioners who are interested in the area of rough sets and their applications. This journal is devoted to the entire spectrum of issues related to rough sets, from logical and mathematical foundations, through all the aspects of rough set theory and its applications, such as data mining, knowledge discovery, and intelligent information processing, to relations between rough sets and other approaches to uncertainty, vagueness, and incompleteness, such as fuzzy sets and the theory of evidence.

Articles that appear in the ***ERCJ*** include foundations and applications of rough sets as well as foundations and applications of hybrid methods combining rough sets with other approaches important for development of intelligent systems.

Egyptian Rough Sets Working Group

The first one Day Workshop on Rough Sets and their Applications – WRSTA2006

Alexandria University, Faculty of Science, Mathematics Department, 13 August, 2006

New Approaches for Data Reduction in Generalized Multi-valued Decision Information System:

Case study of Rheumatic Fever Patients[1]

A.M. Kozae, M.M.E. Abd El-Monsef, and S. Abd El-Badie

Department of Mathematics, Faculty of Science,
Tanta University, Egypt

Abstract: A multi-valued information system (MIS) is a generalization of the idea of a single valued information system (SIS). In a multi-valued information system, attribute functions are allowed to map elements to sets of attribute values. In this paper, we initiated a new approach for data reduction in Generalized Multi–Valued Decision Information System (GMDIS). In the beginning we converted the Single-Valued Decision Information System (SDIS) by collecting the attributes to a GMDIS. Two general relations are defined on condition attributes and decision attribute. We constructed new classes using the general relations which are used for data reduction. The measure of decision dependency on the condition attributes is studied in our approach. To evaluate the performance of the approach, an application of, rheumatic fever datasets has been chosen and the reduct approach have been applied to see their ability and accuracy.

Keywords: Multi-Valued Information System, Rough Sets, Reduction

1 Introduction

The theory of rough sets (RS) [4,6,7] 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 [2,10] enable to approximate the decision classes using possibly large and simplified patterns.

This theory become very popular among scientists around the world and the rough set is now one of the most developing intelligent data analysis [11]. Unlike other intelligent methods such as fuzzy set theory, Dempster -- Shafer theory or statistical methods, rough set

[1] The authors greatly appreciate the valuable comments of Dr. K. E. Sturtz, Veridian, and the good advice of Prof. Dr. E. Rady and Dr. A. S. Salama.

analysis requires no external parameters and uses only the information presented in the given data.

According to rough set theory (RST) philosophy an element x of U is labeled as being in a set X called positive region, and some elements may be clearly labeled as not being in set X called negative region, limited information prevents us from labeling all possible cases clearly. The remaining elements cannot be distinguished and lie in what is known as the boundary region. The positive region in RST is known as lower approximation of set X, which yields no false positives. The positive region plus the boundary region make up an upper approximation of set X, which yields no false negatives. The practical end of rough sets deals with being continuous inputs and deciding which areas of the input space should be included in these various approximations but it can be viewed that most of the studies on rough sets applied on a tabular data related with equivalence relations. It is not easy to confirm a non-equivalence relation on any data table. So in the case of a multi-valued information system (MIS) applying a non-equivalence relation [6], binary non-equivalence relations was chosen from many kinds of non-equivalence relation mentioned in Orlowska book, 1998.

The objective of this paper is to initiate a new approach for data reduction in Generalized Multi–Valued Decision Information System by collecting the attributes to a GMDIS due to the converted Single-Valued Information System (SIS). If there exist an addition attribute D, then we call it a Single-Valued Decision Information System (SDIS).

This paper is organized as follows. In Section 2, we give an overview of the information system and approximation. The characteristic of Rheumatic Fever data is presented in section 3. Section 4, discusses the concept of mutli valued information system. Section 5, discusses the generalization of the multi valued information concept. Then a reduction approach is discussed.

2. Information Systems and Approximation

The starting point of rough set theory is the indiscernibility relation [7], generated by information about objects of interest. The indiscernibility relation is intended to express the fact that due to the lack of knowledge we are unable to discern some objects employing the available information. It means that, in general, we are unable to deal with each particular object but we have to consider clusters of indiscernible objects, as fundamental concepts of RST, now we present above consideration more formally

Suppose we have two finite, non-empty sets U and A, where U is the universe of objects, and A- a set of attributes. The pair (U, A) is an information system. With every attribute $a \in A$ we associate a set V_a, of its values, called the domain of a. Any subset B of A determines a binary relation R on U, called an indiscernibility relation, defined as follows:

$$xRy \text{ } iff \text{ } a(x) = a(y) \text{ } for \text{ } every \text{ } a \in B \tag{1}$$

Where $a(x)$ denotes the value of attribute a for the object x.

Obviously R is an equivalence relation. The family of all equivalence classes of R, i.e. the partition determined by B, will be denoted by U / R, an equivalence class of R, i.e. the block of the partition U / B, containing x will be denoted by $B(x)$.

If $(x, y) \in R$ it will be said that x and y are B-indiscernible. Equivalence classes of the relation R (i.e., blocks of the partition U / B) are referred to as B-elementary sets. In the rough set approach the elementary sets are the basic building blocks (concepts) of the knowledge about the considered problem. The unions of B-elementary sets are called B-definable sets.

The indiscernibility relation will be further used to define basic concepts of RST, this leads to the definition of the B-lower and the B-upper approximations respectively as follows,

$$\underline{B}(X) = \{x \in U : B(x) \subseteq X\} \tag{2}$$

$$\overline{B}(X) = \{x \in U : B(x) \cap X \neq \Phi\} \tag{3}$$

Assigning to every subset X of the universe U. The B-boundary region of X will defined by,

$$BN_B(X) = \overline{B}(X) - \underline{B}(X) \tag{4}$$

If the boundary region of X is the empty set , i.e., $BN_B(X) = \Phi$, then the set X is crisp (exact) with respect to B; in the opposite case , i.e. , if $BN_B(X) \neq \Phi$, the set X is referred to as rough (inexact) with respect to B. Also we can define the positive and negative regions of a set X ,

$$\begin{aligned} POS_B(X) &= \bigcup_{x \in U} \underline{B}(X) \quad , \quad \textit{The positive region of } X \\ NEG_B(X) &= U - \overline{B}(X) \quad , \quad \textit{The negative region of } X \end{aligned} \tag{5}$$

Rough set can be also characterized numerically by the following coefficient

$$\alpha_B(X) = \frac{|\underline{B}(X)|}{|\overline{B}(X)|}, \tag{6}$$

Where $\alpha_B(X)$ be the accuracy of approximation, where $|X|$ denotes the cardinality of X. Obviously $0 \le \alpha_B(X) \le 1$. If $\alpha_B(X) = 1$ then X is crisp with respect to B (X is precise with respect to B) , and otherwise , if $\alpha_B(X) \prec 1$ then X is rough with respect to B (X is vague with respect to B).

3. Rheumatic Fever Data: Characteristic

In this section, we briefly describe the Rheumatic Fever datasets used in this study. No doubt that the rheumatic fever is a very common disease and it has many symptoms differs from patient to another though the diagnosis is the same. So, we obtained the following example on seven rheumatic fever patients from Tanta University Hospital, Egypt. All patients are between 9-12 years old with history of Arthritis began from age 3-5 years. This disease has many symptoms and it is usually started in young age and still with the patient along his life.

Table (1) introduced the seven patients characterized by 8 symptoms (attributes) using them to decide the diagnosis for each patient (decision attribute). Table (1) represents the attributes as described in the rheumatic fever datasets.

Attribute Symbol	Refers to ?	Attribute Values	Refers to ?
S	*Sex*	s_1	*Male*
		s_2	*Female*
F	*Pharyngitis*	f_1	*Yes*
		f_2	*No*
A	*Arthritis*	a_0	*No arthritis*
		a_1	*Began in the knee*
		a_2	*Began in the ankle*
R	*Carditis*	r_1	*Affected*
		r_2	*Not affected*
K	*Chorea*	k_1	*Yes*
		k_2	*No*
E	*ESR*	e_1	*Normal*
		e_2	*High*
P	*Abdominal pain*	p_1	*Absent*
		p_2	*Present*
H	*Headache*	h_1	*Yes*
		h_2	*No*
D	*Diagnosis*	d_1	*Rheumatic arthritis*
		d_2	*Rheumatic carditis*
		d_3	*Rheumatic arthritis and carditis*

Table 1: Rheumatic Fever Data

History / *Patients*	**S**	**F**	**A**	**R**	**K**	**E**	**P**	**H**	**D**
x_1	s_2	f_1	a_1	r_1	k_1	e_1	p_1	h_2	d_3
x_2	s_1	f_1	a_1	r_1	k_1	e_2	p_1	h_1	d_3
x_3	s_2	f_1	a_2	r_1	k_2	e_1	p_1	h_2	d_3
x_4	s_1	f_1	a_1	r_2	k_2	e_1	p_1	h_2	d_1
x_5	s_1	f_2	a_0	r_1	k_2	e_1	p_2	h_2	d_2
x_6	s_1	f_1	a_1	r_1	k_2	e_2	p_1	h_2	d_3

x_7	s_1	f_1	a_2	r_1	k_2	e_1	p_1	h_1	d_3

Table 2: Rheumatic Fever Information System

4. Multi Valued Information System

Let U be a finite set of elements called the universe and At be a non-empty finite set of attributes, $a \in At$ such that $f_a : U \to V_a$. The set V_a is called the range of the attribute a. For an element $x \in U$ and an attribute $a \in At$, the pair $(x, f_a(x))$ indicates that x has the attribute value $f_a(x)$. The pair (U, At) is called an information system

A multi-valued information system is a generalization of the idea of a single valued information system. In a multi-valued information system, attribute functions are allowed to map elements to sets of attribute values [1,8]. More formally, in multi valued information system (U, At), each attribute $a \in At$ implies the function $f_a : U \to P(V_a)$ by setting $f_a(x) = V \subseteq V_a$ $(f_a(x) \in P(V_a))$ This way, each element of the information system can be associated a description by means of a subset $A \subseteq At$.

Here is an illustrated example that describes the multi-valued information concept. Table (3) represents a multi valued information system that contains set of persons $P_1, P_2, \cdots, P_6$ and the goods that these persons bought from two markets M_1 and M_2 (attributes).

U	M_1	M_2
P_1	$\{F,D\}$	$\{B,M,D\}$
P_2	$\{H,R\}$	$\{B\}$
P_3	$\{F,D,S\}$	$\{B,M\}$
P_4	$\{F\}$	$\{B,M\}$
P_5	$\{F,D\}$	$\{B\}$
P_6	$\{F\}$	$\{B\}$

Table 3: Multi-Valued Information System

More formally, let $U = \{P_1, P_2, \cdots, P_6\}$ be the set of objects, $At = \{M_1, M_2\}$ be the set of attributes, and the sets $V_{M_1} = \{B, M, D\}$ and $V_{M_2} = \{D, F, H, R, S\}$ are the values of these attributes which are the goods that these persons bought from those markets. These attributes can be treated as mappings $f_a : U \to P(V_a)$ from the set of objects into the family of subsets of their values; where $f_{M_1}(P_2) = \{H, R\}$, $f_{M_1}(P_3) = \{F, D, S\}$, etc.

Multi-valued information system is any collection of data specified in the form of a structure $(U, At, \{V_a : a \in At\}, f_a)$ where U is a non-empty set (of objects), At is a non-empty set of attributes, $\{V_a : a \in At\}$ is an indexed family of sets (values of the attributes) and $f_a : U \to P(V_a)$ be the information function.

5. Generalized Multi Valued Decision Information System (GMDIS)

Here we will give the main conventions that we will apply in this work. These conventions will be indicated by examples.

The structure GMIS = $(U, At, \{\psi_a : a \in At\}, f_a, \{\eta_B : B \subseteq At\})$ is called generalized multi-valued information system, where U be a non-empty finite set of objects.(persons, planets, cars, digits,…etc) called the universe. Any set $X \subseteq U$ is called a category in U . ψ_a is a collection of value sets corresponding to the attribute $a \in At$. $f_a : U \to P(\psi_a)$ is a total information function such that $f_a(x) \in \psi_a$. η_B is a binary relation defined on U , which is not necessary to be an equivalence relation. In [6], many types of relations are defined, the classes of the relation η_B is up to the practical situation of the problem. If there exists an additional attribute D called the decision attribute, we say that this system is a generalized multi valued decision information system, GMDIS = $(U, At \cup D, \{\psi_a : a \in At\}, f_a, \{\eta_B : B \subseteq At\})$. In this paper, we consider η_a as an example of non-equivalence relation on U which is defined by:

- For $a \in C$,

$$\eta_a = \{(x, y) : f_a(x)^c \subseteq f_a(y)\}, \forall a \in At \quad (7)$$

- For $B \subseteq At$,

$$\eta_B = \{(x, y) : f_a(x)^c \subseteq f_a(y), \forall a \in B, B \subseteq At\} \quad (8)$$

Clearly η_a and η_B are not reflexive, not transitive but they are symmetric.

- For $a \in At$ the class A_{η_a} is defined by,

$$A_{\eta_a} = \{\eta_{ax} : x \in U\} \quad , where \quad \eta_{ax} = \{y : x \eta_a y\} \quad (9)$$

- And for $B \subseteq At$ the class A_{η_B} is defined by,

$$A_{\eta_B} = \{\eta_{Bx} : x \in U\} \quad , where \quad \eta_{Bx} = \{y : x \eta_B y\} \quad (10)$$

If D is the decision attribute, then the generalized decision multi-valued information system will take the form $(U, At \cup D, \{\psi_a : a \in At\}, f_a, \{\eta_B : B \subseteq At\})$. In this case, we suggest the following non-equivalence relation,

$$\begin{aligned} \eta_D &= \{(x, y) : f_D(x) \text{ depends on } f_D(y)\} \\ &= \{(x, y) : f_D(x) \subseteq f_D(y)\} \end{aligned} \quad (11)$$

The concepts of this relation are,

$$\eta_{Dx} = \{y : x \eta_D y\} \quad (12)$$

The set of all concepts is,

$$A_{\eta_D} = \{\eta_{Dx} : x \in U\} \quad (13)$$

Also,

- If D is the decision attribute and for $a \in At$, we have

$$POS_a(D) = \bigcup_{X \in A_{\eta_D}} \underline{\eta}_a(X) \quad (14)$$

- For any subset B of At,

$$POS_B(D) = \bigcup_{X \in A_{\eta_D}} \underline{\eta}_B(X) \tag{15}$$

For any subset $X \subseteq U$ we define,

- For $a \in At$,
  $\underline{\eta}_a(X) = \bigcup\{\eta_{ax} : \eta_{ax} \subseteq X\}$
  $\overline{\eta}_a(X) = \bigcup\{\eta_{ax} : \eta_{ax} \cap X \neq \Phi\}$
- For $B \subseteq At$,
  $\underline{\eta}_B(X) = \bigcup\{\eta_{Bx} : \eta_{Bx} \subseteq X\}$
  $\overline{\eta}_B(X) = \bigcup\{\eta_{Bx} : \eta_{Bx} \cap X \neq \Phi\}$

5.1 Reduct Approach and Topological Spaces

Let us take $\{A_{\eta_a} : a \in At\}$ as a sub base of a topological space τ_a (the set of all finite intersections and arbitrary unions of members of A_{η_a}) and $\{A_{\eta_B} : B \subseteq At\}$ as a subbase of a topological space τ_B, the decision makes the topology τ_D which has $\{A_{\eta_D}\}$ as a subbase.

Now, we will define the following,

1- The set of attributes $B \subseteq At$ is called a reduct if $\tau_B \leq \tau_D$ and B is minimal.
Where $\tau_B \leq \tau_D$ *iff* $\forall G \in \tau_B, \exists G' \in \tau_D$ *s.t.* $G \subset G'$, $G, G' \neq U$.

2- The attribute $a \in At$ is called the principal attribute (PA) if $|\tau_a| \succ |\tau_b|$, $\forall a, b \in At, b \neq a$ and if $\tau_a = \tau_b$, then a and b are in the PA.

We will give a new vision for the GMDIS as we can have a SDIS and converting it to a GMDIS as it will be cleared in the worked example.

5.2 Dependencies among Attributes

Discovering dependencies between attributes in GMDIS is an important issue in (Knowledge Data Discovery) KDD [3,5,9]. A set of attribute *B* depends totally on a set of attributes *A* (In general for any two subsets *A* and *B* of *At*), denoted $A \Rightarrow B$ if all values sets of attributes from *B* are uniquely determined by values sets of attributes from *A*. Let *A* and *B* be subsets of *At*. We will say that *B* depends on *A* in a degree K, $0 \leq K \leq 1$,
Denoted by $A \Rightarrow_K B$ if,

$$K = \gamma(A, B) = \frac{|POS_A(B)|}{|U|} \tag{16}$$

$POS_A(B) = \bigcup_{X \in A_{\eta_B}} \underline{\eta}_A(X)$ called a positive region of B.

Obviously,

$$K = \gamma(A,B) = \sum_{X \in A_{\eta_B}} \frac{|\underline{\eta}_A(X)|}{|U|} \tag{17}$$

If $K = 1$, we call B is depends totally on A.
If $K < 1$, we call B is depends partially (in a degree K) on A.

If we take $A = At$ and $B = D$ in the above two issues where At is the set of condition attributes and D is the decision attribute , then we say that, *D* depends totally on At, denoted by $At \Rightarrow D$ if all values of attributes from *D* are uniquely determined by values sets of attributes from *At*. We will say that *D* depends on *At* in a degree K, $0 \le K \le 1$, denoted by $At \Rightarrow_K D$ if ,

$$K = \gamma(At, D) = \frac{|POS_{At}(D)|}{|U|} \tag{18}$$

Where $POS_{At}(D) = \bigcup_{X \in A_{\eta_D}} \underline{\eta}_{At}(X)$ called a positive region of D .

$$K = \gamma(At, D) = \sum_{X \in A_{\eta_D}} \frac{|\underline{\eta}_{At}(X)|}{|U|} \tag{19}$$

If $K = 1$, we call D is depends totally on At .
If $K < 1$, we call D is depends partially (in a degree K) on At.

After getting the degree of dependencies of all condition attributes At with respect to the decision attribute D. Then, the set of attributes of equal highest degree of dependency is the PA of our system.

5.3 Reduct with Minimal Number of Attributes

This section discusses the reduction approach the set of reducts for the GMDIS system using topological spaces and the degree of dependency. If the set of all reducts of our SDIS is, $\Psi = \{R_1, R_2, \cdots, R_n\}$, and the set of reducts for the GMDIS system using topological spaces, and the degree of dependency is, $\Psi' = \{R_1', R_2', \cdots, R_n'\}$. Then, we will say that Ψ is more precise than Ψ', if $\forall R_i' \in \Psi', \exists R_i \in \Psi \ s.t. \ R_i' \subseteq R_i$. The following algorithms describe the reduct approach of GDMIS system and the PA algorithm:

Algorithm 1: GMDIS Reduct Algorithm

<u>Input</u>

A GMDIS = $(U, At \bigcup D, \{\psi_a : a \in At\}, f_a, \{\eta_B : B \subseteq At\})$

<u>Processing</u>

1. $R \leftarrow \{\}$, R=reduct
2. Do
3. $GMDIS \leftarrow R$
4. $Loop\ a \in (At - R)$

5. If $\tau_{R\cup\{a\}} \leq \tau_D$
6. $GMDIS \leftarrow R \bigcup \{a\}$
7. $R \leftarrow GMDIS$
8. Until $\tau_R \leq \tau_D$
9. Return R

Output R: A set of minimum attribute subset; $R \subseteq At$.

Algorithm 2: GMDIS PA Algorithm

Input

A GMDIS = $(U, At \bigcup D, \{\psi_a : a \in At\}, f_a, \{\eta_B : B \subseteq At\})$

Processing

1. $PA \leftarrow \{\}$
2. Do
3. Loop $a \in At$
4. Loop $b \in At$
5. If $|\tau_a| \succ |\tau_b|$
6. $PA \leftarrow PA \bigcup \{a\}$
7. End Loop
8. End Loop
9. Return PA

Output

PA: A set of principal attribute subset, $PA \subseteq At$.

6. Worked Example

In this section, we will give an example to illustrate our approach for attribute reduction compared with the discernibility matrix approach [6].

Let us start by constructing the reduction using the discernibility matrix for the attributes given in Table (2). First, let set the coded table for the attributes give in Table (2). (Table (4) shows the results) by the following assumptions:

- Put Sex (S) ={M,F}={0,1}.
- Put Pharyngitis (F) = {yes, no}={1,0} .
- Put Arthritis(A)={a_0,a_1,a_2}={0,1,2}.
- Put Carditis (R)={affected , not affected}={1,0}.
- Put Chorea (K)= {yes, no}={1,0}.
- Put ESR (E)= {normal ,high}={0,1}.
- Put Abdominal Pain (P)= {absent, present}={0,1}.
- Put Headache (H)={yes, no}={1,0}.
- Put The decision attribute is Diagnosis (D)= {rheumatic arthritis , rheumatic carditis , rheumatic arthritis and carditis}={d_1 , d_2 , d_3}.

History / Patients	S	F	A	R	K	E	P	H	D
x_1	1	1	1	1	1	0	0	0	d_3
x_2	0	1	1	1	1	1	0	1	d_3
x_3	1	1	2	1	0	0	0	0	d_3
x_4	0	1	1	0	0	0	0	0	d_1
x_5	0	0	0	1	0	0	1	0	d_2
x_6	0	1	1	1	0	1	0	0	d_3
x_7	0	1	2	1	0	0	0	1	d_3

Table 4: The Coded Data

The discerniblity matrix of the data given in Table (4) is represented in Table (5).

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	Φ						
x_2	Φ	Φ					
x_3	Φ	Φ	Φ				
x_4	$\{S,R,K\}$	$\{R,K,E,H\}$	$\{S,A,R\}$	Φ			
x_5	$\{S,F,A,K,P\}$	$\{F,A,K,E,P,H\}$	$\{S,F,A,P\}$	$\{F,A,R,P\}$	Φ		
x_6	Φ	Φ	Φ	$\{R,E\}$	$\{F,A,E,P\}$	Φ	
x_7	Φ	Φ	Φ	$\{A,R,H\}$	$\{F,A,P,H\}$	Φ	Φ

Table 5: The Discernibility Matrix

Then the discernibility function will be as following,

$$f_A = \{S \vee R \vee K\} \wedge \{R \vee K \vee E \vee H\} \wedge \{S \vee A \vee R\} \wedge \{S \vee F \vee A \vee K \vee P\}$$
$$\wedge \{F \vee A \vee K \vee E \vee P \vee H\} \wedge \{S \vee F \vee A \vee P\} \wedge \{F \vee A \vee R \vee K \vee P\} \wedge \{R \vee E\}$$
$$\wedge \{F \vee A \vee E \vee P\} \wedge \{A \vee R \vee H\} \wedge \{F \vee A \vee P \vee H\}$$

Applying the algebraic absorption law [9] then f_A takes the following form,

$$f_A = \{S \vee R \vee K\} \wedge \{S \vee A \vee R\} \wedge \{S \vee F \vee A \vee P\} \wedge \{F \vee A \vee R \vee K \vee P\} \wedge \{R \vee E\} \\ \wedge \{F \vee A \vee E \vee P\} \wedge \{A \vee R \vee H\} \wedge \{F \vee A \vee P \vee H\}$$

From the above function we can see that we have 8 reducts with out any intersections so we do not have any core of this example. So, the reduction set is defined as follows:

$$\mathrm{Re}d\,(A) = \{\{S \vee R \vee K\}, \{S \vee A \vee R\}, \{S \vee F \vee A \vee P\}, \{F \vee A \vee R \vee K \vee P\}, \{R \vee E\} \\ , \{F \vee A \vee E \vee P\}, \{A \vee R \vee H\}, \{F \vee A \vee P \vee H\}\} \quad (20)$$

Now, after getting the reduction for the SDIS given in Table (2) using the discernibility matrix. We will convert it to a GMDIS by using the following new system.

Consider the following system:

Attribute Symbol	Refers to ?	Attribute Values	Refers to ?
α	$\{S, K\}$	α_1	$S \to s_1$
		α_2	$K \to k_1$
		α_3	$\{S, K\} \to \{s_2, k_2\}$
β	$\{F, A, E\}$	β_1	$F \to f_1$
		β_2	$A \to a_1$
		β_3	$A \to a_2$
		β_4	$E \to e_1$
		β_5	$\{F, A, E\} \to \{f_2, a_0, e_2\}$
δ	$\{R, P, H\}$	δ_1	$R \to r_1$
		δ_2	$P \to p_1$
		δ_3	$H \to h_1$
		δ_4	$\{R, P, H\} \to \{r_2, p_2, h_2\}$
D	*Diagnosis*	d_1	*Rheumatic arthritis*
		d_2	*Rheumatic carditis*
		d_3	*Rheumatic arthritis and carditis*

Table 6: Converted Data Description

Then we constraint the GMDIS as in the following form satisfied in Table (6).

U	α	β	δ	**D**
x_1	$\{\alpha_2\}$	$\{\beta_1,\beta_2,\beta_4\}$	$\{\delta_1,\delta_2\}$	**{d_3}**
x_2	$\{\alpha_1,\alpha_2\}$	$\{\beta_1,\beta_2\}$	$\{\delta_1,\delta_2,\delta_3\}$	**{d_3}**
x_3	$\{\alpha_3\}$	$\{\beta_1,\beta_3,\beta_4\}$	$\{\delta_1,\delta_2\}$	**{d_3}**
x_4	$\{\alpha_1\}$	$\{\beta_1,\beta_2,\beta_4\}$	$\{\delta_2\}$	**{d_1}**
x_5	$\{\alpha_1\}$	$\{\beta_4\}$	$\{\delta_1\}$	**{d_2}**
x_6	$\{\alpha_1\}$	$\{\beta_1,\beta_2\}$	$\{\delta_1,\delta_2\}$	**{d_3}**
x_7	$\{\alpha_1\}$	$\{\beta_1,\beta_3,\beta_4\}$	$\{\delta_1,\delta_2,\delta_3\}$	**{d_3}**

Table 7: Rheumatic Fever Data in Multi-Valued Information System

The collecting of single valued attributes to multi-valued attribute is the method we suggested to convert the single valued decision information system (SDIS) into generalized multi-valued decision information system (GMDIS). For example, Table (2) can be converted to Table (7) as follows:

1- Firstly, we collect the attributes of Table (2) as it shown in Table (6), where the choice of $\alpha = \{S,K\}$, $\beta = \{F,A,E\}$ and $\delta = \{R,P,H\}$ are according to the case of our study and to the decision of the expert .
2- To get the corresponding values of the multi-valued attributes α , β and δ we define a domain for each of them according to the number of values of the single attributes constructed that attributes. For simplicity, the values we suggested for the attribute α are α_1 , α_2 and α_3.
3- We supposed that α_1 means " S *takes the value* s_1 " , α_2 means "K *takes the value* k_1" and α_3 means " S *takes the value* s_2 *and* K *takes the value* k_2". Not existing any value of the attribute values of α means that it takes the other value automatically.
4- The assumption in Step 3 is not unique and does not depend on the decision of the expert.
5- The reminder attributes β and δ are illustrated in the pre two last rows of Table (6).

Now we will apply the above contributions on Table (7), where

$U=\{x_1,x_2,x_3,x_4,x_5,x_6,x_7\}$ be set of condition attributes $At=\{\alpha,\beta,\delta\}$ and the decision attribute is the diagnosis denoted by D. First, getting the classes of the relation A_{η_a} defined in (3) then,

- For $a=\alpha$,

$A_{\eta_\alpha}=\{\eta_{\alpha x}:x\in U\}$,where $\eta_{\alpha x}=\{y:x\eta_\alpha y\}$ and $\eta_\alpha=\{(x,y):f_\alpha(x)^c\subseteq f_\alpha(y)\ \}$

$$\eta_{\alpha x_1}=\{\Phi\},\ \eta_{\alpha x_2}=\{x_3\},\ \eta_{\alpha x_3}=\{x_2\},\ \eta_{\alpha x_4}=\{\Phi\}=\eta_{\alpha x_5}=\eta_{\alpha x_6}=\eta_{\alpha x_7}$$

$$\Rightarrow A_{\eta_\alpha}=\{\eta_{\alpha x_1},\eta_{\alpha x_2},\eta_{\alpha x_3},\eta_{\alpha x_4}\}$$
$$\Rightarrow A_{\eta_\alpha}=\{\{\Phi\},\{x_2\},\{x_3\}\},$$
$$A_{\eta_\beta}=\{\Phi\}\ \ for\ \ a=\beta\ \ and$$
$$A_{\eta_\delta}=\{\Phi\}\ \ for\ \ a=\delta.$$

- For $B=\{\alpha,\beta\}$

$A_{\eta_B}=\{\eta_{Bx}:x\in U\}\quad ,where\quad \eta_{Bx}=\{y:x\eta_B y\}\ and\ \eta_B=\{(x,y),\ f_B(x)^c\subseteq f_B(y)\}$

$\Rightarrow A_{\alpha\beta}=\{\Phi\},$
$A_{\alpha\delta}=\{\Phi\}\quad for\ B=\{\alpha,\delta\},$
$A_{\beta\delta}=\{\Phi\}\quad for\ B=\{\beta,\delta\}\ and$
$A_{\alpha\beta\delta}=\{\Phi\}\quad for\ B=\{\alpha,\beta,\delta\}$

From the relations we can get the topological spaces for each relation ($\tau_a\ for\ A_{\eta_a}\ and\ \tau_B\ for\ A_{\eta_B}$) as below,

1. $\tau_\alpha=\{U,\Phi,\{x_2\},\{x_3\},\{x_2,x_3\}\}$
2. $\tau_\beta=\{U,\Phi\}$
3. $\tau_\delta=\{U,\Phi\}$
4. $\tau_{\alpha\beta}=\{U,\Phi\}$
5. $\tau_{\alpha\delta}=\{U,\Phi\}$
6. $\tau_{\beta\delta}=\{U,\Phi\}$
7. $\tau_{\alpha\beta\delta}=\{U,\Phi\}$

Now we will deal with the decision attribute D applying the relation that we referred to before where,

$A_{\eta_B}=\{\eta_{Bx}:x\in U\}\quad ,where\quad \eta_{Bx}=\{y:x\eta_B y\}$ and
$\eta_D=\{(x,y):f_D(x)\ depends\quad on\ f_D(y)\}$
$=\{(x,y):f_D(x)\subseteq f_D(y)\}$

$\eta_{D x_1} = \{x_1, x_2, x_3, x_6, x_7\} = \eta_{D x_2} = \eta_{D x_3} = \eta_{D x_6} = \eta_{D x_7}$,

$\eta_{D x_4} = \{x_1, x_2, x_3, x_4, x_6, x_7\}$, $\eta_{D x_5} = \{x_1, x_2, x_3, x_5, x_6, x_7\}$

$\Rightarrow A_{\eta_D} = \{\{x_1, x_2, x_3, x_6, x_7\}, \{x_1, x_2, x_3, x_4, x_6, x_7\}, \{x_1, x_2, x_3, x_5, x_6, x_7\}\}$,

Then we get τ_D as follows:

$$\tau_D = \{U, \Phi, \{x_1, x_2, x_3, x_6, x_7\}, \{x_1, x_2, x_3, x_4, x_6, x_7\}, \{x_1, x_2, x_3, x_5, x_6, x_7\}\}$$

We observe that, $\tau_\alpha \leq \tau_D$, this leads to from the above contributions that $\{\alpha\}$ is the reduct and since, $|\tau_\alpha| \succ |\tau_a|, \forall a \in At$ this leads to that $\{\alpha\}$ is the PA. Then we can get the positive regions for each subset of At with respect to D as follows:

$POS_a(D) = \bigcup_{X \in A_{\eta_D}} \underline{\eta}_a(X)$ for $a \in At$ and

$POS_B(D) = \bigcup_{X \in A_{\eta_D}} \underline{\eta}_B(X)$ for any subset B of At.

$\Rightarrow POS_\alpha(D) = \{x_2, x_3\}$,

$POS_\beta(D) = \{\Phi\} = POS_\delta(D) = POS_{\alpha\beta}(D) = POS_{\alpha\delta}(D) = POS_{\beta\delta}(D) = POS_{\alpha\beta\delta}(D)$

Then we can get the degree of dependency for each attribute as follows,

$$\gamma(a, D) = \frac{|POS_a(D)|}{|U|}$$

For $a = \alpha$, we get $\gamma(\alpha, D) = \frac{2}{7}$, for $a = \beta$, we get $\gamma(\beta, D) = 0$ and for $a = \delta$, $\gamma(\delta, D) = 0$.

But if we get the degree of dependencies for the other attributes we will find that

$$\gamma(\{\alpha, \beta\}, D) = \gamma(\{\alpha, \delta\}, D) = \gamma(\{\beta, \delta\}, D) = \gamma(C, D) = 0$$

Thus, the set of attributes of equal highest degree of dependency is the PA of our system. So we conclude that $\{\alpha\}$ is the PA of our system and this is the same result that we got using the topology.

$$RED(At) = \{\alpha\} = \{S, K\} \tag{21}$$

Now, from (20) and (21) we observe that the reduction that we got by using the GMDIS is contained in the reduction that we got using the discernibility matrix. Since, $\{S, K\}$ is a subset of the most reducts founded by the discernibility matrix that is means our method for getting the reduction is more precise than using the discernibility matrix method in the SDIS.

7. Conclusion

This paper introduced a new approach for obtaining reducts in generalized multi-valued decision information system in a general case. This approach extended Pawlak approach if the system is single valued and the relation is equivalence. The approach opens the way for other approaches of data reduction if we use the general topological recent concepts such as Pre-open sets, Semi-open sets, etc. In many real life situations, the use of attributes in a single fashion is not represetable for the actual effect of attributes. So, it is necessary to consider subsets of the attributes as multi criteria. Our approach uses the general topological structures as a tool for defining reduction and basic RS concepts. The usual RS approach is a special case of our generalization. Since it depends on a certain type of topologies in which every open set is closed. Thus, we believe that the general structure is more near to reality. An application of, rheumatic fever datasets has been chosen and the reduct approach has been applied to see their ability and accuracy. According to the natural of data, which we get from the application, we suggest the suitable conversion. Hence, we choose some condition attributes to be joined to make a multi-condition attribute and this choice not unique, but according to the expert.

References

[1] Abd El-Monsef M.M.E. (2004), "Statistics and Roughian", Ph.D. Thesis, Faculty of Science, Tanta University, Egypt.

[2] Bazan J., Skowron A., and Synak P., “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, 1994, Lecture Notes in Artificial Intelligence, N. 869, Springer-Verlag, Berlin, pp. 346-355, 1994.

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

[4] Grzymala-Busse J., Pawlak Z., Slowinski R., and Ziarko W. (1999). “Rough Sets”, Communications of the ACM, Vol.38, No. 11.

[5] Kryszkiewicz M., and Rybinski H., “Finding Reducts in Composed Information Systems, Rough Sets, Fuzzy Sets Knowledge Discovery” In Proceedings of the International Workshop on Rough Sets, Knowledge, Discovery, W.P. Ziarko (ed.) Banff, Alta., Canada, Springer-Verlag, pp. 261-273,1994.

[6] Orlowska E. (1998). "What You Always Wanted to Know about Rough Sets", Incomplete Information Set Analysis, Orlowska (Ed.) , Physica- Verlag.

[7] Pawlak Z. (1991). "Rough Sets – Theoretical Aspects of Reasoning About Data", Vol. 9 of System Theory, Knowledge Engineering and Problem Solver Kluwer Academ ic Publishers, Dordrecht.

[8] Salama A. S. (2004), “Topology and Information Systems”, Ph.D. Thesis, Faculty of Science, Tanta University, Egypt .

[9] Skowron A. and Zhong N. (2000). "Rough Sets in KDD: Tutorial Notes". Warsaw University and Maebashi Institute of Technology.

[10] Starzyk J. A., Nelson D. E. and Sturtz K. (1999). "Reduct Generation in Information Systems", Bulletin of International Rough Set Society, Vol. 3., No. 1/2.

[11] Zhong N., Skowron A., and Ohsuga S. (eds.) (1999). "New Directions in Rough Sets, Data Mining and Granular-Soft Computing", LNAI 1711, Springer.

Egyptian Rough Sets Working Group

The first one Day Workshop on Rough Sets and their Applications – WRSTA2006

Alexandria University, Faculty of Science, Mathematics Department, 13 August, 2006

A Multi Criteria Decision Analysis Using Fuzzy Logic

F.F. Farahat
Dept. of Computer Science and Information Systems
Sadat Academy, Al-Maadi- Cairo (Egypt)
E-mail: farahat123@yahoo.com

Abstract: The fuzzy sets (FSs) and rough sets (RSs) are very important research areas, especially in Hybrid and Integrated Intelligent Systems, and Multi-Criteria Decision Analysis Problem. In the real economic life, the optimum moment to replace equipment with a new one plays an important role. One can find that in the classical financial mathematics almost all models have as a goal to find the optimum moment to replace the equipment under the condition of the minimum expenses. These classical mathematical models do not keep some quantitative and qualitative parameters together on the one hand, and ignore some uncertainty, on the other hand. In this paper, a trial is presented to eliminate these deficiencies by applying fuzzy models. Three fuzzy models will be proposed, here. The first one is developed to find the best moment of the equipment replacement, while the second fuzzy model is given to select new equipment. The third one is presented to determine multiple Internal Revenue Rate (IRR), based on J.T.C. Mao's algorithm. In all cases, the quantitative and the qualitative criteria will be taken into consideration. These models are based on the following principles: (1) the rigorous methods to transform all criteria into some fuzzy sets on the same universe; (2) the use of the appropriate aggregation operators (AOs) of the type of generalized means; and (3) the decision – making in the multifactor framework. Fuzzy models are accessible ones, easy to simulate and do not get a computation complexity. Some examples will be presented for persons who are working in this area.

Keywords: Rough Sets, Fuzzy Sets, Economical Models, Fuzzy Models, Aggregation Operators, Decision – Marking, Efficient Solutions, *IRR, Financial Analysis.*

1. Introduction

The rapid changes that have taken place globally on the economic, social and business fronts characterized the 21st century. The magnitude of these changes has formed an extremely complex and unpredictable decision-making framework, which is difficult to model through traditional approaches. The most recent advances in the development of innovative techniques for managing the uncertainty that prevails in the global economic and management environments were presented in, [1, 2]. These techniques originate mainly from FSs theory. However, the integration of FSs with other decision support and mode- ling disciplines,(such as multi criteria decision aid, neural networks, genetic algorithms, machine learning, chaos theory, etc) is explored in [1,2]. The presentation of the advances in these fields and their real world applications adds a new perspective to the broad fields of management science and economics. In the real economic life, there are many important problems that are playing important roles in this life such as: (1) The optimum moment for replacing equipment with a new one,[13] , (2) The selection of new equipment,[13], (3) Decision Making, Management and Marketing,[2] , and (4) Multiple Fuzzy IRR in the Financial Decision Environment (S F González et al.[5]).

In this paper we are going to solve these problems using fuzzy logic. In [2], the following problems are studied, namely, (1) Decision Making, Management and Marketing, (2) Algorithms for Orderly Structuring of Financial "Objects" (J Gil-Aluja), (3) A Fuzzy Goal Programming Model for Evaluating a Hospital Service Performance (M Arenas et al.) ,(4) A Group Decision Making Method Using Fuzzy Triangular Numbers (J L GarcíaLapresta et al.), (5) Developing Sorting Models Using Preference Disaggregating Analysis: An Experimental Investigation (M Doumpos & C Zopounidis), (6) Stock Markets and Portfolio Management, (7) The Causality Between Interest Rate, Exchange Rate and Stock Price in Emerging Markets: The Case of the Jakarta Stock Exchange (J Gupta et al.), (8) Fuzzy Cognitive Maps in Stock Market (D Koulouriotis et al.), (9) Neural Network vs. Linear Models of Stock Returns: An Application to the UK and German Stock Market Indices (A Kanas), (10) Corporate Finance and Banking Management, (11) Expertons and Behavior of Companies with Regard to the Adequacy Between Business Decisions and Objectives (A Couturier & B Fioleau), (12) Multiple Fuzzy IRR in the Financial Decision Environment (S F González et al.) , and (13) An Automated Knowledge Generation Approach for Managing Credit Scoring Problems (M Michalopoulos et al.) .

2. Mathematical Model

There are three model categories that reflect some kinds of certainty or uncertainty, namely, (i) Deterministic Mathematical Models (DMM),(ii) Probabilistic /Random Mathematical Models (PMM), and (iii) Fuzzy Mathematical Models (FMM) . Randomness is a deficiency of the causality's law, and fuzziness is a deficiency of the law of the excluded middle. Probability theory applies the random concept to generalized laws of causality – laws of probability. FSs theory applies the fuzzy properties to the generalized law of the excluded middle, the law of membership from fuzziness, [3,4].For solving any multi-criteria decision problem, using the classical mathematical models, do not keep some quantitative and qualitative parameters together on the one hand, and ignore some uncertainty, on the other hand. In this paper, a trial is presented to eliminate these deficiencies by fuzzy models. Three kinds of fuzzy models will be proposed, here. The first model is proposed to find the best moment of the equipment replacement [4], the second one is proposed to select new equipment, [4], and the third one is proposed to determine multiple Internal Revenue Rate (IRR). Based on J.T.C. Mao's Algorithm, [5]. In the first and the second models, the quantitative criteria (such as the price of acquisition, the amount of running expenses) and the qualitative ones (as: the reliability, the productivity, the color and so on) will be taken into consideration. These models are based on the following facts: (1) the rigorous methods to transform all criteria into some FSs on the same universe; (2) the use of the appropriate AOs of the type of generalized means; and (3) the decision – making in the multifactorial framework. These models are characterized by the following merits, namely, (1) they are accessible ones, (2) they are easy to simulate and (3) they do not get a laborious calculus.

On applying the FMMs, there are two main problems, namely, (i) The determination of the MFs, and(ii) the utilization of an appropriate AO. Consequently, the fuzzy statistical methods and the method of comparisons will be used to determine the MFs. Also, the t-norms, t-co norms and weighted generalized means will be used to aggregate the FSs. Many solved examples will be presented for good understanding,

3. A Fuzzy Multiple Attributes Decision Making Problem

3.1. Definitions and Abbreviations

(1) A fuzzy set (FS) in the universe is defined as a pair {U, A}, where $A:U\rightarrow[0,1]$ is MF, and A (u) is the degree of membership of u to the FS: A. For simplicity, it is denoted by the same letter, A, for the FS as well as its MF. The collection of the FSs in, U, will be denoted by F(U) [13].

(2) A fuzzy number 'A' is defined by Carlsson and Fuller [5], as a FS of the Real line with a normal convex (MF) of bounded support. .
(3) The set of fuzzy numbers (FN) is denoted by F.
(4) A FN with a single maximal element is called a quasi-triangular (QT) FN.
(5) A FS ' A 'is called a symmetric triangular (ST FN) with center 'a' and width'α' > 0 if its MF has the following form:

$$A(t)=\begin{cases}1-\dfrac{|a-t|}{\alpha} & \text{if } |a-t|\le\alpha, \ldots\ldots\ldots\ldots\ldots\ldots \\ 0 & \text{otherwise}\end{cases} \qquad (1)$$

(6) Following Carlsson and Fuller, [5], one can use the notation A =(a , α) ,to denote such STFN . If α =0 then A collapses to the characteristic function of { a } $\subset$ *IR*, and we write A = a . A triangular fuzzy number (TFN) with center 'a', may be seen as a fuzzy quantity “x is approximately equal to a ”.
(7) An aggregation operator (AO) is a mapping M: [0,1] n →[0,1].

3.2. Description of the Problems

Let U be a set of alternatives or strategies and G= {A_1, A_2… Am} , is a set of goods or objectives or criteria. Some of these objectives should be linguistic variables. By an appropriate method one can transform every A_i, i $\in$ {1,2,…,m} into FS in, U, that is to find the MF, A_i : U →[0,1].

In this way, one has to define a Victorian function V: U →[0,1] m, where V(u) = (A_1(u),A_2(u),…,A_m(u)). Through an AO, one can synthesize this vector into scalar,

that is the function M: $[0,1]m \rightarrow [0,1]$. If there is $u_0 \in U$ so that : $u_0 = U \sup \in M (V(u))$, then u_0 is the "good" alternative. This reasoning is summarized in Fig .(1) .
In the classical financial mathematics almost all models have the target of finding the optimum moment to replace the equipment under the condition of the minimum expenses. Next, three kinds of models are proposed: (1) the first one is for the replacement of the equipment , (2) the second one is for the choice of new equipment, and (3) the third one is to determine multiple IRRs, based on J.T.C. Mao's Algorithm. These fuzzy models take into account more conditions that can be either quantitative or qualitative in nature (got by linguistic variables). All the criteria are turned into the FSs of the same universe.

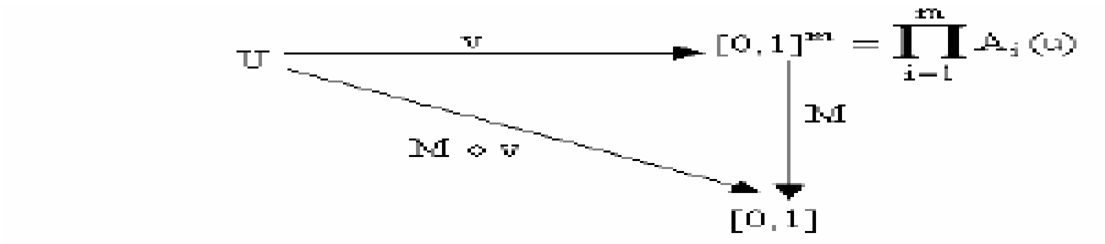


Figure 1

3.3. A Multifactor Fuzzy Model

Assuming the following criteria:

(1) $T = [0,a]$, $a > 0$ be an interval of time,
(2) E, is the equipment
(3) The beginning of the operation is at the time $t = 0$,
(4) a_1 is the residual value (value's recovery),
(5) a_2 is the reliability,
(6) a_3 is the technological wear,
(7) a_4 is the scientific depreciation of E,
(8) a_5 is the running's expenses,
(9) a_6 is the upkeep's expenses , ……,(n+3) a_n: ………………………

Corresponding to these criteria, one can obtain the FSs: A_i in the interval, T, that is: $A_i \in F(T)$, $i = 1,2,\ldots,n$. For example, A_3 (t) is the degree of the technological wear at the moment "t", $t \in T$.

Now, how one can establish the MFs in order to keep the underlying properties of phenomenon? .The incremental method can be an appropriate one for the A_1, A_2, A_3,etc.

The fuzzy statically experiments, considered under various forms (i.e. the method of comparison, preferred, absolute comparison etc.) lead to a good, MF, [3, 4]. All MFs can be approximated by piecewise linear fuzzy quantities, refer to Fig.(2). The MFs, A_i , are supposed to be continuous functions on the time interval, T. From the nature of criteria, two kinds of MFs are to be distinguished, namely, (a) non-increasing, and (b) non-decreasing. Let us consider the AOs : D & C are defined according to the following two equations : (2)& (3) .

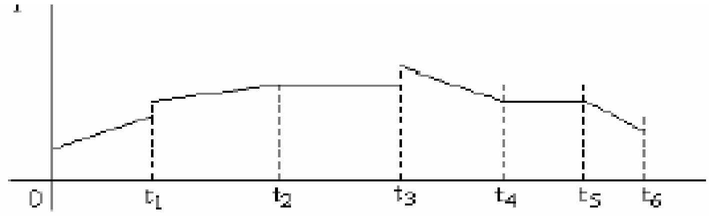


Figure 2

$$D(t) = \sum \alpha_{ik} A_{ik}(t), \ldots\ldots\ldots\ldots \quad (2)$$

$$C(t) = \sum \beta_{jk} A_{jk}(t), \ldots\ldots\ldots\ldots \quad (3)$$

For D(t) : $\alpha_{ik} \varepsilon [0,1]$, $\sum \alpha_{ik} = 1$, $i_k = 1,\ldots..,p$, and A_{ik} is non-increasing , and for C(t) : $\beta_{ik} \varepsilon [0,1]$, $\sum \beta_{jk} = 1$, $j_k = 1,\ldots..,q$, and A_{jk} is non-decreasing .
It must be noted that: $p + q = n$, and $t \varepsilon [0,T]$. From the above two equations (2) and (3), it is clear that the operator, C ,is a non-decreasing continuous function and D, is a non-increasing continuous function. These properties guarantee the existence of the solutions for the following equation:

$$C(t) = D(t), \ldots\ldots\ldots\ldots \quad (4)$$

If there exists a single point :t_0 ε (0,T), as a unique solution of the last equation (4), then ' t_o ' is "the best" moment to replace the equipment , refer to Fig. (3). If the set of solutions of the Equation (4) is an interval $H = [t_1,t_2]$, then every point $t \varepsilon [t_1, t_2]$, can be considered as a "good" moment ,refer to Fig. (4).

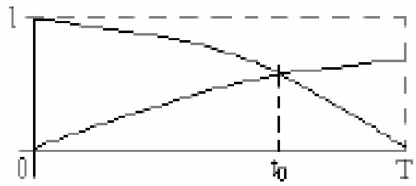


Figure 3

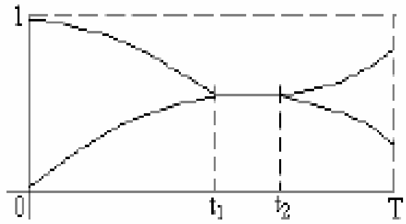


Figure 4

3.3.1. Numerical Example:

Let T = [0, 6] , be the time interval. For simplicity, let us consider the following attributes: (i) the residual value, (ii) the reliability, (iii) the technological wear and (iv) the running's expenses. Let us suppose that the corresponding MFs are defined as follows:

$A_1(t) = 1- (t^2 / 36)$, $A_2(t) = 1- (t^2/25)$, $A_3(t) = t^2/ 25$, $A_4(t) = t^2 / 49$

It is clear that: (i) A_1 & A_2 are non-increasing functions and (ii) A_3 & A_4 are non-decreasing functions. Then, taking the average value (mean), and solving the Eq. (4) in time, t, one can get that :

$$(1-(t^2/25))/2 + (1-(t^2/36))/2 - t^2(1/25 + 1/49)/2 = 0 \,, \ldots\ldots \quad (5)$$

The last Equation,(5) , has in the interval: T = [0,6] ,the unique solution: $t_0 \doteq 4{,}03$, which is the best moment to replace the equipment , refer to Fig. (5) .

It must be noted that for higher order MF (the degree is more than two), one can solve this problem using numerical methods to find the best moment for the equipment replacement by the simulation aid.

3.4. A Fuzzy Model for New Equipment: Selection:

Let us consider the following assumptions and criteria: (i) the existence of a DB about the type of the equipment that will be selected , (ii)Let $U = \{E_1, E_2 \ldots E_m\}$, be the set of the supply equipments , (iii)It is necessary to choose an equipment, E_j , looking for the following criteria : (1) b_1 is the price of acquisition , (2) b_2 is the operation's expenses , (3) b_3 is the maintenance's expenses , (4) b_4 is the reliability in the running, (5) b_5 is the productivity, (6) b_6 is the longevity , ……………………………
(iv)Corresponding to these features, one can obtain the FSs: B_1, B2… $B_6,\ldots,B_n$ respectively a "good price", a "good longevity", defined on the universe $U = \{E_1, E_2,\ldots,E_n\}$. For example, $B_1(E_1)$ is the degree that the equipment,E_1 ,has a "good price". Similarly, one can interpret $B_i(E_k)$, as the degree that the E_k has a "good B_i". But, how can one find the degree of membership of E_k to the FS "good B_i"? . This can be known from a DB, where the values of ,B_i ,can be found for every E_k. For example, the price of acquisition of the equipment E_k is P_k . We order the set $P = \{P_k\}$ into an increasing row: $\{P_{1k,}, P_{2k,} \ldots, P_{mk}\}$, and then define :

$$B_1 (E_{k1}) = P_{k1} / P_{k1} = 1 \text{ , and } B_1 (E_{kj}) = P_{k1} / P_{kj} \,, \ldots\ldots (6)$$

Similarly, one can proceed for B_2 and B_3. In the cases of B_4, B_5, B_6 the corresponding values must decrease in order. For example, we have: $T = \{t_k\}$, the set longevity of the equipments E_k (k = 1,2,…,m). If we denote, tk, as the maximum value of $\{t_k\}$,i.e. $t_{k=} \max\{t_k\}$, then:

$$B_6(E_{k1}) = t_{k1} / t_{k1} = 1 \text{ , and } B_6 (E_{kj}) = t_{kj} / t_{k1} \,, \ldots\ldots \quad (7)$$

Obviously for some $B_{j,}$ one can use other methods in accordance with the nature of the criteria b_j. There is a class of non-decreasing and a class of non-increasing MFs. In this way, one can obtain a mapping V from U to [0, 1], defined as follows:

$$V(E_k) = (B_1(E_k), B_2(E_k), \ldots, B_n(E_k)) \,, \ldots\ldots \quad (8)$$

Now all information keep by V (E_k) can be synthesized through an operator
M: [0,1]n→[0,1], defined by:

$$M(V(E_k)) = \sum a_i B_i (E_k) \,; \ldots\ldots \quad (9)$$

The weights, a_i , reflects the degree of the criterion importance in the decision-making, and $\sum a_I = 1$. This operator, M, led to a new FS in U "the best equipment". The number M (V (E_k)) is the degree of membership of E_k to this set. If M (V (Eq)) is the maximum, that is:

$$M(V(Eq)) = \max_k (V(E_k)) ,\ldots \qquad (10)$$

Then, the equipment Eq is selected.

3.4.1. Numerical Example:

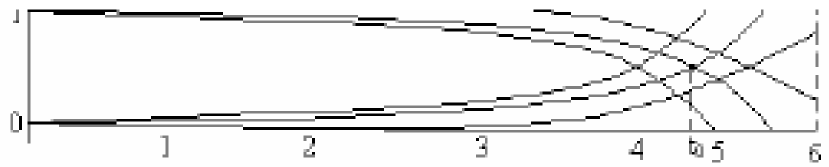


Figure 5

It is required to choose "the best equipment" between '5 'supplies taking into account the following criteria: (i)the price of acquisition, (ii)the expenses of operation, (iii) the reliability and (iv) the productivity. From the supply information and the estimation of the experts the vectors V(E_k) are obtained, which are represented by the rows given in the following Table (1):

	E_1	E_2	E_3	E_4	E_5
B_1	.7	.3	.5	.8	.4
B_2	.4	.7	.6	.3	.8
B3	.3	.7	.7	.4	.7
B_4	.5	.6	.8	.3	.7

Table (1)

Taking the weights: a_1=1/2, a_2= 1/3, a_3= 1/12, and a_4= 1/12 , ($\sum ai_i = 1$) , then the vector M(V(E_k)) from Eq. (8) will have the following form :

M(V(Ek))=(0.550, 0.491,575,-5583,· 5833).,……………………………….(11)

It is clear that: M(V(E5)) = max M(V(Ek)) = · 5833. Therefore, the equipment E_5 is "the best one". It must be noted that for large scale Tables (for example m*n , where 'm' is the number of criteria, and 'n' is the number of equipments), then one can use the digital computer for solving this problem .

3.5. Multiple Fuzzy IRR in the Financial Decision Environment:

In [5], a new fuzzy methodology was presented to determine multiple IRRs, based on

J.T.C. Mao's Algorithm. Also, an alternative algorithm that exhibits a high level of efficiency and efficacy to solve the multiple IRR problem, was also presented ..The analysis and algorithms presented there have not been reported so far in the fuzzy literature. It is known for all that the goal of any investment project evaluation is to determine a measure of the investment,. [Terceño Gómez A. et al [6]). That measure is an indicator that leads to a decision to reject or accept the investment. The financial evaluation of any company requires four factors , to efficiently guide the evaluation criteria to be applied., namely, (i) the determination of the cash flow, (ii) the planning horizon (lifetime), (iii)the interest rate, and (iv) the behavior of the cash flow with time.[5] According to Mendoza [5], all companies search the efficient assignment of financial resources (the necessary assets to be productive), pursuing the goal at long term, from a financial perspective. Many investment projects can be justified, but not all of them can be accomplished. That is the main reason to establish a hierarchy and select to most profitable ones. To reach this goal, you need to evaluate each of the multiple investment possibilities present to the company at a given moment. The traditional criteria are efficient when the information is well-behaved or it can be analyzed with probabilities. Nevertheless, these perceptions have taken place in several occasions, through reasoning based in the concept of precision and have been formalized through the classical mathematical schemes. The result is a set of models that constitute a modified reality that adapts to our mathematical knowledge, instead of the other way around, an adaptation of model to the facts. That is the reason why the main mathematical tool to handle uncertainty is fuzzy theory, with all of its variants. On the other hand, likelihood is treated with probability theory. An analysis of investment evaluation in the presence of multiple financial decisions in a fuzzy environment was presented,[5,6] .The fuzzy IRR with multiple cash flow was studied in,[5], using fuzzy cash flow and interest rates to determine the Internal Revenue Rate (IRR). The analysis uses the fuzzy number criteria.

4. Conclusion

Fuzzy theory has not been well known, or even unknown, as much by the pure mathematicians as by the specialists in applied mathematics. Recently, there has been a change and now it is known better how to separate and associate, when necessary, what is measurable, and what is not, [5]. In this paper, three problems were anal sized and solved using fuzzy models. It is concluded that : (a) These fuzzy models are based on the following principles :(1) the rigorous methods to transform all criteria into some fuzzy sets on the same universe; (2) the use of the appropriate AOs of the type of generalized means; and (3) the decision – making in the multifactor frame-work. (b) These Fuzzy models are accessible ones, easy to simulate and do not get a computation complexity.

(c) Many problems are solved using this fuzzy logic (such as in Management, Economics, Marketing, Engineering, etc. Some numerical examples were presented for persons who are working in this area.

5. References

1. http://www.worldscibooks.com/browse.shtml.

2. Constantin Zopounidis *(Technical University of Crete, Greece)*, Panos M Pardalos *(University of Florida, USA)* & George Baourakis *(Mediterranean Agronomic*

Institute of Chania, Greece), Fuzzy Sets in Management, Economics and Marketing.
3. Hong Xing Li - Fuzzy Sets and Fuzzy Decision-Making CRC Press. Boca Raton, New York, Vincent C. Yen London Tokyo – 1995.

4. Toader T. Buhăescu "Dunărea de Jos" University of Galați – România,"Fuzzy Models for Equipment Replacement ", Proceedings of the Int'l Conference AMSE 2005.

5. González Santillo F., Flores Romero B, Flores R.J., A Multiple Fuzzy IRR in the Financial Decision Environment, *Proceedings of Int'l AMSE Conference MS'05-2005, (Morelia, México)*.

6.. J.Gil Aluja: - Toward a new paradigm of investment selection in uncertainty. Fuzzy Sets and Systems, 84 (1996) pp. 187-198.

7. T.T. Buhăescu: - Application of uninorm aggregation operators in the theory of approximate prices. Proceedings of the Int. Conf. On Modeling and Simulation in Distributed Applications. MS'2001 – Changsha – China, pp. 58-61.
8. T.T. Buhăescu: - Fuzzy models of stock management Proceedings of the Int. Conf. on modeling and Simulation. MS'2002 – Girona – Spain, pp. 439-443.
9. P. Halpern - Canadian managerial Finance – forth edition 1994, J.F. Weston: Dryden Harcourt Brace & Company, Canada. E.F. Bringham
10. Hong Xing Li - Fuzzy Sets and Fuzzy Decision-Making CRC Press.Boca Raton, New York, Vincent C. Yen London Tokyo – 1995.
11. D. Ramirez - On the significance of fuzzy logic. Proceedings of the Int. Conf. on Modeling and Simulation - MS'2002 – Girona – Spain, pp 445-450.
12. Al.P.Tacu - Fuzzy Systems in Economy and Engineering. Publishing House of H.N.Teodorescu: Romanian Academy 1994.
13. Toader T. Buhăescu "Dunărea de Jos"University of Galați –Romania," Fuzzy Models for Equipment Replacement ", Proceedings of the Int'l Conference AMSE 2005.

Egyptian Rough Sets Working Group

The first one Day Workshop on Rough Sets and their Applications – WRSTA2006

Alexandria University, Faculty of Science, Mathematics Department, 13 August, 2006

A Novel Rough Net Approach for Rules Representation and Verification

Hala S. Own

National Research Institute of Astronomy and Geophysics,
Helwan, Cairo, Egypt
Email: hala@sci.kuniv.edu.kw

Abstract: Rule based systems store knowledge as a set of rules and reason over them to solve problems. A new hype rough net approach for rule generation, representation and reasoning of knowledge-based system using rough sets and Petri nets is presented in this paper. A Rough Petri Net (RPN) is proposed for representing knowledge and formalism for the verification of rule-based systems. The main stages of the proposed approach are: Rules will be first generated and normalized, then transform the normalized rules into a Rough Petri net, and finally we verify these normalized rules. A rough Petri net model (RPN) is presented to represent the rule of a rule-based system in which a rough production rule describes the rough relation between two propositions. An algorithm is presented for checking the consistency of a rough knowledge based via a set of reduction rules that preserve the properties of the RPN.

Keyword: Reasoning, Petri net, rough sets, Rule generation, Knowledge-based system

1. Introudction

Knowledge representation and reasoning (KRR) is the study of how knowledge about the real world domain can be represented and what kinds of reasoning can be done with that knowledge [12]. Many of the broad topics within Artificial Intelligence and Computer Science in general (such as reasoning, knowledge engineering, data acquisition, search techniques, human-computer interaction, and others) rely on representations of knowledge [3,9].

Petri Nets (PNs) have ability to represent and analyze in an easy way concurrency and synchronization phenomena, like concurrent evolutions, where various processes that evolve simultaneously are partially independent[1,2,10]. Furthermore, PN approach can be easily combined with other techniques and theories such as object-oriented programming, fuzzy theory, neural networks,[6,11] etc. These modified PNs are widely used in computer;

manufacturing, robotic, knowledge based systems, process control, as well as other kinds of engineering applications. PNs have an inherent quality in representing logic in intuitive and visual way. The reasoning path of expert systems can be reduced to simple sprouting trees if Fuzzy Petri Nets (FPN)-based algorithms are applied as an inference engine[4,8]. Many results prove that FPN is suitable to represent and reason misty logic implication relations. FPN is widely applied in knowledge system representation and redundancy reduction.

In this paper we introduce a rough Petri net model for rule representation. Based on this model we can answer quarries with imprecise information. Moreover we can impose the degree of truth onto transition of Petri net and compute the truth value of associated transition in algebraic form based on the state equation.

The paper is organized as follows: Section 2. introduce brief introduction of rough set. A formal description of the model RPNM is introduced in section 3. In section 4, we introduce Attribute and Rule Reduction in RPNM. A RPNM Verification Methodology is described in section 5. A Production Rough Rule and Reasoning Algorithm will be introduced in section 6.Finally a conclusion and future work will be discussed in section 7.

2. Rough Sets: Basic Concepts

Rough sets theory has been proposed by Pawlak [18,19] for knowledge discovery in databases and experimental data sets. The main objective of rough set mediated data analysis is based on the concept of an upper and a lower approximation of a set, the approximation space and models of sets. The structure of data is represented in the form of information system or, more precisely, the special case of information system called decision table.

An information system can be represented as

$$S = \{U, Q, V, f\} \tag{1}$$

Where U is the universe, a finite set of N objects $\{x_1, x_2, x_3, ..., x_n\}$ (a nonempty set), Q Is a finite set of attributes, $V = \bigcup_{q \in Q} V_q$ (where V_q is a domain of the attribute q), $f : UxQ \to V$ is the total decision function (called the in, $f : Ux(c \cup D) \to V$ formation function) such that $f(x, q) \in V_q$ for every $q \notin Q, x \in U$. A subset of attributes $A \subseteq Q$ an equivalence relation (called an indiscernibility relation) on U such that

$IND(A) = \{(x, y) \in U : \forall a \in A, f(x, a) = f(y, a)\}$,

For more information about rough set one can refer to [18,19].

3. Rough Petri Net Model (RPNM)

This section introduces the rough Petri net approach to model the knowledge representation and verification.

Formal Definition (RPNM): Rough Net structure can be defined as a 10 –tuple:

$$RPNM = (P, T, F, \eta, I, R_{in}, R_{out}, Cer, Cov, Str).$$

Where

- $P = \{p_1, p_2, ..., p_n\}$ is the finite set of places
- $T = \{t_1, t_2, ..., t_n\}$ is the finite set of transitions
- $F \subseteq (PxT) \bigcup (TxP)$ is the set of arcs
- $\eta = \{\eta_1, \eta_2, ..., \eta_n\}$: where $\eta_i \in [0,1]$ represent the degree of dependency of a attribute represented by place P_i with respect to the class decision.
- $I : P \to \{0,1\}$ Is a rough marking function it represents the distribution of token over places. The rough marking function illustrates the degree of completion of the rough event as a result of the processes of the rough reasoning rules. $I_i, i = 1,2,3,..........n$ where $I_i = 1$ if there is a token in P_i , $I_i = 0$ if P_i is not marked . The initial marking is denoted by I^0 .
- R_{in} is a mapping $PxT \to \{0,1\}$ Corresponding to the set of directed arcs from proposition to rules.
- $R_{in}(P_i, T_j) = 1$ if there is a directed arc from P_i to T_j for $i = 1,2,......n$ and $j = 1,2,......n$ and $R_{in}(P_i, T_j) = 0$ if there is no directed arc from P_i to T_j .
- R_{out} is a mapping $PxT \to \{0,1\}$ corresponding to the set of directed arcs from rule to proposition
  $R_{out}(P_i, T_j) = 1$ if there is a directed arc from T_j to P_i for $i = 1,2,......n$ and $j = 1,2,......n$ and $R_{out}(P_i, T_j) = 0$ if there is no directed arc from T_j to P_i .
- $cer : T \to [0,1]$ Is a certainly factor.
- $\text{cov} : T \to [0,1]$ Is a coverage factor.
- $str : T \to [0,1]$ Is a strength factor.

Execution rules of RPNM:

An essential feature of RPNM is that it can be executed; here we will introduce the enabling and firing rule of RPNM. A rule T_i is enabled means that the condition of the rule has been satisfied and the associated rule is activated [17]. This condition is fulfilled if and only if each input proposition is marked by token.
Enabling rule T_i at a marking *I*, resulting a new marking I' such that

- A token will be added to each output place of the enabling transition.

- *for* each $P_i \in R_{out}(T_i)\ \eta(P_i) = \max(\eta_j)$, $j = 1,2,...\ k$ such that $P_j \in R_{in}(T_j)$

- $Cer(T_i) = cer(T_i) * \eta(P_i)$ where $\eta(P_i) = \max(\eta_j)$, $j = 1,2,...\ k$ such that $P_j \in R_{in}(T_j)$
  $Cov(T_i) = Cov(T_i) * \eta(P_i)$ where $\eta(P_i) = \max(\eta_j)$, $j = 1,2,...\ k$ such that $P_j \in R_{in}(T_j)$
  $Str(T_i) = str(T_i) * \eta(P_i)$ where $\eta(P_i) = \max(\eta_j)$, $j = 1,2,...\ k$ such that $P_j \in R_{in}(T_j)$

Now we will illustrate how RPNM model can be used to model the structure properties and describe the dynamic behavior of the following rough rule.

Disease (yes) AND Age (old) => Test (+)

The description of places and transitions of the MFPN model shown in figure 1 are as follows:

P1: represent the antecedent proposition 1 (disease is yes) with its degree of dependency η_1

P2: represent the antecedent proposition 2 (age is old) with its degree of dependency η_2
T1: represents the min composition operation of the antecedent propositions of the rule with coverage, strength, and certainly factors Cov(t1), Str(t1), and Cer(t1) respectively.

P3: represent the degree of truth of test is +.

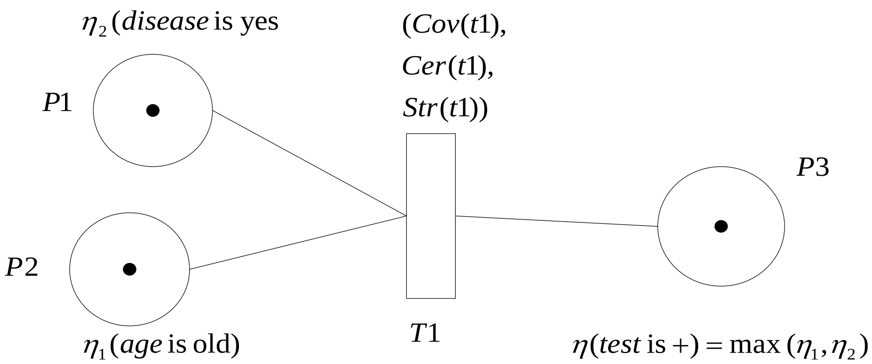


Figure 1. RPNM model of the rough rule Disease (yes) AND Age (old) => Test (+)

The generated model is developed according to the following steps:

Step1: The input attributes for rules is represented as a set of places.

Step2: Construct a set of input object distribution transitions, where these transitions are used to distribute the desired input attributes to the common propositions of the second antecedent part of the rule.

Step3: Calculate the firing strength for each rule according to execution rules of RPNM.

Step 4: Determine the winning rule that has highest confidence among the activated rules.

4 Attribute and Rule Reduction in RPNM

4.1 Attribute Reduction Rules

The place P_i can be suppressed (deleted with its arcs) according to the following rules [5]:

- Rule 1- the output transition of place P_i have no other input places than P_i and P_i is pure i.e. there are no transition which are both input and output transition of P_i.

- Rule 2- If the place P_i is implicit i.e. the marking in this place never forms an obstacle to the firing of its output transition.
- Rule 3- The marking of P_i can be deduced from the marking of other places by relation $M(p_i) = \sum_{k \neq i} (\alpha_k M(p_k)) + \beta$ where α_k is a rational number positive or null, and β is a rational number.

4.2 Rule Reduction algorithm

A transition T_i and its input and output arcs can be suppressed according to the following rules[15]:

Rule 1- the set of input place of a transition T_i is identical to the set of its Output places (i.e.

$$if\ \exists \mathrm{T_k} \neq T_i \text{ such that Post } (\mathrm{P_i}, T_k) \geq \text{pre}(\mathrm{P_i}, T_i) \text{ for every place } \mathrm{P_i} \in \mathrm{R_{in}}(T_i).$$

Rule 2-

$$If\ \exists\, \mathrm{T_k} \text{ such that } \mathrm{R_{in}}(T_k) = R_{in}(T_k) \text{ and } \mathrm{R_{out}}(T_{ki}) = R_{out}(T_{kj}).$$

Rule 3- If Ti is a self loop transition ($i.e. \mathrm{R_{in}}(T_i) = P_i = R_{out}(T_i)$). Suppressing Ti includes first to delete its input and output arc to and from Pi, and delete the transition Ti if it is isolated.

5. RPTNM Verification Methodology

In this section we will use RPTNM as knowledge representation formalism where structural and behavioral properties of the net can be used to verify the knowledge base integrity. In rough set environment, the issues of concern when talking about integrity checking are the definition of concepts such as inconsistency, redundancy and completeness [14].

Petri net provide a tool to investigate all the properties of the net which is reachability tree [13,16]. Reachability tree is made up of nodes which correspond to the reachable markings and each directed edge represents transition firing and connect one node to another resulting in the passing from one marking to another.

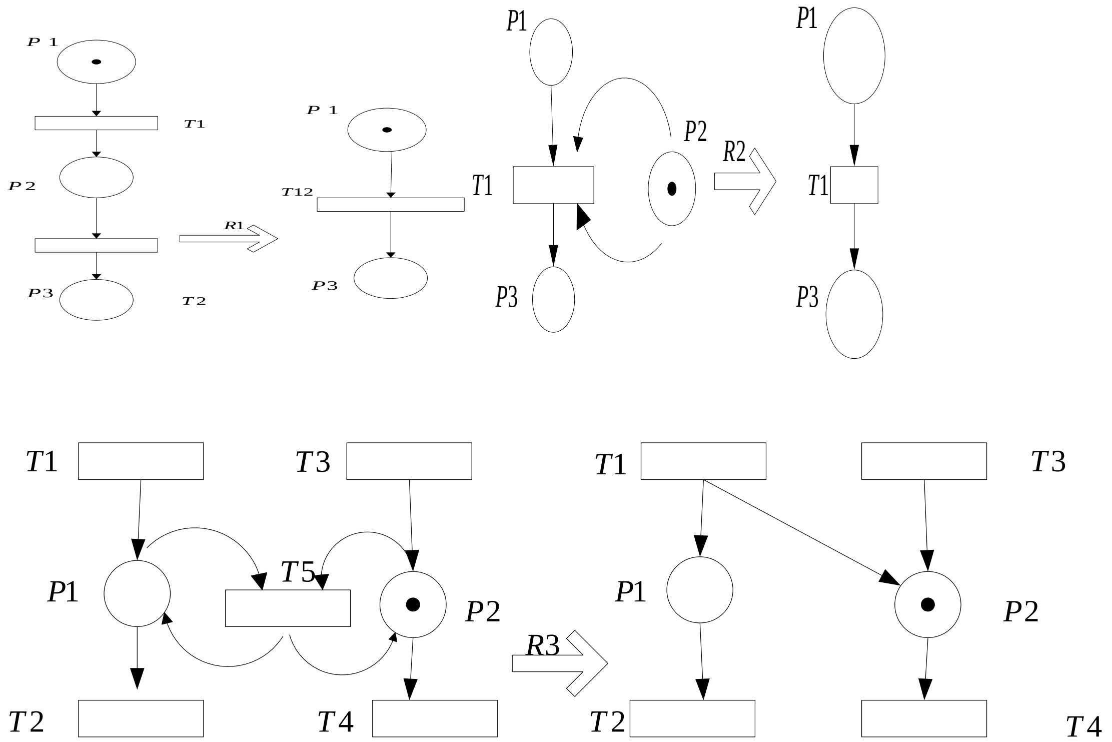


Figure 2 show examples of the 3 rules reduction

The reachabilty tree of RPTNM algorithm is given as follows:

Input: a vector V of nodes denoted the initial marking.
Processing:
Step 1: label the initial marking M_0 as the root

Step2: for all enabled transition $t_i, i = 1,2,3,........n$, n = number of enabled transitions under the current marking M
Step 3: loop

2.1 Generate a new marking and update V according to the new marking, and add an edge labeled by $T = \{t_1, t_2, ..., t_n\}$ from the original node to that node.

2.2 If there is a marking M_j=M_i on the path from M to M_i, then M_i has no successor.

2.3 If there is a marking M_j on the path from M to M_k such that M_k>M_j, then w is placed for each of the components greater than the component M_j. This component ω remain with no change for any firing transition.

2.4 If the new marking $M' = M$, then the set of transitions do not change the current marking. We add an edge labeled T from the current node to itself.

2.5 Go back to step 2.

Output: reachability tree

6. Production Rough Rule and Reasoning Algorithm

The rough rule production and reasoning algorithm is given in the following section.

Based on the model and execution rules of RPNM, we can study the reasoning algorithm to obtain the final reasoning results from the initial truth degree of propositions.

Input: Decision table $\Gamma =< U, \Omega, V_q, f_q >_{q\varepsilon\Omega}$
Output: Reasoning model corresponding to the decision system

Processing:

Step 1: build RPNM for the generated rule.

Step 2: while there is an enabling transition loop

Step 2.1: apply firing execution rule these step generate a new marking M_i

Step 2.2: check if $M_i = M_{i-1}$, if true go to step 3, otherwise go to step 2

Step 3: Apply attribute and rule reduction. Drive a reduct model
Step 4: Verify the model and drive a verified model by:

- o Remove the inconsistent rules
- o Remove the conflict rules
- o Remove redundant rules

Output: Reasoning model corresponding to the decision system

7. Conclusion and Future Work

A rough net model is presented to represent the rough production rule of a rule-based system in which a rough production rule describes the relation between two propositions. Based on the rough net model, an algorithm is proposed to perform rule verification automatically. It can determine whether an antecedent-consequence relationship exists between two propositions. The formal description of the model and the rule verification algorithm are shown in detail with examples. In the future work an investigation of using the RPNM with the proposed reasoning algorithm to predict a class decision in some medical; applications

References

1. T. Agerwala, "Putting Petri nets to work," Computer, vol. 12, no. 12, pp. 85-94, Dec. 1979.

2. W. Brauer, W. Reisig, and G. Rozenberg, Eds., Petri Nets: Applications and Relationships to Other Models of Concurrency. Berlin: Springer-Verlag, 1987.

3. A . J. Bugarin and S. Barro, Fuzzy reasoning supported by Petri nets”, IEEE Trans. Fuzzy Syst., vol. 2, pp. 135-149, Apr. 1994.

4. B . Chandrasekaran, "On evaluating AI systems for medical diagnosis," AI Mag., vol. 4, no. 2, pp. 34-37, Summer 1983.

5. R . David and H.Alla: Petri nets and Grafcet, Prentice Hall, 1992.

6. F. Feldbrugge , K Jensen, Petri net tool overview 1986, Advances in Petri nets 1986, part II on Petri nets: applications and relationships to other models of concurrency, p.20-61, February 1987.

7. B. Fryc, K. Pancerz, and Z.Suraj, “Approximate Petri Nets for Rule-Based Decision Making”, In: Proceedings of Rough Sets and Current Trends in Computing: 4th International Conference, RSCTC 2004, Uppsala, Sweden, June 1-5, 2004, pages 733-742. Volume 3066 of Lecture Notes in Computer Science / Shusaku Tsumoto, Roman Slowinski, Jan Komorowski, et al. (Eds.) --- Springer-Verlag, June 2004.

8. A. Giordana , L. Saitta, Modeling production rules by means of predicate transition networks, Information Sciences: an International Journal, v.35 n.1, p.1-41, March 1985

9. F. Hayes-Roth, Rule-based systems, Communications of the ACM, v.28 n.9, p.921-932, Sept. 1985 .

10. X. Li, W. Yu, and F. L. Rosano, “Dynamic Knowledge interence and learning under adaptive Fuzzy Petri net framework”, IEEE Syst., Man, Cybern, vol. 30, pp.442-450. Nov. 2000.

11. C.G. Looney, Fuzzy Petri nets for rule-based decision making, IEEE Transactions on Systems, Man and Cybernetics, v.18 n.1, p.178-183, January/February 1988.

12. C. G.Looney , A. R. Alfize, Rule-based reasoning as Boolean transformations, IEEE Transactions on Systems, Man and Cybernetics, v.17 n.6, p.1077-1082, Nov./Dec. 1987

13. T. Murata and K. Matsuyama, "Inconsistency check of a set of clauses using Petri net reductions," Franklin Institute, vol. 325, no. 1, pp. 73-93, 1988.

14. T. A. Nguyen, "Verifying consistency of production systems," in Proc. Third IEEE Conf. Artificial Intell. Appl., Orlando, FL, Feb. 1987, pp. 4-8.

15. T. A. Nguyen, W. A. Perkins, T. J. Laffey, D. Pecora, "Checking expert system knowledge bases for consistency and completeness," in Proc. Ninth Int. Joint Conf. Artificial Intell. Los Angeles, CA, Aug. 1985, pp. 375-378.

16. R. M. O'Keefe, O. Balci, and E. P. Smith, "Validating expert system performance," IEEE Expert vol. 2, pp. 81-90, Winter 1987.

17. K. Pancerz, and Z. Suraj, “Synthesis of Petri Net Models: A Rough Set Approach”, Fundamental informaticae XX , IOS Press, 2003.

18. Pawlak Z., “Rough Sets- Theoretical aspect of Reasoning about Data” Kluwer Academic Publishers, 1991.

19. Pawlak Z., Grzymala-Busse J., Slowinski R., Ziarko, W., “Rough sets Communications of the ACM, Vol. 38, No. 11, pp. 89-95, 1995.

Egyptian Rough Sets Working Group
The first one Day Workshop on Rough Sets and their Applications – WRSTA2006

Alexandria University, Faculty of Science, Mathematics Department, 13 August, 2006

Decision Analysis via Granulation Based on General Binary Relation

N. M. Kilany[1] and M.M.E. Abd El-Monsef[2]

Commercial Technical for Computer Sciences Institute in Suez, Egypt[1]
Mathematics Department, Faculty of Science, Tanta University, Egypt[2]

Abstract: Decision theory considers how best to make decisions in the light of uncertainty about data. There are several methodologies that may be used to determine the best decision. In rough set theory, the classification of objects according to approximation operators can be fitted into the Bayesian decision–theoretic, with respect to three regions (Positive, Negative and Boundary region). Granulation using equivalence classes is a restriction that limits the decision makers. In this paper, we introduced a generalization and modification of decision–theoretic rough set model by using granular computing on general binary relations. We obtain two new types of approximation that enable us to classify the objects into five regions instead three regions. The classification of decision region into five areas will enlarge the range of choice for decision maker

.

Keywords: Decision Analysis, Granulation, Binary Relations.

1. Introduction

Making decisions is a fundamental task in data analysis. Some methods have appeared to make a decision. Yao et. al. (1992) proposed and studied a more general type of rough set approximations via Bayesian decision theory. In Section 2 we give a brief overview of granulation structures on the universe. One is defined by an equivalence relation due to Pawlak (1982) and the other by a general relation proposed by Rady et. al. (2004). Approximation structures are discussed for each type of granulation. Section 3 discusses a decision-theoretic model of rough sets under equivalence relations given by Yao et. al. (1992). Our main contribution is to introduce a general decision theoretic model of rough sets using a general relation. The resulted granulation induces approximation different from that due to Pawlak. This enables us to construct two new approximations, namely, semi lower and semi upper approximation which is useful in the partition of boundary region specially and the universe in general with respect to any subset of the universe.

2. Granulation of universe and rough set approximations

In rough set theory, indiscernibilitiy is modeled by an equivalence relation. A granulated view of the universe can be obtained from equivalence classes. By generalizing equivalence relations to binary relations, one may obtained a different granulation of the universe. For any kind of relations, a pair of rough set approximation operators, known as lower and upper approximation operators, can be defined in many ways [Pawlak (1982), Rady et. al. (2004)] .

2.1. Granulation by equivalence relation:

Let $E \subseteq U \times U$ be an equivalence relation on finite non-empty universe U. The equivalence class,

$$[x]_E = \{y \in U : yEx\}$$

Consists of all elements equivalent to x, and is also the equivalence class containing x .

In an approximation space apr = (U, E), ***Pawlak*** (1982) defined a pair of lower and upper approximations of a subset $A \subseteq U$, written as $\underline{apr}(A)$ and $\overline{apr}(A)$ or simply $\underline{A}$ & $\overline{A}$ as follows :

$$\underline{A} = \{x \in U : [x]_E \subset A\}$$

$$\overline{A} = \{x \in U : [x]_E \cap A \neq \Phi\}$$

The lower and upper approximations have the following properties:

For every A and $B \subset U$ and every approximation space $apr = (U, E)$

1. $\underline{apr}(A) \subseteq A \subseteq \overline{apr}(A).$
2. $\underline{apr}(U) = \overline{apr}(U) = U.$
3. $\underline{apr}(\phi) = \overline{apr}(\phi) = \phi.$
4. $\overline{apr}(A \cup B) = \overline{apr}(A) \cup \overline{apr}(B).$
5. $\underline{apr}(A \cup B) \supseteq \underline{apr}(A) \cup \underline{apr}(B).$
6. $\overline{apr}(X \cap B) \subseteq \overline{apr}(X) \cap \overline{apr}(B).$
7. $\underline{apr}(A \cap B) = \underline{apr}(A) \cap \underline{apr}(B).$
8. $\overline{apr}(-A) = -\underline{apr}(A).$
9. $\underline{apr}(-A) = -\overline{apr}(A).$
10. $\overline{apr}(\overline{apr}(A)) = \underline{apr}(\overline{apr}(A)) = \overline{apr}(A).$
11. $\underline{apr}(\underline{apr}(A)) = \overline{apr}(\underline{apr}(A)) = \underline{apr}(A).$
12. *If* $A \subseteq B$, *then* $\overline{apr}(A) \subseteq \overline{apr}(B)$ *and* $\underline{apr}(A) \subseteq \underline{apr}(B)$.

Moreover, for a subset $A \subseteq U$, a rough membership function is defined by Pawlak et. al.(1994):

$$\mu_A(x) = \frac{|[x]_E \cap A|}{|[x]_E|}$$

Where $|.|$ denotes the cardinality of a set. The rough membership value $\mu_A(x)$ may be interpreted as the conditional probability that an arbitrary element belongs to A given that the element belongs to $[x]_E$.

2.2. Granulation by general relation

Let U be a finite universe set and E is any binary relation defined on U, and S be the set of all elements which are in a relation with certain x in U, for all $x \in U$. In symbols,

$$S = \{\{xE\}, \forall x \in U\} \quad , \quad \text{Where} \quad \{xE\} = \{y : xEy; x, y \in U\}$$

Define β as the general knowledge base (GKB) using the arbitrary intersections of the members of *S*. The member that will be equal to any union of some members of β must be omitted. That is, $\beta = \{\beta_i = S_i \cap S_j \; ; \; S_i, S_j \subset S \textit{ and } \beta_i \neq \bigcup S_i \textit{ for some } i\}$. The pair $apr_\beta = (U, E)$ will be called the general approximation space based on the general knowledge base β .

Rady et. al. (2004) extend the classical definition of the lower and upper approximations of any subset A of U to take these general forms

$$\underline{apr}_\beta(A) = \bigcup\{\beta_x : \beta_x \subset A\} \quad \text{and} \quad \overline{apr}_\beta(A) = \bigcup\{\beta_x : \beta_x \cap A \neq \Phi\}$$

Where β_x denotes the subset of β containing x .

These general approximations satisfy all the properties introduced in **(2.1)** except for properties (8, 9, 10 and 11). This is the main deviation that will help to construct our new approach.

For granulation by any binary relation, ***Lashein*** et. al. (2005) defined a rough membership function as follows:

$$\mu_A(x) = \frac{A \cap (\cap \beta_x)}{\cap \beta_x} \; .$$

2.2.1. Granulation by general relation in multivalued information system

For a generalized approximation space ***Ezzat*** (2004) defined a multivalued information system. This system is an ordinary information system whose elements are sets. Each object has number of attributes with attribute subsets related to it. The attributes are the same for all objects but the attribute set-valued may differ.

A multivalued information system (*IS*) is an ordered pair (U, Ψ), where U is a non-empty finite set of objects (the universe), and Ψ is a non-empty finite set of elements called attributes. Every attribute $q \in \Psi$ has a multivalued function Γ_q, which maps into the power set of V_q, where V_q is the set of allowed values for the attributes. i.e., $\Gamma_q : U \times \Psi \to \mathcal{P}(V_q)$.

The multivalued information system may also be written as

$$\boldsymbol{MIS} = \left\langle U, \Psi, V_q, \Gamma_q \right\rangle_{q \in \Omega}$$

With a set $P \subseteq \Psi$ we may associate an indiscernibility relation on U, denoted by $\beta(P)$ and defined by

$$(x, y) \in \beta(P) \ \boldsymbol{iff} \ \Gamma_q(x) \subseteq \Gamma_q(y), \forall Q \in P.$$

Clearly, this indiscernibility relation doesn't perform a partition on U.

Worked Example (1):

In Table (1), we have ten persons (objects) with attributes reflecting each situation of life. Consider that we have three condition attributes namely, Spoken Languages, Computer Programs and Skills. Each one was asked about his adaptation by choosing between {English, German, French} in the first attribute; {Word, Excel, Access, Power Point} in the second attribute; {Typing, Translation} in the third attribute. Let a_i be the i^{th} value in the first attribute, b_j be the j^{th} value in the second attribute and c_k be the k^{th} value in the third attribute.

	Spoken Languages (T_1)	***Computer programs*** (T_2)	***Skills*** (T_3)
x_1	$\{a_2, a_3\}$	$\{b_1, b_3, b_4\}$	$\{c_1, c_2\}$
x_2	$\{a_2\}$	$\{b_1, b_4\}$	$\{c_2\}$
x_3	$\{a_1, a_2\}$	$\{b_1, b_2, b_4\}$	$\{c_2\}$
x_4	$\{a_1\}$	$\{b_1, b_3\}$	$\{c_1\}$
x_5	$\{a_1, a_2, a_3\}$	$\{b_2\}$	$\{c_1\}$
x_6	$\{a_2\}$	$\{b_1, b_3, b_4\}$	$\{c_1, c_2\}$
x_7	$\{a_1, a_3\}$	$\{b_1, b_3\}$	$\{c_1\}$
x_8	$\{a_1, a_3\}$	$\{b_1, b_3\}$	$\{c_1, c_2\}$
x_9	$\{a_1\}$	$\{b_2, b_3\}$	$\{c_2\}$
x_{10}	$\{a_1, a_2\}$	$\{b_1, b_2\}$	$\{c_2\}$

Table 1: Multivalued information system

The indiscernibility relation for C= $\{T_1, T_2, T_3\}$ will be

$$\beta(C) = \{(x_1, x_1), (x_2, x_1), (x_2, x_2), (x_2, x_3), (x_2, x_6), (x_2, x_2), (x_3, x_3), (x_4, x_4), (x_4, x_7), (x_4, x_8), (x_4, x_9), (x_5, x_5), (x_6, x_6), (x_7, x_7), (x_7, x_8), (x_8, x_8), (x_9, x_9), (x_{10}, x_3), (x_{10}, x_{10})\}$$

It's easy to see that $\beta(C)$ doesn't perform a partition on U in general. This can be seen via

$$U/\beta(C) = \{\{x_1\}, \{x_1, x_2, x_3, x_6\}, \{x_3\}, \{x_4, x_7, x_8, x_9\}, \{x_5\}, \{x_6\}, \{x_7, x_8\}, \{x_9\}, \{x_{10}, x_3\}\}.$$

Obviously, the $U/\beta(C)$ is the set S defined in the general approach in Section 2.2.

3. Bayesian decision-theoretic framework for rough sets

In this section, the basic notion of the Bayesian decision procedure is briefly reviewed (Duda & Hart, 1973). We present a review of results that are relevant to decision-theoretic modeling of rough sets induced by an equivalence relation. A generalization and modification of decision-theoretic modeling induced by general relation is applied on the universe.

3.1. Bayesian decision procedure

Let $\Omega = \{\omega_1, ..., \omega_s\}$ be a finite set of s states of nature, and let $\mathcal{A} = \{a_1, ..., a_m\}$ be a finite set of m possible actions. Let $P(\omega_j / X)$ be the conditional probability of an object x being in state ω_j given the object is described by X .

Let $\lambda(a_i / \omega_j)$ denote the loss for taking action a_i when the state is ω_j. For an object with description X, suppose that an action a_i is taken. Since $P(\omega_j / X)$ is the probability that the true state is ω_j given X, the expected loss associated with taking action a_i is given by:

$$R(a_i / X) = \sum_{j=1}^{s} \lambda\,(a_i / \omega_j)\, P(\omega_j / X)$$

and also called the conditional risk. Given description X, a decision rule is a function $\tau(X)$ that specifies which action to take. That is, for every X, $\tau(X)$ assumes one of the actions, $a_1, ..., a_m$. The overall risk **R** is the expected loss associated with a given decision rule. Since $R(\tau(X)/X)$ is the conditional risk associated with the action $\tau(X)$, the overall risk is defined by:

$$\mathbf{R} = \sum_{X} R(\tau\,(X)/X)\, P(X)$$

where the summation is over the set of all possible description of objects ,i.e., entire knowledge representation space. If $\tau(X)$ is chosen so that $R(\tau(X)/X)$ is as small as possible for every X, the overall risk R is minimized. Thus, the Bayesian decision procedure can be formally stated as follows. For every X, compute the conditional risk

$R(a_i / X)$ $for\ i = 1,...,m$ and then selected the action for which the conditional risk is minimum.

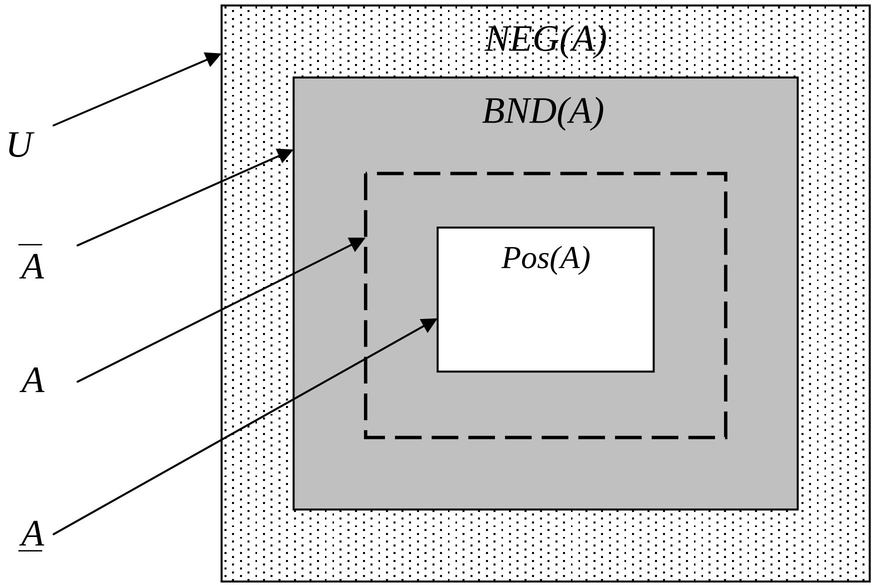
3.2. Decision- theoretic approach of rough sets (under equivalence relations)

Let apr = (U, E) be an approximation space where E is equivalence relation on U. With respect to a subset A $\subseteq$ U one can divide the universe U into three disjoint regions, the positive region *POS(A),* the negative region *NEG(A)* and the boundary region *BND(A).*

$$POS(A) = \underline{apr}(A)$$

$$NEG(A) = U - \overline{A}$$

$$BND(A) = \overline{apr}(A) - \underline{apr}(A)$$



In an approximation space apr = (U, E), the equivalence class containing x, $[x]_E$, is considered to be description of x. The classification of objects according to approximation operators can be easily fitted into Bayesian decision-theoretic framework (***Yao***, et. al. (1992)). The set of states is given by $\Omega = \{A, -A\}$ indicating that an element is in A and not in A, respectively. With respect to the three regions, the set of actions is given by $\mathcal{A} = \{a_1, a_2, a_3\}$, where $a_1, a_2, and\ a_3$ represent the three actions in classifying an object, deciding *POS (A)* deciding *NEG (A)* and deciding *BND (A)* respectively.

Let $\lambda(a_i / A)$ denote the loss incurred for taking action a_i when an object in fact belongs to A, and $\lambda(a_i / -A)$ denote the loss incurred for taking the same action when the object does not belong to A, the rough membership values $\mu_A(x) = P(A/[x]_E)$ and $\mu_{A^c}(x) = 1 - P(A/[x]_E)$ are in fact the probabilities that an object in equivalence class

$[x]_E$ belongs to A and –A respectively. The expected loss $R(a_i/[x]_E)$ associated with taking the individual actions can be expressed as:

$$R(a_1/[x]_E) = \lambda_{11}P(A/[x]_E) + \lambda_{12}P(-A/[x]_E)$$
$$R(a_2/[x]_E) = \lambda_{21}P(A/[x]_E) + \lambda_{22}P(-A/[x]_E)$$
$$R(a_1/[x]_E) = \lambda_{31}P(A/[x]_E) + \lambda_{32}P(-A/[x]_E)$$

where $\lambda_{i1} = \lambda(a_i / A)$, $\lambda_{i2} = \lambda(a_i / -A)$ and $i = 1,2,3$.

The Bayesian decision procedure leads to the following minimum-risk decision rules:

(P) *If* $R(a_1/[x]_E) \le R(a_2/[x]_E)$ *and* $R(a_1/[x]_E) < R(a_3/[x]_E)$ ***decide*** $POS(A)$;
(N) *If* $R(a_2/[x]_E) \le R(a_1/[x]_E)$ *and* $R(a_2/[x]_E) < R(a_3/[x]_E)$ ***decide*** $NEG(A)$;
(B) *If* $R(a_3/[x]_E) \le R(a_1/[x]_E)$ *and* $R(a_3/[x]_E) \le R(a_2/[x]_E)$ ***decide*** $BND(A)$.

Based on $P(A/[x]_E) + P(-A/[x]_E) = 1$, the decision rules can be simplified by using only probabilities $P(A/[x]_E)$.

Consider a special kind of loss with $\lambda_{11} \le \lambda_{31} < \lambda_{21}$ and $\lambda_{22} \le \lambda_{32} < \lambda_{12}$. That is, the loss for classifying an object *x* belonging to *A* into the positive region is less than or equal to the loss of classifying *x* into the boundary region, and both of these losses are strictly less than the loss of classifying *x* into the negative region. For this type of loss functions, the minimum -risk decision rules (P)-(B) can be written as:

(P') *If* $P(A/[x]_E) \ge \gamma$ *and* $P(A/[x]_E) \ge \alpha$ *,* ***decide*** $POS(A)$;
(N') *If* $P(A/[x]_E) < \beta$ *and* $P(A/[x]_E) \le \gamma$ *,* ***decide*** $NEG(A)$;
(B') *If* $\beta < P(A/[x]_E) \le \alpha$ *,* ***decide*** $BND(A)$.
where,

$$\alpha = \frac{\lambda_{12} - \lambda_{32}}{(\lambda_{31} - \lambda_{32}) - (\lambda_{11} - \lambda_{12})}$$

$$\gamma = \frac{\lambda_{12} - \lambda_{22}}{(\lambda_{21} - \lambda_{22}) - (\lambda_{11} - \lambda_{12})}$$

$$\beta = \frac{\lambda_{32} - \lambda_{22}}{(\lambda_{21} - \lambda_{22}) - (\lambda_{31} - \lambda_{32})}$$

From the assumptions, $\lambda_{11} \le \lambda_{31} < \lambda_{21}$ and $\lambda_{22} \le \lambda_{32} < \lambda_{12}$ it follows that $\alpha \in [0,1]$, $\gamma \in (0,1)$ and $\beta \in [0,1)$. Note that the parameters λ_{ij} should satisfy the condition $\alpha \ge \beta$. This ensures that the results are consisted with rough set approximations. That is, the boundary region may be non- empty.

3.3. Generalized Decision theoretic approach (under general relation)

Define an approximation space apr = (U, E), where E be any binary relation defined on U. Since the general approximations cannot satisfy the properties (10, 11) in (**2.1**), we can define tow new approximations, namely, semi lower and semi upper approximation as follows:

$$semiL(A) = A \bigcap \underline{(\overline{A}_\beta)}_\beta$$

$$semiU(A) = A \bigcup \overline{(\underline{A}_\beta)}_\beta$$

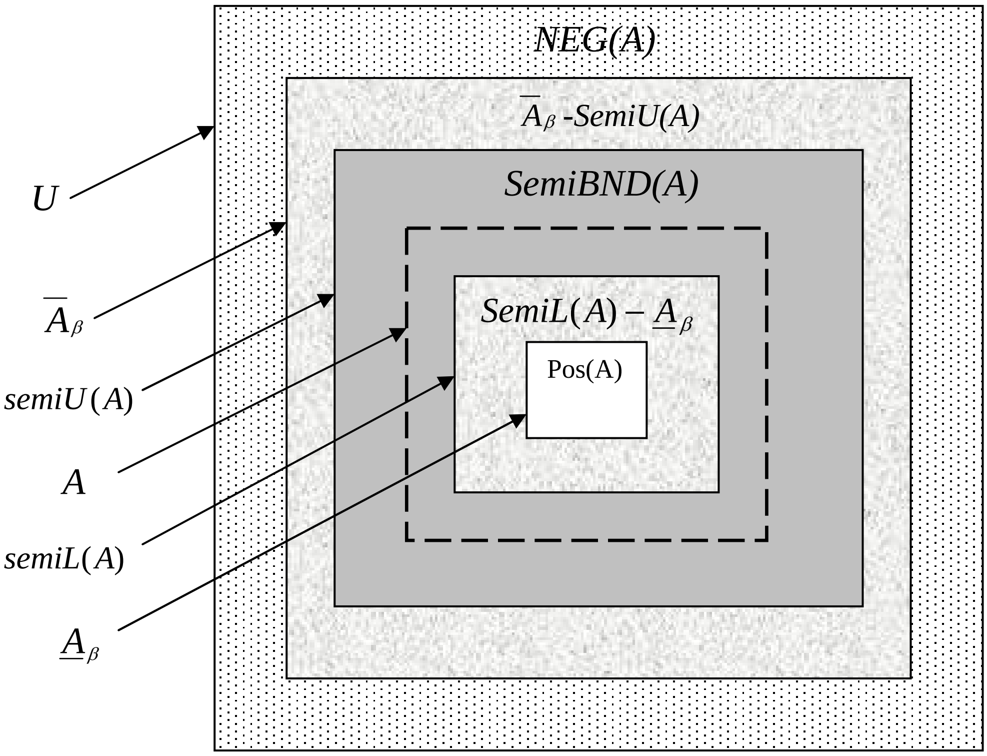
where $\underline{A}_\beta \subseteq semiL(A) \subseteq A \subseteq semiU(A) \subseteq \overline{A}_\beta \subseteq U$.

This enables us to divide the universe U into five disjoint regions as follows:

POS(A) , *SemiL(A)*- $\underline{A}_\beta$, *SemiBND(A)* , $\overline{A}_\beta$-*SemiU(A)* , *NEG(A)*

where,

SemiBND(A) = SemiU(A) - SemiL(A)



In this case, the set of states remains $\Omega = \{A, -A\}$ but the set of actions becomes $\mathcal{A} = \{a_1, a_2, a_3, a_4, a_5\}$ where a_1, a_2, a_3, a_4 and a_5 represent the five actions in classifying an object deciding *POS*(A), deciding *SemiL(A)*- $\underline{A}_\beta$, deciding *SemiBND(A)*, deciding $\overline{A}_\beta$-*SemiU(A)*,and deciding *NEG*(A),respectively

In an approximation space apr = (U, E), where E is a binary relation, an element x is viewed as β_x (a subset of GKB containing x). Since β doesn't perform a partition on U in general, then we consider that $\bigcap \beta_x$ be a description x. The rough membership values $\mu_A(x) = P(A/\bigcap \beta_x)$ and $\mu_{A^C}(x) = 1 - P(A/\bigcap \beta_x)$ are in fact the probabilities that an object in $\bigcap \beta_x$ belongs to A and –A respectively. The expected of loss $R(a_i/\bigcap \beta_x)$ associated with taking the individual actions can be expressed as:

$$R(a_1/\beta_x) = \lambda_{11} P(A/\bigcap \beta_x) + \lambda_{12} P(-A/\bigcap \beta_x)$$
$$R(a_2/\beta_x) = \lambda_{21} P(A/\bigcap \beta_x) + \lambda_{22} P(-A/\bigcap \beta_x)$$
$$R(a_3/\beta_x) = \lambda_{31} P(A/\bigcap \beta_x) + \lambda_{32} P(-A/\bigcap \beta_x)$$
$$R(a_4/\beta_x) = \lambda_{41} P(A/\bigcap \beta_x) + \lambda_{42} P(-A/\bigcap \beta_x)$$
$$R(a_5/\beta_x) = \lambda_{51} P(A/\bigcap \beta_x) + \lambda_{52} P(-A/\bigcap \beta_x)$$

The Bayesian decision procedure leads to the following minimum-risk decision rules:

(1) $If\ R(a_1/\bigcap \beta_x) \le R(a_2/\bigcap \beta_x), R(a_1/\bigcap \beta_x) \le R(a_3/\bigcap \beta_x), R(a_1/\bigcap \beta_x) \le R(a_4/\bigcap \beta_x),$
$R(a_1/\bigcap \beta_x) \le R(a_5/\bigcap \beta_x), \textbf{\textit{decide}}\ POS(A).$

(2) $If\ R(a_2/\bigcap \beta_x) \le R(a_1/\bigcap \beta_x), R(a_2/\bigcap \beta_x) \le R(a_3/\bigcap \beta_x), R(a_2/\bigcap \beta_x) \le R(a_4/\bigcap \beta_x),$
$R(a_2/\bigcap \beta_x) \le R(a_5/\bigcap \beta_x), \textbf{\textit{decide}}\ SemiL(A) - \underline{A}_\beta.$

(3) $If\ R(a_3/\bigcap \beta_x) \le R(a_1/\bigcap \beta_x), R(a_3/\bigcap \beta_x) \le R(a_2/\bigcap \beta_x), R(a_3/\bigcap \beta_x) \le R(a_4/\bigcap \beta_x),$
$R(a_3/\bigcap \beta_x) \le R(a_5/\bigcap \beta_x), \textbf{\textit{decide}}\ SemiBND(A).$

(4) $If\ R(a_4/\bigcap \beta_x) \le R(a_1/\bigcap \beta_x), R(a_4/\bigcap \beta_x) \le R(a_2/\bigcap \beta_x), R(a_4/\bigcap \beta_x) \le R(a_3/\bigcap \beta_x),$
$R(a_4/\bigcap \beta_x) \le R(a_5/\bigcap \beta_x \beta_x), \textbf{\textit{decide}}\ \overline{A}_\beta - SemiU(A).$

(5) $If\ R(a_5/\bigcap \beta_x) \le R(a_1/\bigcap \beta_x), R(a_5/\bigcap \beta_x) \le R(a_2/\bigcap \beta_x), R(a_5/\bigcap \beta_x) \le R(a_3/\bigcap \beta_x),$
$R(a_5/\bigcap \beta_x) \le R(a_4/\bigcap \beta_x), \textbf{\textit{decide}}\ NEG(A).$

Since $P(A/\bigcap \beta_x) + P(-A/\bigcap \beta_x) = 1$, the above decision rules can be simplified such that only the probabilities $P(A/\bigcap \beta_x)$ are involved.

Consider a special kind of loss function with
$\lambda_{11} \le \lambda_{21} \le \lambda_{31} \le \lambda_{41} < \lambda_{51}$, $\lambda_{52} \le \lambda_{42} \le \lambda_{32} \le \lambda_{22} < \lambda_{12}$.
For this type of loss functions, the minimum-risk decision rules (1)-(5) can be written as:

$(1')$ $If\ P(A/\bigcap \beta_x) \ge b, P(A/\bigcap \beta_x) \ge c, P(A/\bigcap \beta_x) \ge d, P(A/\bigcap \beta_x) \ge e, \textbf{\textit{decide}}\ POS(A)$

$(2')$ $If\ P(A/\bigcap \beta_x) \le b, P(A/\bigcap \beta_x) \ge f, P(A/\bigcap \beta_x) \ge g, P(A/\bigcap \beta_x) \ge l, \textbf{\textit{decide}}\ SemiL(A) - \underline{A}_\beta$

$(3')$ $If\ P(A/\bigcap \beta_x) \le c, P(A/\bigcap \beta_x) \le f, P(A/\bigcap \beta_x) \ge m, P(A/\bigcap \beta_x) \ge n, \textbf{\textit{decide}}\ SemiBND(A)$

$(4')$ $If\ P(A/\bigcap \beta_x) \le d, P(A/\bigcap \beta_x) \le g, P(A/\bigcap \beta_x) \le m, P(A/\bigcap \beta_x) \ge q, \textbf{\textit{decide}}\ \overline{A}_\beta - SemiU(A)$

$(5')$ $If\ P(A/\bigcap \beta_x) \le e, P(A/\bigcap \beta_x) \le l, P(A/\bigcap \beta_x) \le n, P(A/\bigcap \beta_x) \le q, \textbf{\textit{decide}}\ NEG(A)$

where,

$$b = \frac{\lambda_{12} - \lambda_{22}}{(\lambda_{21} - \lambda_{22}) - (\lambda_{11} - \lambda_{12})}$$

$$c = \frac{\lambda_{12} - \lambda_{32}}{(\lambda_{31} - \lambda_{32}) - (\lambda_{11} - \lambda_{12})}$$

$$d = \frac{\lambda_{12} - \lambda_{42}}{(\lambda_{41} - \lambda_{42}) - (\lambda_{11} - \lambda_{12})}$$

$$e = \frac{\lambda_{12} - \lambda_{52}}{(\lambda_{51} - \lambda_{52}) - (\lambda_{11} - \lambda_{12})}$$

$$f = \frac{\lambda_{22} - \lambda_{32}}{(\lambda_{31} - \lambda_{32}) - (\lambda_{21} - \lambda_{22})}$$

$$g = \frac{\lambda_{22} - \lambda_{42}}{(\lambda_{41} - \lambda_{42}) - (\lambda_{21} - \lambda_{22})}$$

$$l = \frac{\lambda_{22} - \lambda_{52}}{(\lambda_{51} - \lambda_{52}) - (\lambda_{21} - \lambda_{22})}$$

$$m = \frac{\lambda_{32} - \lambda_{42}}{(\lambda_{41} - \lambda_{42}) - (\lambda_{31} - \lambda_{32})}$$

$$n = \frac{\lambda_{32} - \lambda_{52}}{(\lambda_{51} - \lambda_{52}) - (\lambda_{31} - \lambda_{32})}$$

$$q = \frac{\lambda_{42} - \lambda_{52}}{(\lambda_{51} - \lambda_{52}) - (\lambda_{41} - \lambda_{42})}$$

A loss function should be chosen in such a way to satisfy the conditions:

$$b \geq f, b \geq g, b \geq l$$
$$c \geq m, c \geq n, f \geq m, f \geq n$$
$$q \leq d, q \leq g, q \leq m.$$

These conditions imply that $\left(SemiL(A) - \underline{A}_{\beta}\right) \cup SemiBND(A) \cup \left(\overline{A}_{\beta} - SemiU(A)\right)$ *is* not empty, that is, the boundary region is not empty.

Worked Example (2)

In *Example (1),* we choose any subset A from U, say;

$$A = \{x_1, x_3, x_4, x_6\}$$

Now, we can decide the region for each object by using the generalized decision-theoretic approach that proposed in ***Section 3.3***. This approach can be apply on a multivalued information system and give us the ability to divide the universe U into five regions which help in increasing the decision efficiency. The result given by general rough sets model can be viewed as a special case of our generalized approach.

In our example, the set of states is given by $\Omega = \{A, -A\}$ indicating that an element is in A and not in A, respectively. With respect to five regions, the set of actions is given by $\mathcal{A} = \{a_1, a_2, a_3, a_4, a_5\}$

To apply our proposed technique, consider the following loss function:

$\lambda_{11} = \lambda_{52} = 0$, $\lambda_{21} = \lambda_{42} = 0.25$, $\lambda_{31} = \lambda_{32} = 0.5$, $\lambda_{41} = \lambda_{22} = 1$, $\lambda_{51} = \lambda_{12} = 2$

There is no cost for a correct classification, 2 units of cost for an incorrect classification, 0.25 unit cost for an object belonging to A is classified in $SemiL(A) - \underline{A}_\beta$ and for an object does not belong to A classified into $\overline{A}_\beta - SemiU(A).$, 0.5 unit cost for classifying an object into boundary region and 1 unit cost for classifying an object belong to A into $\overline{A}_\beta - SemiU(A).$ and for an object does not belong to A into $SemiL(A) - \underline{A}_\beta$,(note that a loss function supplied by user or expert) . According to these losses, we have

b = 0.8 g = 0.5
c = 0.75 l = 0.36
d = 0.64 m = 0.33
e = 0.5 n = 0.25
f = 0.67 q = 0.2

By using the decision rules $(1') - (5')$, we get the following results

	P(A/ $\cap \boldsymbol{\beta}_x$ ***)***	***Decision***
x_1	1	$POS(A)$
x_2	0.75	$SemiL(A)$- $\underline{A}_\beta$
x_3	1	$POS(A)$
x_4	0.25	$\overline{A}_\beta$ -*SemiU(A)*
x_5	0	*NEG(A)*
x_6	1	$POS(A)$
x_7	0	*NEG(A)*
x_8	0	*NEG(A)*
x_9	0	*NEG(A)*
x_{10}	0.5	*SemiBND(A)*

Thus we have,

POS (A) = $\{ x_1, x_3, x_6 \}$
SemiL(A) - $\underline{A}_\beta$ = $\{ x_2 \}$
SemiBND(A) = $\{ x_{10} \}$
$\overline{A}_\beta$ -*SemiU (A)* = $\{ x_4 \}$

NEG (A) = { x_5, x_7, x_8, x_9 }

Now we apply the decision theoretic technique proposed by ***Yao*** et. al. (1992) to classify the decision region into five areas. The set of actions is given by $\mathcal{A} = \{a_1, a_2, a_3\}$, where $a_1, a_2, and\ a_3$ represent the three actions in classifying an object, deciding *POS(A)* deciding *NEG(A)* and deciding *BND(A)* respectively. To make this, consider that there is 0.25 unite cost for a correct classification, 3 units of cost for an incorrect classification and 0.5 unit cost for classifying an object into boundary region, i.e.,

$$\lambda_{11} = \lambda_{22} = 0 \quad , \quad \lambda_{31} = \lambda_{32} = 0.5 \quad , \quad \lambda_{21} = \lambda_{12} = 3 \ .$$

These losses give us that,

$$\alpha = 0.9 \quad , \quad \gamma = 0.5 \quad , \quad \beta = 0.1$$

By using the decision rule (P')-(C'), and replacing $P(A/[x]_E)$ by $P(A/\beta_x)$, we get on the following results:

	P(A/ $\cap$ ***β_x)***	***Decision***
x_1	1	$POS(A)$
x_2	0.75	*BND(A)*
x_3	1	$POS(A)$
x_4	0.25	*BND(A)*
x_5	0	*NEG(A)*
x_6	1	$POS(A)$
x_7	0	*NEG(A)*
x_8	0	*NEG(A)*
x_9	0	*NEG(A)*
x_{10}	0.5	*BND(A)*

This means that

POS (A) = { x_1, x_3, x_6 }
BND (A) = { x_2, x_4, x_{10} }
NEG (A) = { x_5, x_7, x_8, x_9 }

From comparesion between two approches, we note that the our approach (classification of decision region into five eareas) give us the ability to divid *BND (A)* = { x_2, x_4, x_{10} } into *SemiL(A)* - $\underline{A}_\beta$ = { x_2 }**,** which closer to the positive region**,** $\overline{A}_\beta$ *-SemiU (A)* = { x_4 }, which closer to the negative region, and *SemiBND(A)* = { x_{10} }.

4. Conclusion

The decision theoretic rough set theory is a probabilistic generalization of standard rough set theory and extends the application domain of rough sets. The decision model can be interpreted in terms of more familiar and interpretable concept known as loss or cost. One can easily interpret or mesure loss or cost according to real application. We have proposed in this paper a genralized decision theoretic approach, which applied under granulated the universe by any binary relation. This approach enable us to classify the decision region into five areas. This classification will enlarge of choice for decision maker and help in increasing the decision efficiency.

Refrences

[1] Duda, R.O. and Hart, P.E.(1973). Pattern Classification and Scene Analysis, Wiley, New York.

[2] Ezzat, M. M. (2004). Statistis and Roughian. Phd. Thesis, Faculty of Science, Tanta University, Egypt.

[3] Lashein, E., Kozae, A.M., Abo Khadra, A. and Medhat, T. (2005). Rough Set Theory for Topological Spaces. Int. J. of Approx. Reasoning, 40, 35-43.

[4] Pawlak, Z.(1982). Rough Sets. International Jornal of Computer Information Sciences,11,314-356.

[5] Pawlak, Z. and Skowron, A. (1994). Rough Membership Functions, in: Advances in the Dempster-Shafer Theory of Evidence, R.R. Yager and M. Fedrizzi and J. Kacprzyk (Eds.), John Wiley and Sons, New York, 251-271.

[6] Rady, E.A, Kozae, A.M and Abd El-Monsef, M.M.E.(2004). Generalized Rough Sets. Chaos, Solitons, & Fractals, 21, 49-53.

[7] Yao,Y.Y.(2003). Information Granulation and Approximation in a Decision-theoretic Model of Rough Sets, in: Rough-Neural Computing: Techniques for Computing with Words, Pal, S.K., Polkowski, L., and Skowron,A.(Eds), Springer, Berlin, pp. 491-518.

[8] Yao, Y.Y, Wong, S.K.M and Lingras,P.(1990). A Decision-theoretic Rough Set Model. In Z.W.Ras,M.Zemankova & M.L.Emrich,Eds. Methodologies for Intelligent System,5,17-24,New York:North-Holanda.

[9] Yao, Y.Y and Wong, S.K.M.(1992). A decision Theoretic Framework for Approximating Concepts. International Gournal of Man-machine Studies,37,793-809.

Egyptian Rough Sets Working Group

The first one Day Workshop on Rough Sets and their Applications – WRSTA2006

Alexandria University, Faculty of Science, Mathematics Department, 13 August, 2006

Evolution Rough Sets

Osama Badawy[1], Yasser Foaud [2], Moustafa Fahmy[1], Wedad Sagar [2]

[1]College of Computing and Information Technology, Arab Academy for Science &Technology & Maritime Transport, Alexandria, Egypt
[2]Department of Mathematics (Computer, Science), Alexandria University, Egypt

Abstract: Rough set theory, which emerged about 20 years ago, is nowadays a rapidly developing branch of artificial intelligence and soft computing. In this paper, we present a new approach (RS_DT) for construction of decision tree based on rough set theory, which will be induced a simplified tree. This tree is transformed into an initial population of genetic probabilistic rough induction (GA+GDT-RS), which is a combination between probabilistic rough induction and genetic algorithm. We will use probabilistic rough induction for modeling a classification system and applying genetic operators to a population of chromosomes. However, it is interesting to try to incorporate these approaches into the hybrid system. The challenge is to get as much as possible from this association.

Keywords: Generalization Distribution Table, Rough Sets, Genetic Algorithm, Machine learning, Data mining

1. Introduction

Historically different approaches for knowledge extraction evolved [14], such as symbolic approaches and computational learning theory. Among them we can find many classical approaches, like decision trees, rough sets, case based reasoning, neural networks, support vector machines, different fuzzy methodologies, ensemble methods [5], but they all have some advantages and limitation. Evolutionary approaches (EA) are also a good alternative, because they are not inherently limited to local solutions [7]. Recently, taking into account the limitations of classical approaches many researches focused their research on hybrid approaches.

Current studies show that the selection of appropriate method for data analysis can be crucial for the success. Therefore, for a given problem, different methods should be tried to increase the quality of extracted knowledge. According to the previous paragraph a logical step would also be to combine different methods into one more complex methodology in order to overcome the limitation of a single method.
The paper is organized as follows: section 2 introduces a basic model, section 3 shows the new approach (RS_DT) for construction of decision tree based on rough set

theory, section 4 discusses our hybrid method in detail. Applications and results in section 5, section 6 concludes our paper.

2. Basic Models

2.1 Rough Set Methodology

In the rough set methodology for rule discovery, data base is regarded as a decision table, which is denoted $T = (U, A, \{Va\}_{a\in A}, f, C, D)$, where U is a finite set of instances (or objects), called the universe, A is a finite set of attributes, each Va is the set of values of attribute a, f is a mapping from $U \times A$ to $V(= \bigcup_{a\in A} Va)$, C and D are two subsets of A, called the sets of attributes and decision attributes, respectively, such that $C \cup D = A$, and $C \cap D = \phi$. Equivalence classes in U/C and U/D are called condition classes and decision classes, respectively [9], [10], [12], [13].

The process of rule discovery is that of simplifying a decision table and generating minimal decision algorithm. In general, an approach for decision table simplification consists of the following steps:

(1) Elimination of duplicate condition attributes. It is equivalent to elimination of some columns from the decision table.

(2) Elimination of duplicate rows.

(3) Elimination of superfluous values of attributes.

A representative approach for the computation of reducts of condition attributes is to represent knowledge in the form of a discernibility matrix [12], [13]. The basic idea can be briefly presented as follows:

Let $T = (U, A, \{Va\}_{a\in A}, f, C, D)$, be a decision table with U= $\{u_1, u_2 \ldots, u_n\}$. By a discernibility matrix of T, denoted M(T), we will mean nxn matrix defined as:

$$mij = \begin{cases} \{c \in C : c(ui) \neq c(uj) \text{ if } \exists d \in D\,[d(ui) \neq d(uj)] \\ \lambda \qquad \text{if } \forall d \in D\,[d(ui) = d(uj)] \end{cases}$$

for i, j=1,2,…,n such that u_i or u_j belongs to the C-Positive region of D.

Thus entry mij is the set of all the condition attributes that classify objects u_i and u_j into different decision classes in U/D. Since M(T) is symmetric and $mii = \phi$, M(T) are represented only by elements in the lower triangle, that is, the mij with $1 \le j < i \le n$. Furthermore, $mij = \lambda$ denotes that this case does not need to be considered. Hence it is interpreted as logic truth. The discernibility function f_T for T is defined as follows:

for any $ui \in U$ $\quad f_T(ui) = \wedge_j \{\vee mij : j \neq i, j \in \{1,2,\ldots,n\}\}$ where

i) $\vee mij$ is the disjunction of all variables such that $c \in mij$ *if* $mij \neq \phi$

ii) $\vee mij = \perp$ (*false*), *if* $mij = \phi$

iii) $\vee mij = T$ (*true*), *if* $mij = \lambda$

Each logical product in the minimal disjunctive normal form (DNF) of f_T(ui) is called a reduct of instance ui. Generating minimal decision algorithm is to eliminate the superfluous decision rules associated with the some decision class. It is obvious that some decision rules can be dropped without disturbing the decision-making process, since some other rules can take over the job of the eliminated rules.

2.2 GDT-RS Principles

GDT-RS is a soft hybrid induction system for discovering classification rules from databases with uncertain and incomplete data [15]. The system is based on hybridization of GDT and the rough set methodology.

2.2.1 Generalization Distribution Table (GDT)

Any GDT consists of three components: possible instances, possible generalizations of instances, and probabilistic relationships between possible instances and possible generations. Here the possible instances are all possible combinations of attribute values in a database; the possible generations for instances are all possible cases of generation for all possible instances; the probabilistic relationships between possible instances and possible generations, represented by entries Gij of a given GDT, are defined by means of a probabilistic distribution describing the strength of the relationship between any possible instance and any possible generation. The prior distribution is assumed to be uniform, if background knowledge is not available. Thus, it is defined by the following Equation:

$$G_{ij} = p(PI_j / PG_i) = \begin{cases} 1/N_{PGi} \text{ if } PI_j \supset PGi \\ 0 \quad \text{otherwise} \end{cases} \tag{1}$$

Where PIj is the jth possible instance, PGi is the ith possible generalization. and N_{PGi} is the number of the possible instances (PI) satisfying the ith possible generation, i.e.,

$$N_{PGi} = \prod_{j}^{m} n_j \tag{2}$$

Where j=1,..m and j is not equal the attribute that is contained by the ith possible generalization.

2.2.2 Rule Strength

The rules are expressed in the following from:

P→ Q with S

i.e., "if P then Q with the strength S" where P denotes a conjunction of conditions (i.e., P ⊆ C), Q denotes a concept that the rule describes (i.e., Q ⊆ D), S is a "measure of strength" of the rule. Furthermore, S consists of three parts: s(P), accuracy and coverage, where s(P) is the strength of the generalization P (i.e., the condition of the rule), the accuracy of the rule is measured by a noise rate function: r(P → Q), coverage denotes how many instances are covered by the rule. If some instances covered by the rule also belong to another class, the coverage is a set :{ number of instances belonging to the class, number of instances belonging to another class}.

The strength of a given rule reflects the incompleteness and uncertainly in the process of rule inducing influenced both by unseen instances and noise. The strength of the generalization P=PG is given by the following equation under that assumption that the prior distribution is uniform:

$$s(P)=\sum_j p(PI_j / P)=card([P]) \times 1/N_P \qquad (3)$$

where card([P]) is the number of observed instances satisfying the generation P. The strength of the generalization P represents explicitly the prediction for unseen instances since possible instances are considered. On other hand, the noise rate is given by Eq. (3)

$$r(P\rightarrow Q)=card([P]\cap[Q])/card([P]) \qquad (4)$$

where card([P]∩[Q]) is the number of all instances from class Q within the instances satisfying the generalization P.

2.2.3 Searching Algorithm for an Optimal Set of Rules

In [16] they outline the idea of searching algorithm for a set f rules. They use a sample decision table shown in Table 1 to illustrate the idea. Let Tnoise be a threshold value.

Table 1. A sample database

No	a	b	c	d
u_1	a_o	b_o	c_1	y
u_2	a_o	b_1	c_1	y
u_3	a_o	b_o	c_1	y
u_4	a_1	b_1	c_0	n
u_5	a_o	b_o	c_1	n
u_6	a_o	b_2	c_1	n
u_7	a_1	b_1	c_1	y

- **Step 1.** Create GDT, if prior background knowledge is not available, the prior distribution of a generalization is calculated using Eqs.(1) and (2).
- **Step 2**. Consider the indiscernibility classes with respect to the condition attribute set C (such as u_1,u_3 and u_5 in the sample database of Table 1) as one instance, called the compound instance (such as $u_1'=[u_1]IND(a,b,c)$ in the following Table 2).

Table 2: Table 1 after makes compound instances

U \ A	a	b	c	d
$u_1'(u_1,u_3,u_5)$	a_o	b_o	c_1	y,y,n
u_2	a_o	b_1	c_1	y
u_4	a_1	b_1	c_o	n
u_6	a_o	b_2	c_1	n
u_7	a_1	b_1	c_1	y

- **Step 3.** For any compound instance u' (such as the instance u_1' in the above table), let d(u') be the set of the decision classes to which the instances in u' belong. Furthermore, Let $Xv=\{x\in U:d(x)=v\}$ be the decision class corresponding to the decision value v. The rate r_v can be calculated by Eq(4). If there exists a $v\in d(u')$ such that $r_v(u')=\min\{ r_v(u') \mid v'\in d(u')\} <$ Tnoise, then they let the compound instance u' point to the decision class corresponding to v. If there is no $v \in d(u')$ such that $r_v(u')<$ Tnoise, they treat the compound instance u' as a contradictory one, and set the decision class of u' to $\perp$ (uncertain).

For example, we have

U \ A	a	b	c	d
u_1'(u_1,u_3,u_5)	a_o	b_o	c_1	⊥

Let U' be the set of all the instances except the contradictory ones.

- **<u>Step 4.</u>** Select one instance u from U'. Using the idea of discernibility matrix, create a discernibility vector (i.e. the row or the column with respect to u in the discernibility matrix) for u. For example, the discernibility vector for instance u_2:$a_0b_1c_1$ is as follows:

U \ U	u_1'(⊥)	u_2(y)	u_4(n)	u_6(n)	u_7(y)
u_2(y)	b	**Ø**	a,c	b	Ø

- **<u>Step 5.</u>** Compute all the so-called local relative reducts for instance u by using the discernibility function. For example, from instance u_2:$a_0b_1c_1$, they obtain two reducts: {a,b} and {b,c}.
- **<u>Step 6.</u>** Construct rules from the local reducts for instance u, and revise the strength of each rule using (3). For example, the following rules are acquired:

  {a_0 b_1}⟶y with S=1 X ½ = 0.5 , and

  {b_1 c_1}⟶y with S=2 X ½ =1

  for instance u_2 :$a_0b_1c_1$
- **<u>Step 7.</u>** Select the best rules from the rules (for u) obtained in step 6 according to its priority :
  - Selecting the rules that contain as many instances as possible.
  - selecting the rules that contain as little attributes as possible, if they cover the same number of instances
  - Selecting the rules with larger strength, if they have same number of condition attributes and cover the same number of instances.

For example, the rule '{b_1c_1}⟶y' is selected for the instance u_2:$a_0b_1c_1$ because it matches more instances than the rule '{a_0b_1}⟶y' .

- **<u>Step 8.</u>** U'=U'-{u}. If U ≑ Ø , then go back to step 4. Otherwise, goto setp 9.
- **<u>Step 9.</u>** Finish if the number of rules selected in step 7 for each instance is 1. Otherwise find a minimal set of rules, which contains all of the instances in the decision table.

The following Table 3, shows the result for the sample database shown in Table 1.

Table 3: Results for a sample database

U	Rules	Strength
u_2,u_7	$b_1 \wedge c_1 \rightarrow y$	1
u_4	$c_o \rightarrow n$	0.167
u_6	$b_2 \rightarrow n$	0.25

2.3 Genetic Algorithm

The origin of Genetic Algorithms (GAs) is attributed to Holland's [3] works on cellular automata. There has been significant interest in GA over last two decades. The range of applications of GA includes such diverse areas as job shop scheduling, training neural nets, image feature extraction, and image feature identification [2].

A genetic algorithm is a search process that follows the principles of evolution through natural selection [2]. The domain knowledge is represented using a candidate solution called an organism or chromosome. Typically, an organism is a single genome represented as a vector of length n: $c = (c_i | 1 \le i \le n)$, where c_i is called a gene.

An abstract view of a generational GA is given in Figure 1. A group of organisms is called a population. Successive populations are called generations. A generational GA starts from initial generation G(0) , and for each generation G(t) generates a new generation G(t+1) using genetic operators such as mutation and crossover. The mutation operator creates new genomes by changing values of one or more genes at random as shown in Figure 2. The crossover operator joins segments of two or more genomes to generate a new genome. Figure 3 depicts an example of a crossover operation.

Genetic Algorithm:

```
Generate initial population   G(0);
Evaluate G(0);
For (t=1;solution is not found;t++)
{ generate G(t) using G(t-1);
evaluate G(t); }
```

Figure 1. Abstract view of a generational genetic algorithm

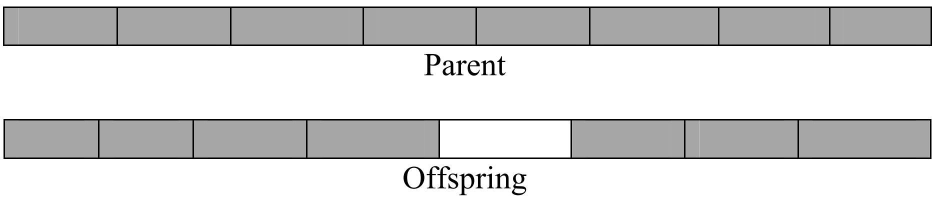


Figure 2. Mutation operation

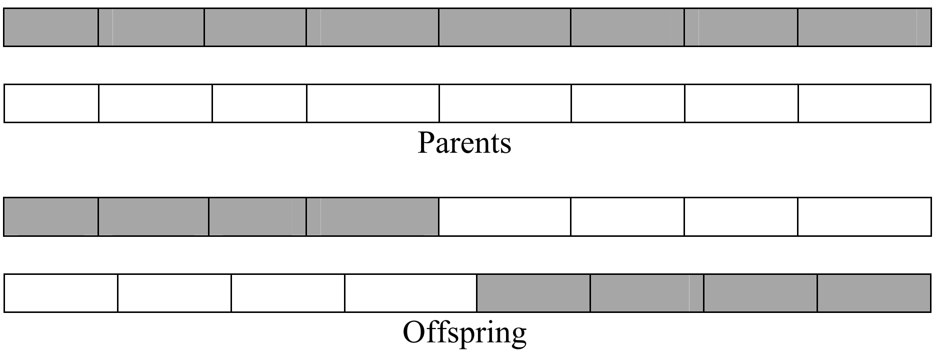


Figure 3. Crossover operation

3. Proposed Method of Decision Tree (RS_DT)

In many literatures [4], [9], rough set has been used for selecting attributes, consequently a reduct of attributes will be found which is regarded as the best reduction of attributes. The goal is to reduce the volume of data. In this paper, we introduced a new approach of decision tree construction based on rough sets [1]. The basic algorithm for construction of decision tree is as follows. The input of the algorithm for construction of decision tree based on discernibility vector is a collection of data. The output is a decision tree corresponding to the data.

The algorithm RS_DT is presented in recursive form.

RS_DT algorithm

Input: U= a data collection.
A = condition attributes.
RS_DT (U), parameter U indicates collection of data.

Processing:

- **Step 1:** Compute the discernibility vector for all instances in U.
- **Step 2:** Compute the redundancy of all condition attributes in all discernibility vectors.
- **Step 3:** Choose the attribute with the maximal number as the node of this layer of current branch.
- **Step 4:** Construct decision tree on the current branch according to the possible values of the selected attribute that appear in U.
- **Step 5:** For each branch of the selected attribute of the decision tree, if it has not reached leaf node then call RS_DT(U) with the collection of the data of this branch as its parameter.
- **Step 6:** Return.

Output: decision tree.

Worked Example:

We will illustrate the main idea of our method by way of an example. Suppose, we have information table that shown in Table 4.

Table 4. A sample Information Table

U \ A	a	b	c	d
u_1	a_o	b_1	c_1	y
u_2	a_o	b_o	c_1	y
u_3	a_1	b_1	c_o	n
u_4	a_o	b_2	c_1	n
u_5	a_1	b_1	c_1	y

Suppose, we have information table that shown in table4. Let U be a set of all instances, select one instance u from U. Using the idea of discernibility matrix, create a discernibility vector (i.e., the row or column with respect to u in the discernibility matrix) for u. For the example $u_1=a_0\ b_1\ c_1$ is as in table 5, where λ means that we don't care the element since it is with the same class. And so on u_2(y), u_3(n), u_4(n), u_5(y) compute discernibility vector.

Table 5: First discernibility vectors of instances

U \ U	U_1(y)	u_2(y)	u_3(n)	u_4(n)	u_5(y)
u_1 (y)	λ	λ	a,c	b	λ
u_2 (y)	λ	λ	a,b,c	b	λ
u_3 (n)	a,c	a,b,c	λ	λ	c
u_4 (n)	b	b	λ	λ	a,b
u_5 (y)	λ	λ	c	a,b	λ

Now, compute the redundancy of each attributes condition in all discernibility vectors. The attribute which has maximal number is the root. The redundancy of 'b' is 8 and redundancy of 'a' the same as redundancy of 'c' is 6. Therefore attribute 'b' is selected as the root of the decision tree, and the data set will be partitioned into three data subsets. In each data subset, all tuples have the same value of the attribuate 'b', if all tuples in the same branch have the same class label then no action is needed to construct the decision tree to a deeper layer. Otherwise, another attribute will have to be chosen as node to construct the decision tree to further depth until the tree reaches leaf node that all tuples in the leaves have the same class label, see Figure 4.

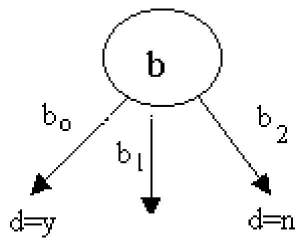


Figure 4. The root of decision tree

Since condition attribute 'b' has been used in upper layer of the tree, only the other two condition attributes (attribute 'a' and 'c') are tested on branch $b=b_1$, we obtain the discernibility vector of those instances u_1, u_3, u_5 as see in Table 6, and compute the reducency of attributes condition 'a' and 'c'.

Table 6: Second discernibility vectors of instances

U \ U	u_1(y)	u_3(n)	u_5(y)
u_1 (y)	λ	a,c	λ
u_3 (n)	a,c	λ	c
u_5 (y)	λ	c	λ

Since condition attribute 'c' has maximal number (equal 4) , 'c' is selected as the winner of branch $b=b_1$. the class labels of the leaf node of the two branches $c=c_0$ and $c=c_1$ are labelled in accordance with the class labels of the tuples of the corresponding branches, i.e labelled to be d=y and d=n.

Since $b=b_0$ and $b=b_2$ need no further classification, the final decision tree can be obtained as shown in Figure 5.

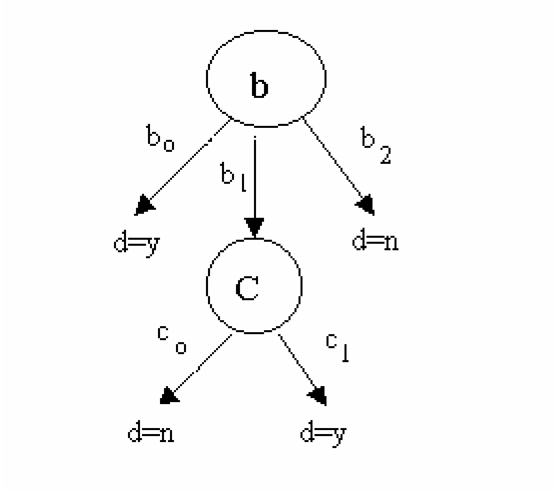


Figure 5. Final of decision tree

From a root node to leaf node, rules can be obtained

r_1. ($b=b_0$) → ($d=y$)

r_2. ($b=b_2$) → ($d=n$)

r_3. ($b=b_1$, $c=c_1$) → ($d=y$)

r_4. ($b=b_1$, $c=c_0$) → ($d=n$)

4. Hybrid Method (Evolution Rough Sets)

In this Section we describe the main characteristics of our method for classification. This is a hybrid method that combines decision tree (RS_DT) and genetic probabilistic rough induction (GA+GDT_RS). The basic idea is to use a decision tree (RS_DT) algorithm to produce rules before run the algorithm of genetic probabilistic rough induction (GA+GDT_RS). The method discovers rules into two training phases. In the first phase it runs RS_DT decision tree induction algorithm. The induced, simplified tree by Fisher's exact test [6] is transformed into a set of rules. This rule set can be thought of as expressed in an initial population. The second phase consists of using a GA+GDT_RS genetic probabilistic rough induction to discover rules. Show the algorithms in Figure 6.

4.1 Genetic Operators

To evolve a population of individuals we need to apply genetic operators that produce new individuals (which will from the next generation) from the best individuals of the current generation. We briefly describe below the genetic operators used by our hybrid system. We used a population size of 100 individuals a parameter value often used in the literature [3] and adequate for our experiment.

The reproduction operator was roulette wheel selection with elitism. Out of the 100 individuals, 98 are reproduced by roulette wheel selection and the best 2 individuals of the current population are passed on, unaltered, to the next generation, in an elitism-reproduction scheme. Elitism avoids the danger of possible loss of the best two individuals of the current generation. Note that this loss might occur if roulette wheel selection was used alone (without elitism), due to the stochastic nature of roulette wheel selection. The individuals selected for reproduction undergo the application of two other stochastic operators, namely mutation and crossover. In our experiments reported in this paper mutation is applied with 0.05 probability and crossover with 0.8 probabilities.

4.2 Fitness Function and Evaluation of the Set of Rules

Decision tree (RS_DT) method induces a set of rules. First we simplified these rules. Second, we delete from this set of rules any duplicate ones. Third, we use these rules as an initial population of a hybrid system genetic probabilistic rough induction GA+GDT_RS, after the number of iteration. Any weak rule which its accuracy value is less than some threshold, will be removed from the set of rules. Therefore, we evaluated the whole set of rules as Rosetta [17]. We only use training cases to generate rules and testing cases to measure the efficient of the method.

We define the fitness function of the individual depending on rule coverage by using this equation:

$$con=card([P]\cap[Q])/card([Q]) \qquad (5)$$

Where P,Q as we explained in section 2.2.2.

Depending on the fitness value, the genetic operators select the individual to be processed (the better the fitness, the more likely the individual is to be selected).

Algorithm of the Method

We now describe an idea of searching algorithm for a set of rules. We use a sample database shown in Table 4 to illustrate the idea. The input is data in the form of a decision table structure, the output is the set of rules, and the process is as follows.

- **Step 1**: Read the decision table, determine the thresholds Tnoise and λ for classification accuracy. If the decision table contains inconsistent data do steps1-3 in algorithm of GDT_RS, otherwise go to step 2.
- **Step 2**: By decision table generate a decision tree by running RS_DT algorithm, see section 3, simplified the tree and convert the tree as an initial population.
- **Step 3**: Iteratively perform the following substeps until the maximum of generation, G, is reached.

a. Evaluate each individual in the population via the following:

i. Calculate the classification accuracy and rule coverage for each rule (chromosome) as we explained in section 2.2.2, see Table 7, and remove the rule if its accuracy is less than the thresholds λ.

ii. Use a Eq.(5) as a fitness measure to assign a value for this set of rules.

Table 7: Results for a sample database

U	Rules	s(P)	Accuracy	Coverage
u_1'	$b_0 \rightarrow y$	0.25	0.67	1
u_2,u_7	$b_1 \wedge c_1 \rightarrow y$	1	1	2
u_4	$c_0 \rightarrow n$	0.17	1	1
u_6	$b_2 \rightarrow n$	0.25	1	1

b. Create a new population by applying the following operations:

i. Reproduce an existing individual (selected based on its fitness) and copy it into the new population.

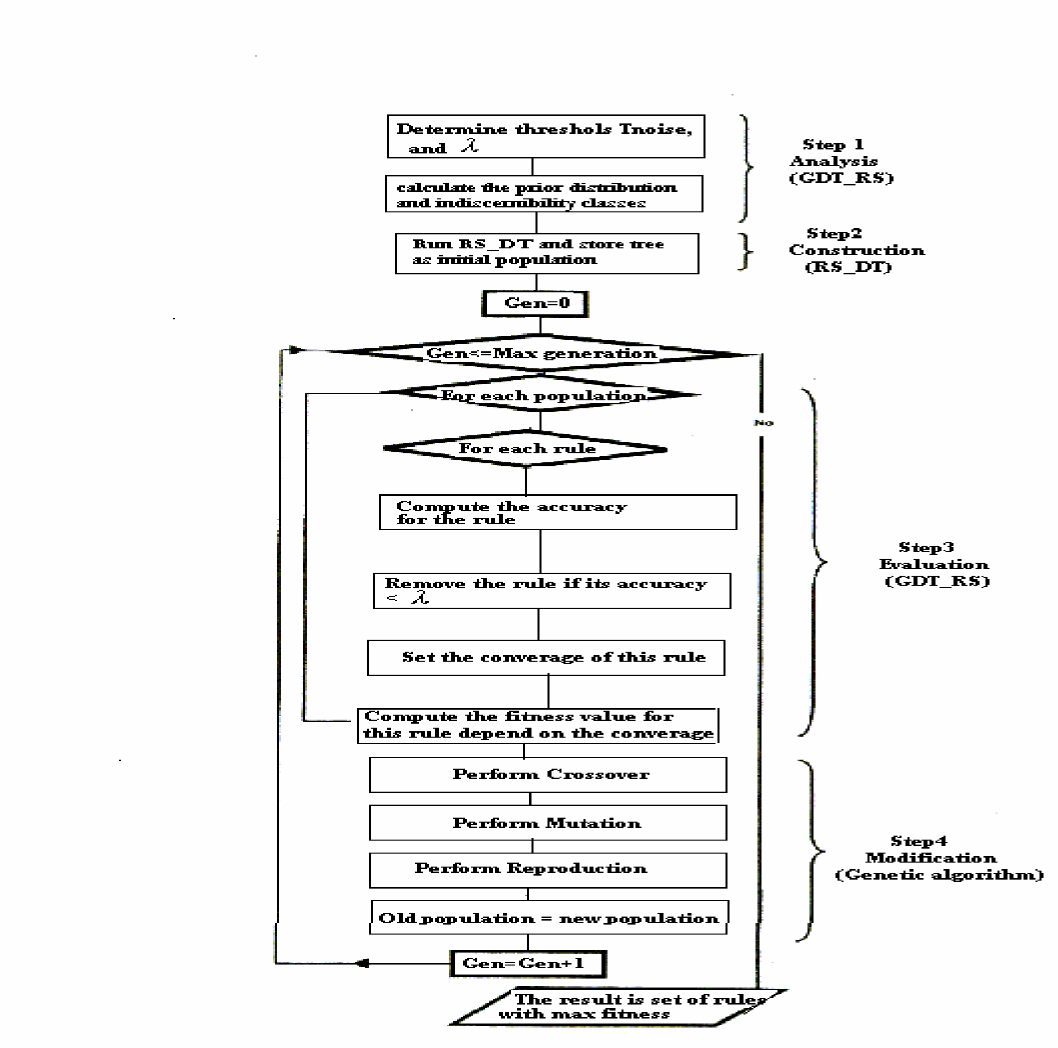


Figure 6: Algorithm of our hybrid method for classification.

ii. Create two new individuals from two existing individuals by genetically recombining randomly chosen parts of two existing chromosomes using crossover operation.

iii. Create a new individual from an existing one by mutating a randomly chosen part of the selected chromosome using the mutation operation.

iv. Insert these chromosomes into the new population.

- **Step 4.** The best-so-far individual (the individual that has the highest fitness value over all the generations) is designated as the result of the run.

5 Applications and Results

We have evaluated the performance of our hybrid method RS_DT/GA+GDT_RS across dataset taken from [11]. The experiment used RS_DT as the decision tree component of our hybrid method. The performance of the hybrid method RS_DT/GA+GDT_RS was compared with the performance of RS_DT method and standard rough set method for the number of rules and accuracy.

We divided this dataset to 80% training and 20% testing, the results of computations of rules have been only done to training data, and the accuracy computed on testing data. Figure 7 shows a comparison of the number of rules obtained from RS_DT method, standard rough set method and our hybrid method. Figure 8 presents the accuracy of dataset on RS_DT method, standard rough set method and our hybrid method.

Since we are started with decision tree RS_DT, and the results with RS_DT are the initial population of hybrid system GA+GDT_RS. So the results obtained from hybrid system GA+GDT_RS are the same as of decision tree RS_DT or better. The set of rules that is produced by our new method contains the fewest rules with better accuracy rate than that is produced by the decision tree RS_DT. For example, if we take Monk's data problem, in the third Monk's problem data there exist a rule produced from decision tree RS_DT: IF attribute#1=2 AND attribute#5=4 THEN Class 0. This rule is reduced in our method to the following: IF attribute#5=4 THEN Class 0. Some rules in decision tree (RS_DT) are presented in the list of rules produced form hybrid system (RS_DT/GA+GDT_RS) such as: IF attribute#2=3 AND attribute#5=2 THEN Class 0, because the fitness function is high, and other rules from RS_DT is omitted, because rules with high fitness will appeared. The hybrid method RS_DT/GA+GDT_RS proposed here is slower than a decision tree RS_DT method in the processing time. But on the other hand hybrid method RS_DT/GA+GDT_RS gives more accuracy and less number of rules than a decision tree RS_DT. So the incorporation of knowledge into evolutionary algorithms is interesting of optimization and machine learning problems.

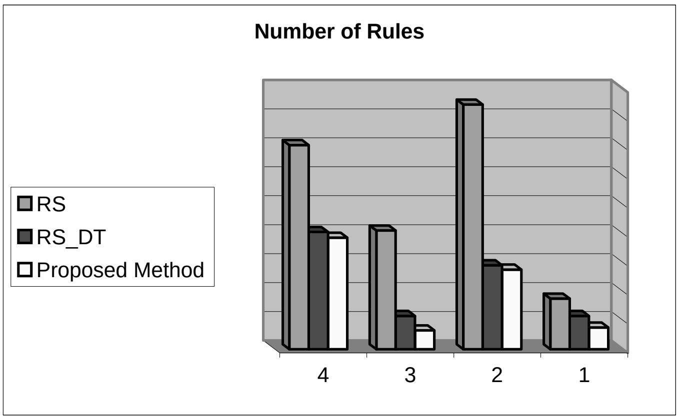


Figure 7 Comparison of number of rules on datasets

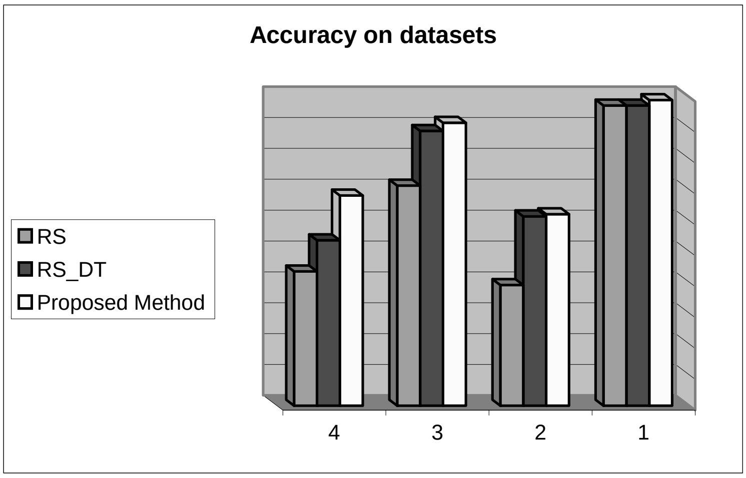


Figure 8 the accuracy on datasets

6 Conclusions

In this paper, we have described a hybrid system RS_DT/GA+GDT_RS and compared its performance with the performance of decision tree RS_DT and standard rough set method on some dataset. We showed that a hybrid system can be used to improve the performance of decision tree. A hybrid system as we saw with three goals: increase the accuracy of dataset; reduce the number of rules if we compared with decision tree (RS_DT) and standard rough set method. Also reduce the processing time if we compared with standard genetic algorithm method.

References

[1] Badawy, O., Foaud, Y., and Sagar, W., Rough Set Approach to Decision Tree Construction, Fourth International Conference on Informatics and System, Cairo, Egypt, 2006

[2] Buckles, B., and Petry, F., Genetic Algorithms, Los Alamitos, California: IEEE Computer Press. 1994

[3] Carvalho, R., and Freitas, A., A Genetic Algorithm with Sequential Niching for Discovery Small-Disjunt Rules, Proceedings Genetic and Evolutionary Computation Conf. (GECCO-2002), Morgan Kaufmann, pp.1035 – 1042, 2002.

[4] Choubey. S., Deogun. J., Raghavan. V., and Sever. H., A Comparison of Feature Selection Algorithms in the Context of Rough Classifiers. In Proceedings of Fifth IEEE International Conference on Fuzzy Systems (New Orleans, LA), Vol. 2, pp. 1122 – 1128. 1996

[5] Dietterich, G., Ensemble Methods in Machine Learning. In: Firt International Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science, Springer-Verlag, pp. 1-15, 2000.

[6] Finney, D., Latscha. R., Bennett. B., and Hsu. P., Tables for Testing Significance in a 2 ×2 Contingency Table, Cambridge University Press. 1963.

[7] Goldberg, D., Genetic Algorithm in search, Optimization and Machine Learning, Addison-Wesley, 1989.

[8] Holland, J., Adaptation in Natural and Artificial Systems, Ann Arbor: University of Michigan Press. 1975.

[9] Kohavi. R., and Frasca, B., Useful Feature Substets and Rough Set Reducts. In Proceedings of the Third International Workshop on Rough Sets and Soft Computing (RSSC'94). San Jose, California, 1994, pp. 310 –317.

[10] Lin, T., and Cercone, N ., (Eds). Rough Sets and Data Mining : Analysis of Imprecise Data . kluwer, 1997.

[11] Merz, C., and Murphy, P., UCI Repository of Machine Learning Database. http://www.ics.uci.edu/~mlearn/mlrepository.html. Dept. Information and Computer Science, University of Californi, Irvine, CA, 1998.

[12] Mollestad, T., and Skowron, A., A Rough Set Framework for Data Mining of Propositional Default Rules, in: Z . W. Ras , M. Michalewicz (Eds.), Proceedings Ninth International Symposium on Methodologies for Intelligent Systems (ISMIS – 96) , Lecture Notes in Artificial Intelligance, Springer, Berlin, Vol . 1079, pp. 448 – 457. 1996.

[13] Pawlak. Z., Rough Sets, Theoretical Aspects of Reasoning about Data. kluwer, 1991.

[14] Thrun, S., and Pratt, L., Learning to Learn, Kluwer Academic, 1998.

[15] Zhong. N., Dong. Z., Ohsuga. S., Data Mining: A Probabilistic Rough Set Approach. In: Polkowski. L., Skowron, A., (Eds.), Rough Sets in Knowledge Discovery, Physica, Vol. 2, pp. 127 – 146. 1998.

[16] Zhong. N., Dong. Z., Ohsuga. S., Rule Discovery by Soft Induction Techniques, Neurogomputing, 36, pp. 171 – 204. 2001.

[17] The Rosetta homepage. [http://www.idi.ntnu.no/~aleks/rosetta/]. Norwegian University of Science and Technology, Department of Computer and Information Science.

Egyptian Rough Sets Working Group

The first one Day Workshop on Rough Sets and their Applications – WRSTA2006

Alexandria University, Faculty of Science, Mathematics Department, 13 August, 2006

Reducing the Response Time for Data Warehouse Queries Using Rough Set Theory

Mahmoud Mohamed Al-Bouraie, Yasser Fouad Hassan, Wesam Fathy Jasser

Department of Math.(Computer Science), Faculty of Science

Email: y_fouad@hotmail.com & wesamjasser@yahoo.com

Abstract: One of the most important problems in relational databases applications like data warehouses is a size of real-world databases. In practical problems, data may contain millions of records in many data tables bounded by relations. One approach to reduce the size of these applications is rough set theory. The Rough set theory is deeply investigated, and an approach for data filtering based on rough set theory is proposed. In addition, to improve processing of star queries on data and processing of aggregation star queries a new optimization technique is the called pre-grouping transformation. Although this transformation is expected to reduce the time needed to answer large aggregation queries to less than 50%, there are several cases where it is not beneficial. In this paper we retry reach to the optimization case by applying the two above theories.

Keywords: Rough sets, data warehouse, Transformation

1. Introduction

The topic of data warehousing encompasses architectures algorithms and tools for bringing together selected data from multiple databases or other information sources into a single repository, called a data warehouse, suitable for direct querying or analysis. In recent years, data warehousing has become a prominent buzzword in the database industry, but attention from the database research community has been limited. In this paper, we try to reduce the size of the database using rough sets theory and reducing the response time at queries using new technique called pre-grouping transformation. Rough sets theory was first introduced by Pawlak in the 1980s and it has been applied in many applications such as machine learning, knowledge discovery, and expert systems. Rough set is especially useful for domains where data collected are imprecise and/or incomplete about the domain objects. It provides powerful tools for data analysis and data mining from imprecise and ambiguous data. Many rough set models have been developed in the rough set community in the last decades. Some of them have been applied in the industry data mining projects such as stock market prediction, patient symptom diagnosis, telecommunication churner prediction, and financial bank customer attrition analysis to solve challenging business problems. These rough set models focus on the extension of the original model proposed by Pawlak and attempt to deal with its limitations such as handling statistical distribution or noisy data.

A methodology recently proposed to improve processing of star queries on data warehouses is the clustering and indexing of fact tables using their multidimensional hierarchies. Due to this improved organization schema, processing of aggregation star queries changes dramatically creating new optimization opportunities. An important optimization technique is the so-called pre-grouping transformation.

The paper is organized as follows. In section 1, we give an overview of rough set theory. Section 2 describes the general concept of pre-grouping transformation, while section 3 shows the measurement results are presented.

2. Rough Set Theory

Rough set theory [1] is an extension of conventional set theory that supports approx. in decision-making. It possesses many features in common fuzzy set theory. Rough Set theory was proposed as a new approach to knowledge discovery from incomplete data. The Rough set itself is the approximation of a vague concept by a pair of precise concept, called lower and upper approximations, which are classification of domain of interest into disjoint categories. The Rough set approach to processing of incomplete data is based on these approximations. The Rough Set is defined as pair of two crisp sets corresponding to approximations. If both approximations of a given subset of the universe are exactly the same then one can say that mentioned above subset is definable with respect to above information otherwise it is roughly definable.

2.1. Indiscernibility

A decision system may be unnecessarily large in part because it is redundant in two ways, the same or indiscernible objects may be represented several times, or some of the attributes may be superfluous. The former case of indiscernible objects is discussed in this section.

If I= [U, A, V, f] be an information system with any $P \subseteq A$ there is associated an equivalence relation, IND (P): IND (P) = $\{(x, y) \in U^2: \forall a \in P \ a(x) = a(y)\}$ IND (P) is called the P-**Indiscerenibility** relation. If $(x, y) \in$ IND (P), then objects x and y are indiscernible from each other by attributes from P. The equivalence classes of the P-Indiscerenibility relation are denoted $[x]_P$.

2.2. Lower and Upper approximations

Let U be a finite set of objects called Universe, $R \subseteq U*U$ be an **equivalence relation** [2, 3] on U, the pair A= (U, R) is called **approximation space**; equivalence classes of the relation R are called **Elementary sets** in A.

Let $X \subseteq U$ be a subset of U, we define lower approximation $\underline{A}(X)$ and upper approximation $\overline{A}(X)$ as:

- $\underline{A}(X) = \{x \in U: [x]_R \subset X\}$
- $\overline{A}(X) = \{x \in U: [x]_R \cap X \neq \varphi\}$

The objects in $\underline{A}$ (X) can be with certainty classified as members of X on the basis of knowledge in R, while the objects in $\overline{A}$ (X) can be only classified as possible members of X on the basis of knowledge in R.

2.3. Positive, Negative and Boundary Region

Given a subset $X \subseteq U$ representing certain concept of interest, we can characterize the approximation space A= (U, R) with three distinct regions:

a. $\overline{A}$ (X) is called **the positive region** of X in A.
b. $\overline{A}$ (X)- $\underline{A}$ (X) is called **the boundary region** of X in A.
c. U- Bnd(X) is called **the negative region**.

2.4. Non-definable Sets

Set X is definable in A iff $\overline{A}$ (X) = $\underline{A}$ (X), otherwise set X is **non-definable** in A. Four different kinds of non-definable (Rough Sets):

a. If $\underline{A}$ (X) ≠ φ and $\overline{A}$ (X) ≠ U, set X is called **roughly definable** in A, it means that we can define set x with some "approximation" i.e. define its lower and upper approximation in A.
b. If $\underline{A}$ (X) ≠ φ and $\overline{A}$ (X) = U, set X is called **externally definable** in A, it means that we are unable to exclude any element $x \in U$ being possibly a member of X.
c. If $\underline{A}$ (X) = φ and $\overline{A}$ (X) ≠ U, set X is called **internally definable** in A, it means that we are unable to say for sure that any $x \in U$ is a member of X.
d. If $\underline{A}$ (X) = φ and $\overline{A}$ (X) = U, set X is called totally **non-definable** in A it means that we are unable to define even its approximations.

3. Pre-Grouping transformation

OLAP data are divided into two main categories. The **measures** (or **facts**) are mainly numeric values, which correspond to measurements of some value related to an event at specific points in time (e.g., amount of money appearing in a line of an invoice at a particular day, or balance of an account at the end of each day, etc.) and are expected to change rapidly. The **dimension** data (or simply **dimensions**) are used to characterize the measures and are considered to be almost static in (or slowly changing with) time. The dimension values characterize a specific measure value in the same way that coordinate values characterize a specific point in a multidimensional space. Our structure will consisting of a central table (the fact table) and surrounding tables (the dimension tables) that link to it through 1: N relationships are known as the star schema. Figure 1 shows **star schema** with dimension tables.

The abstract processing operations can be optimized using various standard optimization techniques such as Eager-Group [4] by transformation. Furthermore, the existence of the h-surrogate allows for a new kind of transformation that is expected in most of the cases to perform better than the initial plan. This, so called **pre-grouping transformation** [4], allows the grouping of fact table tuples before all join operations leading to a significant reduction of both the join and grouping effort.

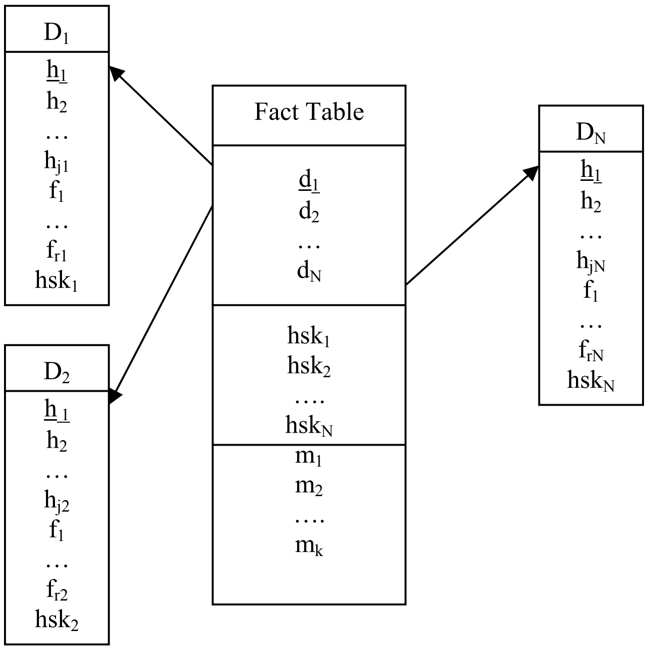


Figure 1: Fact and dimension tables

The pre-grouping transformation exploits the functional relationship among the hierarchy attributes and the corresponding h-surrogate attribute stored in the fact table. This relationship allows us to group and aggregate tuples using a prefix of the h-surrogate value instead of the values of some hierarchy attributes of a dimension. In many cases this early grouping will not correspond exactly to the grouping required by the query. This means that a second grouping operation will have to be performed after the join operations in order to obtain the final result. In those cases the pre-grouping transformation splits the initial grouping and aggregation operation into two grouping and aggregation operations.

3.1 Schema of Table

Let us now illustrate our transformation with an example query on a schema like:

SELECT L.area, P.brand, SUM (F.sales)
FROM SALES_FACT F, LOCATION L, DATE D, PRODUCT P
WHERE F.day = D.day AND F.store_id =L.store_id
AND F.product_id = P.item_id
GROUP BY L.area, D.month, P.brand

And the schema will be as shown in Figure2:

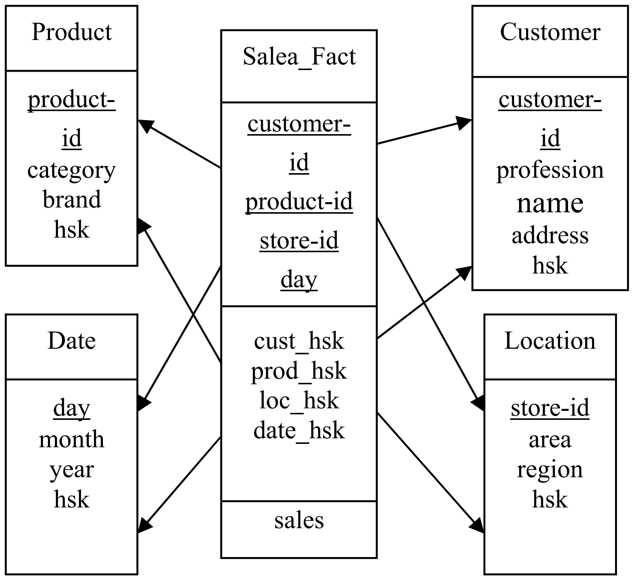


Figure 2. The data warehouse schema for our example

Let us now illustrate our transformation with an example query on the above schema. Assume we want to have a report about sales that took place in Athens containing the average sales value for each year and each profession of customers. In our schema each customer has only one profession. Using the pre-grouping transformation we can modify the plan and split grouping into two stages: one before the residual join with the **CUSTOMER** dimension and one after. In order to do the split we use the fact table's h-surrogate attribute **date_hsk**.

This attribute corresponds to the *DATE* dimension and has the structure: **year/month/day**. Using only the *year* prefix-part of **date_hsk (date_hsk: year)** we can group the fact table tuples on *year* before any residual join operation. This early grouping operation will also be done on the **cust_hsk** attribute of the fact table. This attribute corresponds to the **CUSTOMER** dimension and is required in order to obtain later the result grouped on the profession of customers. Recall that *cust_hsk* is a key attribute and so it functionally defines all attributes of the **CUSTOMER** dimension including the **profession** attribute. After the residual join operation with **CUSTOMER** we group the resulting tuples on **year** and **profession** just like the original plan. However, in this case the input to the final grouping operator will already be grouped on **year** and **customer_id**. Also, the number of tuples that will be involved in the residual joins is expected to be much smaller. Figure 3 illustrate the execution of the pre-grouping transformation [5, 6].

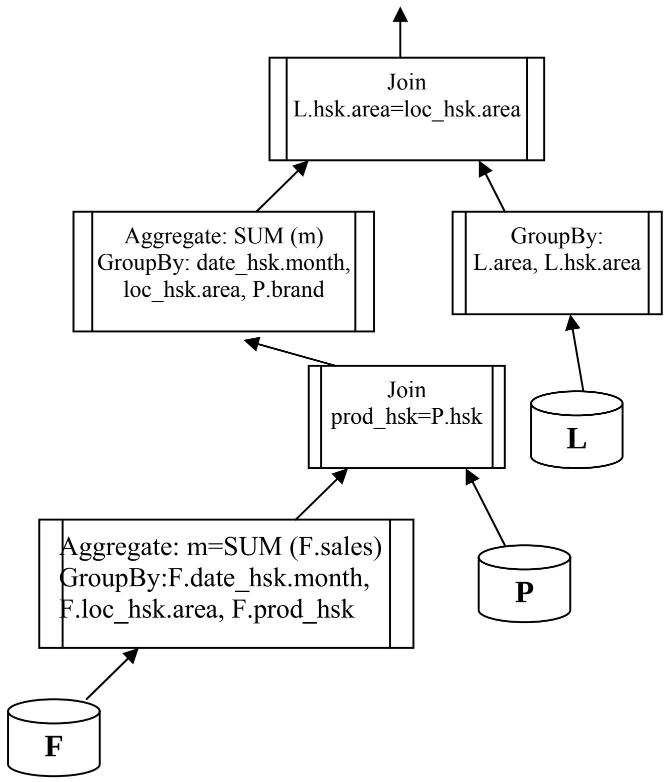


Figure3. Execution of the pre-grouping transformation

4. Our Target

The main purpose for this paper is trying to reach to the optimization case. This case is happened when the size of any database become small as possible (by applying rough set theory), and the response time for processing a query become small as possible too (by applying per-grouping transformation). Using rough set theory and reducts principle we are finding an optimal subset of attributes in a database according to some criterion, so that a classifier with the highest possible accuracy can be induced by learning algorithm using information about data available only from the subset of attributes.

5. Application and Result

We built a simple database depending on **Heart diseases** dataset using **Excel** file. This dataset has 270 numbers of examples and 2 classes, (1) for absence and (2) for presence of heart disease. The numbers of attributes are 13 attributes. The attributes information and their types will be as following:

Attribute Information:

1) age
2) sex
3) chest pain type (4 values)
4) resting blood pressure
5) serum cholesterol in mg/dl
6) fasting blood sugar > 120 mg/dl
7) resting electrocardiograph results (values 0, 1, and 2)
8) maximum heart rate achieved
9) exercise induced angina
10) old peak = ST depression induced by exercise relative to rest
11) the slope of the peak exercise ST segment
12) number of major vessels (0-3) colored by fluoroscopy
13) thal: 3 = normal; 6 = fixed defect; 7 = reversable defect

Attributes types

1) Real: 1,4,5,8,10,12
2) Ordered: 11
3) Binary: 2,6,9
4) Nominal: 7,3,13

Using **Rosetta software** which is used for analyze the data; we caught the database by the **ODBC** connector. Then the database reduced. After that we generate the rules on the best reduct. Finally we filtered the rules using **Quality filtering loop**. The result for our experiment like this in Table 2.

Threshold	Rules	AUC	SE
0.755	3.0	0.53	0.035265
0.754167	4.0	0.537833	0.035198
0.753333	9.0	0.570611	0.034805
0.752083	127.0	1.0	0.0
0.751667	258.0	1.0	0.0

Table2. The Result for the Application

If the best rule which has efficient quality had chose then applying the pre-grouping transformation.

6. Conclusion

In this paper, we have focused on the processing of the most common type of query in data warehouse, the star query. For realistic database sizes a star query may take from a couple of minutes to a few hours to execute, depending on the complexity of the query and the number of tuples retrieved from the fact table. The need for fast answers to ad hoc star queries and small database sizes as possible is a real-world problem for all contemporary business intelligence applications. Rough set approach is an important tool to deal with uncertain or vague knowledge. In this paper, the Rough set theory is deeply investigated, and a based algorithm for data filtering in Information system is proposed. The important feature of this approach is that the internal dependency structure of the system is kept intact, and that no additional parameters are needed. Theoretical analysis and experimental results show this algorithm can effectively reduce granularity of attribute measurement and obtain a higher strength of prediction in terms of the statistical significance of rules.

One of the most promising techniques for efficiently evaluating such queries is the use of fact table organizations that store data clustered according to the dimension hierarchies. A special hierarchical encoding is imposed on star joins are transformed to multidimensional range queries on the underlying multidimensional structures. The conventional star query evaluation plan changes radically and new processing steps are required.

References

[1] Z. Pawlak (1991),"Rough Sets: Theoretical Aspects of Reasoning about Data" Kluwer *Academic* Publishers, Dordrecht.

[2] Honghua Dai, "Rough Sets and Inexact Discovery", CS031 *Lecture Notes*, pp 1-13.

[3] Jan Komorowski, Zdzis law Pawlak, Lech Polkowski and Andrzej Skowron, "Rough Sets: A tutorial", pp. 2-12.

[4] N. Karayannidis, A. Tsois, T. Sellis, R. Pieringer, V. Markl, F. Ramsak, R. Fenk, K. Elhardt and R. Bayer: Processing Star Queries on Hierarchically-Clustered Fact Tables. *VLDB Conference 2002.*

[5] Nikos Karayannidis, Aris Tsois, Timos K. Sellis, Roland Pieringer, Volker Markl, Frank Ramsak, Robert Fenk, Klaus Elhardt, Rudolf Bayer: Processing Star Queries on Hierarchically-Clustered Fact Tables. *VLDB 2002: 730-741*

[6] Roland Pieringer, Klaus Elhardt, Frank Ramsak, Volker Markl, Robert Fenk, Rudolf Bayer, Nikos Karayannidis, Aris Tsois, Timos Sellis: Combining Hierarchy Encoding and Pre-Grouping: Intelligent Grouping in Star Join Processing. *ICDE 2003*

Rough Neural Intelligent Approach for Classification

Sanaa Rashed Abdallah and Yasser Fouad Hassan
Department of Mathematics (computer science),
Faculty of Science, Alex University

This paper describes rough neural network where neural network systems and rough sets theory are completely integrated into a hybrid system and are used cooperatively for decision and classification support. Also, rough sets and neural network are chosen for the combined method because they can discover patterns in ambiguous and imperfect data, and provide tools for data and pattern analysis. The common characteristic of rough sets and neural networks is that both approaches have the ability to learn decision models by examples.

1 Introduction

Intelligent systems comprise various paradigms dedicated to approximately solving real-world problems, e.g., in decision making, classification or learning; among these paradigms are fuzzy sets, rough sets, neural networks, and genetic algorithms. Fuzzy sets provide a natural framework for the process in dealing with uncertainty. It offers a problem-solving tool between the precision of classical mathematics and the inherent imprecision of the real world. The imprecision in an image contained within color value can be handled using fuzzy sets. Neural networks and rough sets are widely used for classification and rule generation. Genetic algorithms (GAs) are involved in various optimization and search processes, like query optimization and template selection. Other approaches like case based reasoning and decision trees are also widely used to solve data analysis problems. Each one of these techniques has its own properties and features including their ability of finding important rules and information that could be useful for data classification.Neural networks provide a robust approach to approximating real-valued, discrete-valued and vector-valued functions. The well-known algorithm Back-propagation, which uses gradient descent to tune network parameters to best fit the training set with input-output pair, has been applied as a learning technique for the neural networks. Rough sets based systems provide domain knowledge expressed in the form of If-then rules and tools for data analysis. Unlike other intelligent systems, rough set analysis requires no external parameters and uses only the information presented in the given data. The combination or integration of more distinct methodologies can be done in any form, either by a modular integration of two or more intelligent methodologies, which maintains the identity of each methodology, or by integrate one methodology into another, or by transforming the knowledge representation in one methodology into another form of representation, characteristic to another methodology. This paper introduces an overview of the rough neural hybrid approach for decision making.

The paper is organized as follow. Section 2, gives a brief introduction to the basic of the methods, namely rough sets and rough neural network. Section 3, describes the rough neural model. Conclusion is given in section 4.

2 Preliminary: Intelligent Techniques

Recently various intelligent techniques and approaches have been applied to handle the different challenges posed by data analysis. The main constituents of intelligent systems include fuzzy logic, neural networks, genetic algorithms, and rough sets. Each of them contributes a distinct methodology for addressing problems in its domain. This is done in a cooperative, rather than a competitive, manner. The result is a more intelligent and robust system providing a human-interpretable, low cost, approximate solution, as compared to traditional techniques. Rough set theory is a relatively new intelligent technique used in the discovery of data dependencies; it evaluates the importance of attributes, discovers the patterns of data, reduces all redundant objects and attributes, and seeks the minimum subset of attributes. Moreover, it is being used for the extraction of rules from databases.

2.1 Rough sets

Rough sets theory has been proposed by Professor Pawlak a new intelligent mathematical tool for extracting classification rules from uncertain and incomplete data-based information [7,8]. It is based on the concept of an upper and a lower approximation of a set, the approximation space and models of sets. Unlike other intelligent methods, rough set analysis requires no external parameters and uses only the information presented in the given data.

An information system (IS) is an ordered pair (U, A), where $U = \{x_1, x_2, \ldots, x_n\}$ is a nonempty finite set of objects called the universe, and $A = \{a_1, a_2, \ldots, a_n\}$is a nonempty set and the elements of A, called attributes (in our case called image features). Decision

Table 1: An example of decision table

	a_1	a_2	a_3	D
x_1	0	0	1	0
x_2	1	1	1	0
x_3	0	2	1	0
x_4	1	2	1	0
x_5	1	0	1	0
x_6	1	2	1	1
x_7	0	0	1	1

System is an information system (IS) for which the attributes in A are further classified into disjoint sets of condition attributes C and decision attributes D..

A small example of decision table can be found in Table 1. The table has seven objects, where a_1, a_2, and a_3 are conational attributes (i.e., image features in our case) and d is the decision attributes.

The discernibility matrix of A is the nxn matrix with (i,j)th entry defined as follows:

$$DM_{ij} = \{a \in A : a(x_i) \neq a(x_j)\}. \quad (1)$$

The discernibility matrix contains all the attributes that differentiate between two given objects x_1 and x_2 Every subsets of attributes of P is associated an indiscernible on on U is defined as follows:

$$I_p = \{(x, y) \in UxU : a(x) = a(y), \forall a \in P\} \quad (2)$$

Where U/I_P is the set of all equivalence classes in the relation I_P, we say that the objects x and y are $P-$indiscernible if $(x, y) \in I_p$.

In Table 1, objects x_1 and x_7 are indiscernible by attributes a_1, a_2 and a_3.

The partition constructed by attributes $\{a_1, a_2, a_3\}$ for the objects in Table 1 is: $\{\{x_1, x_7\}, \{x_2\}, \{x_3\}, \{x_4, x_6\}, \{x_5\}\}$

Due to imprecision which existed in the real world data, there are always conflicting objects contained in a decision table. Here conflicting objects refer to the two or more objects that are indiscernible by employing any set of condition attributes, but they belong

$$\underline{P}X = \{x_2, x_3, x_5\}, \quad \overline{P}X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}, \quad BN_P(X) = \{x_1, x_4, x_6, x_7\}$$

to different decision classes. Such objects are called inconsistent; therefore, the decision table is called inconsistent decision table.

Now a new partition of the universe U can be found by the indiscernibility relation. Let $IS = (U, A)$ is an information system, and let $P \subseteq A$ and $X \subseteq U$. X can be approximated using only the information contained in P by constructing B-lower and B-upper approximation of X. These basic operations in rough sets theory are defined as follows:

$$\underline{P}X = \bigcup\{Y : Y \in U/I_p, Y \subseteq X\} \tag{3}$$

$$\overline{P}X = \bigcup\{Y : Y \in U/I_p, Y \cup X \neq \phi\} \tag{4}$$

Where $\underline{P}X$ is the set of all objects of U that cab be certainly classified by set P as a members of X and $\overline{P}X$ is a set of objects that can be probably classified by P as members of X. The set

$$BN_P(X) = \overline{P}X - \underline{P}X \tag{5}$$

is referred to as the B-boundary region of X and thus consists of those objects that cannot be classified into X on the basis of knowledge X.

As a measure of quality of a partition approximation by attribute set B and a deacons attribute d. It takes the following form:

$$\gamma(B, d) = \frac{\sum_1^n (card\underline{P}X)}{card(U)} \tag{6}$$

Where $card$ denotes the cardinality. It is expresses the ratio of elements that can be properly classified employing attributes in B to all elements of the universe. If $\gamma(B, d) = 1$, it is said that d depend totally on B and if $\gamma(B, d) < 1$, it is said that d depends partially on B. An information system may contain unnecessary attributes. That means, that all conditions attributes are not needed to describe dependencies between conational and decisional attributes. For example, the reduce set of attributes in table 1 is $\{a_1, a_2\}$.
The simplest way of rule generation is to interpret each row of the reduced decision table as

a rule (i.e. the values of ci=ondition attributes imply a certain value of decision attribute. For example, the first row in table 1 can be read as follows; if $a_1 = 0$ and $a_2 = 0$ then $d = 0$

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 $\gamma(C, D) = \gamma(R, D)$. The set of all reducts of A is denoted $Red(A)$. An information may have more than one reduct.

An attribute $C_j \in C$ is a core attribute in C with respect to D if $Lower_{[C]/[D]} \neq Lower_{[C-C_j][D]}$. It is the intersection of all reducts:

$$Core(C) = \bigcap_{R_i \in Red(B)} Red_i, i = 1, 2, \ldots \tag{7}$$

It is now possible to define the significance of an attribute. This is done by calculating the change of dependency when removing the attribute from the set of considered conditional attributes.
Given P, Q and an object $x \in P$, the significant $\sigma_x(P, Q)$ of x in the equivalence relation denoted P and Q is $\sigma_x(P, Q) = \gamma(P, Q) - \gamma(P - \{x\}, Q)$.
Now, attribute reduction involves removing attributes that have no significance to the classification at hand.

2.2 Neural Networks

The effectiveness of artificial neural networks as tools that aid human decision-making in the medical field has been reported in many recent papers. Experimental results indicate that neural networks perform particularly well in solving complex pattern classification problems, due to their ability to model nonlinear relationships. Neural networks are also robust in handling data with noise or missing values, due to their inherently parallel data processing. There are a wide variety of areas in which artificial neural networks have been applied to problems in the sciences. These include pattern recognition, optimal control, adaptive filtering, inversion, target tracking, general purpose modeling, and medical diagnosis. Detailed theory and applications of are readily available in the literature [9].One of the most-used types of ANN architecture is the Feed-forward with the Back-Propagation Neural Network (FFBNN), as shown in Figure (1). The signals flow from neurons in the input layer to the neurons in the output layer, passing through the hidden neurons where there could be more than one hidden layer.

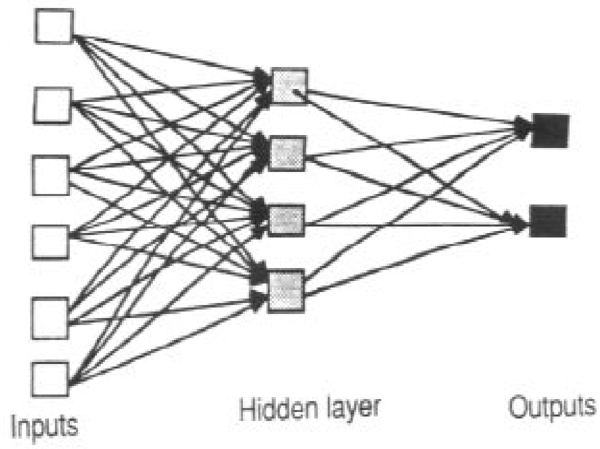


Figure 1: ANN architecture

In this subsection, we only briefly outline the main facts about networks that will be further discussed. In most of the applications presented in this work the classical multilayer feed forward network as described in [3] is utilized. The most commonly used learning algorithm is back-propagation. The signals flow from neurons in the input layer to the neurons in the output layer, passing through the hidden neurons where there could be more than one hidden layer.

By a sigmoidal excitation function for a neuron we will understand a mapping of the form:

$$f(x) = \frac{1}{1 + e^{-\beta x}} \tag{8}$$

where x represents weighted sum of inputs for a given neuron and β is the coefficient called gain, which determines the slope of the function.

Let Icn_i,Ocn_j and w_{ij} are input of the conventional neuron i, output from neuron j, and weight of the connection between neuron i and neuron j, respectively. Icn_i,Ocn_j and w_{ij} are defined as follows:

$$Icn_i = \sum_{j=1}^{n} w_{ij} Ocn_j \tag{9}$$

Neuron i is connected to neuron j.

$$Ocn_i = f(Icn_i) \tag{10}$$

Where f is the sigmoid function defined in equation (8).

3 Rough neural network: rough neuron

Rough neural networks [2,3,6,9,10]used in this study consist of one input layer, one output layer and one hidden layer. The input layer neurons accept input from the external environment. The outputs from input layer neurons are feed to the hidden layer neurons. The hidden layer neurons feed their output to the output layer neurons which send their output to the external environment.

The number of hidden neurons is determined by the following equation:

$$N_{hn} \leq \frac{N_{ts} * T_e * N_f}{N_f + N_o} \tag{11}$$

Where N_{hn} is the number of hidden neurons, N_{ts} is the number of training samples, T_e is the tolerance error, N_f is the number of attributes (features), and N_o is the number of the output. The output of a rough neuron is a pair of upper and lower bounds, while the output of a conventional neuron is a single value.

3.0.1 characteristic of rough neuron

Rough neuron was introduced in 1996 by Lingras [2] It was defined relative to upper bound (U_n), lower bound (L_n) and inputs were assessed relative to boundary values. Rough neuron has three type of connections:

Step 1. Input-Output connection to U_n
Step 2. Input-Output connection to L_n
Step 3. Connection between U_n and L_n

A rough neuron R_n is a pair of usual rough neurons $R_n = (U_n, L_n)$, where U_n and L_n are the upper rough neuron and the lower rough neuron, respectively.

Let (Ir_{L_n}, Or_{L_n}) is the input/output of the lower rough neuron and (Ir_{U_n}, Or_{U_n}) is the input/output of the upper rough neuron. The calculation of the input/output of the lower/upper rough neuron is given by the following equations:

$$Ir_{L_n} = \sum_{j=1}^{n} w_{L_{nj}} On_j \tag{12}$$

$$Ir_{U_n} = \sum_{j=1}^{n} w_{U_{nj}} On_j \tag{13}$$

$$Or_{L_n} = min(f(Ir_{L_n}), f(Ir_{U_n}))) \tag{14}$$

$$Or_{U_n} = max(f(Ir_{L_n}), f(Ir_{U_n}))) \tag{15}$$

The output of the rough neuron (O_{rn}) will be computed using the following equation:

$$O_{rn} = \frac{Or_{U_n} - Or_{L_n}}{avarge(Or_{U_n}, Or_{L_n})} \tag{16}$$

The basic stricture of rough neural network is given in Figure (2).

The rough neural network classification algorithm is described as follows Algorithm-3]:

Algorithm-1: The classification algorithm

Input: A new data to be classified, set of features (i.e., attributes), set of neurons inputs and the set of rules

Processing:

Step-1 For each attribute in the attribute set Do

Step-2 Compute the upper and lower rough neuron

Step-3 Build rough neural networks

Step-4 Compute the relative error

Step-5 Calibrate the rough neural network

Step-6 Repeat 4 and 5 until the error become minimum

Step-7 Return Class with minimum error.

Output: The final classification

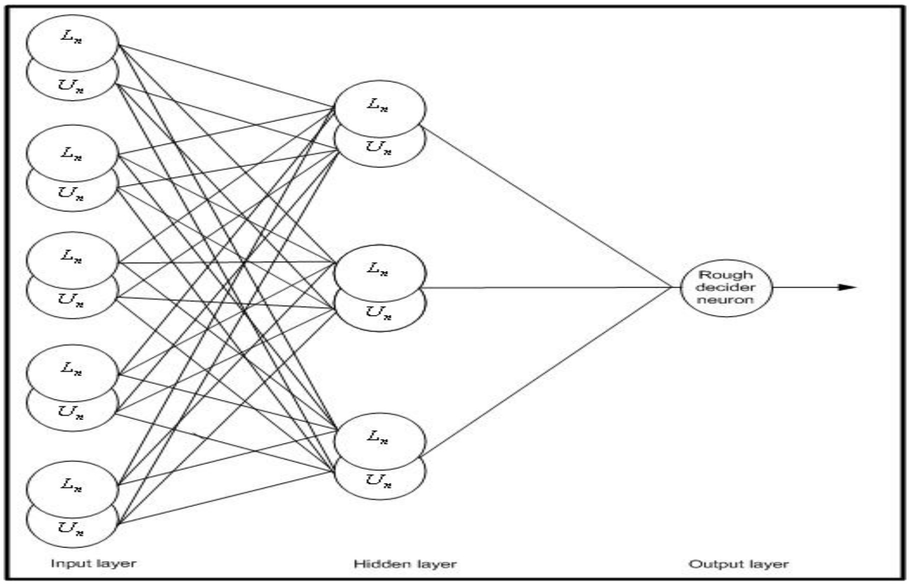


Figure 2: Rough neural network

This study proposed rough sets theory, rough neural networks for estimating rough output patterns from rough input patterns. A rough pattern uses upper and lower bounds of the values as opposed to precise values. The rough neural networks use a combination of rough and conventional neurons. A rough neuron can be viewed as a pair of neurons. One neuron corresponds to the upper bound and the other corresponds to the lower bound. Upper and lower neuron exchange information with each other during the calculation of their outputs. The paper discussed different types of connections to and from rough neurons. The errors in estimation from rough neural network models are significantly lower than the conventional neural network model. Moreover, the addition of rough neurons in hidden layer seems to improve the prediction performance.

4 Summary and Conclusion

This study proposed rough sets theory, rough neural networks for estimating rough output patterns from rough input patterns. A rough pattern uses upper and lower bounds of the values as opposed to precise values. The rough neural networks use a combination of rough and conventional neurons. A rough neuron can be viewed as a pair of neurons. One neuron corresponds to the upper bound and the other corresponds to the lower bound. Upper and lower neuron exchange information with each other during the calculation of their outputs. The paper discussed different types of connections to and from rough neurons. The errors in estimation from rough neural network models are significantly lower than the conventional neural network model. Moreover, the addition of rough neurons in hidden layer seems to improve the prediction performance.

References

[1] Grzymala-Busse J. Pawlak Z. Slowinski R. and Ziarko W. (1999) Rough Sets. In Communications of the ACM, vol.38, no. 11.pp.

[2] Lingras P. J. (1996) Rough neural networks. In: Proc. of the 6th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU96), Granada, Spain,pp.1445-1450.

[3] Aboul Ella Hassanien (2006) Rough Neural Intelligent Approach for Image Classification: A Case of Patients with Suspected Breast Cancer. International Journal of Hybrid Intelligent System, IOS press, 2006 (to appear)

[4] Ning S. Xiaohua H. Ziarko W. and Cercone N. (1994) A Generalized Rough Sets Model. In Proceedings of the 3rd Pacific Rim International Conference on Artificial Intelligence, Beijing, China, Int. Acad. Publishers, Vol. 431, pp. 437-443.

[5] Ning S. Ziarko W. Hamilton J. and Cercone N. (1995) Using Rough Sets as Tools for Knowledge Discovery. KDD'95 Proceedings First International Conference on Knowledge Discovery Data Mining, U.M. Fayyad, R. Uthurusamy (eds.), Montreal, Que., Canada, AAAI, pp. 263-268.

[6] Pal S.K. Polkowski S.K. and Skowron A. (Eds.) (2002) Rough-Neuro Computing: Techniques for Computing with Words. Berlin: Springer-Verlag.

[7] Pawlak Z. (1982) Rough Sets. Int. J. Computer and Information Sci., Vol. 11, pp. 341–356,

[8] Pawlak Z. Grzymala-Busse J. Slowinski R. and Ziarko W. (1995) Rough sets. Communications of the ACM, vol. 38, no. 11, pp. 89-95.

[9] Peters J.F. Liting H. and Ramanna S. (2001) Rough Neural Computing in Signal Analysis. Computational Intelligence vol. 17, no.3: pp. 493-513.

[10] Peters, J.F. Andrzej Skowron, Liting H. and Ramanna S. (2000) Towards Rough Neural Computing Based on Rough Membership Functions: Theory and Application. Rough Sets and Current Trends in Computing 2000, pp. 611-618

[11] Setiono R.(2000)Generating concise and accurate classification rules for breast cancer diagnosis. Artificial Intelligence in Medicine, vol. 18, no. 3, pp. 205-219.

[12] Slowinski R. (1993) Rough set approach to decision analysis. AI Expert, pp. 19-25, March 1995. 27. Stefanowski J., "Classification support based on the rough sets" Foundations of Computing and Decision Sciences, vol. 18, no. 3-4, pp. 371-380.