

USING MODULARITY WITH ROUGH DECISION MODELS

Ahmed T. Shawky¹, Hesham A. Hefny², and Ashraf H. Abd-Elwahab³

¹ Department of Computer and Information Sciences, Cairo University, Egypt

ah_taisser@yahoo.com

² Department of Computer and Information Sciences, Cairo University, Egypt

hehefny@ieee.org

³ Computer engineering, Electronics Research Institute, Cairo Egypt.

awahab@mcit.gov.eg

ABSTRACT

Many real world applications need to deal with imprecise data. Therefore, there is a need for new techniques which can manage such imprecision. Computational Intelligence (CI) techniques are the most appropriate for dealing with imprecise data to help decision makers. It is well known that soft computing techniques like genetic algorithms, neural networks, and fuzzy logic are effective in dealing with problems without explicit model and characterized by uncertainties. Using fuzzy set theory considered as major techniques, which allows decision makers to take a good decision using imprecise inexact data and knowledge. Now using rough set is getting quite necessary to be used for its ability to mining such type of data. In this research, we are looking forward to propose a novel technique, which depends on the integration between fuzzy set concepts and rough set theory in mining relational databases. The proposed model allows introducing modularity mechanism, by building a virtual modular decision tables according to variety of decision makers points of view. And introduce decision grouping mechanism for getting the optimizing decision. This approach provides flexibility in decision making verifies all decision standards and determines decision requirements, through modularizing rough decision table, extraction of rough association rules and developing mechanisms for decision grouping.

KEYWORDS

Rough sets, Fuzzy sets, modularity, Data mining.

1. INTRODUCTION

One of the newest approaches, which are suggested to improve the performance of decision making process based on relational databases, is the integration between fuzzy rules and rough set theory. Such an approach adopts rough set theory with applying fuzzy rules to use in data mining.

It's quite believed that adopting this approach in classical databases can help decision-makers and produce better decisions, especially in the case of uncertain and inexact data. The integration of Fuzzy Set theory and Rough Set theory can achieve the flexibility of manipulation of uncertainty, and the modularity techniques overcome the problem of complexity as they split the rough decision table to smaller decision tables, which simplify reduction process by decreasing the number of attributes.

Using rough set theory is a new promising technique for data mining. Unfortunately, there are serious challenges that need to be enforced for activating this technique in largest manner. They are:-

- Complexity as a result of the current huge amount of data available to build decision tables is the first problem of using rough set in data mining.
- Variety of decision makers with the differences of their points of view considered as major problem for using rough sets in decision making.

2. ROUGH SETS

The rough-set theory, proposed by Pawlak (1982, 1996), can serve as a new mathematical tool for dealing with data classification problems. In this research we assume that data are presented in the form of decision tables. So, Rough set theory is the best tool for dealing with data for helping decision makers. The decision table as shown later in this section consists of Rows and columns. Rows of the decision table represent cases, while columns represent variables. The set of independent variables are called conditional attributes and a dependent variable is called a decision attribute. In this section some concepts of rough set theory will be represented as decision table, indescribability relation, and decision rules.

A rough set theory is a new mathematical tool to deal with uncertainty and vagueness of decision system and it has been applied successfully in all the fields. It is used to identify the reduct set of the set of all attributes of the decision system. The reduct set is used as pre-processing technique for classification of the decision system in order to bring out the potential patterns or association rules or knowledge through data mining techniques. "[7]

2.1 Information Systems

A data set is represented as a table, where each row represents a case or object, patient as an example. Every column represents an attribute (a variable, an observation, a property, etc.) that can be measured for each object; the attribute may be also supplied by a human expert or user. This table is called an information system.

Table 1. Example of Decision table.

Object	Conditional Attributes			Decision
	Age	Height	Gender	Accepted
X1	Young	Tall	Male	Yes
X2	Baby	Tall	Female	Yes
X3	Young	Tall	Female	Yes
X4	Old	Medium	Female	No
X5	Baby	Short	Male	Yes
X6	Old	Medium	Male	NO

An information system (sometimes called a data table, an attribute-value system, acknowledge representation system, etc.), as a basic concept in rough set theory, provides a convenient framework for the representation of objects in terms of their attribute values. An information system S is a pair (U, A), where U is a non-empty, finite set of objects and is called the universe

International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.1, January 2012
 and A is a non-empty, finite set of attributes. V is the set of all attribute values, such as
 $V_a: U \times A \rightarrow V$ for each $x \in U$ [6]. In this example $U = \{X1, X2, X3, X4, X5, X6\}$, $A = \{\text{Age, Height, Gender, Accepted}\}$, and $V(x1, \text{Age}) = \text{Young}$.

As seen information table are divided into three sections; the universe U (objects), conditional attributes for simple attributes and decision attribute for simple decision. Equivalence relations can be defined by either conditional or decision attributes, for example:

$$R_1 = \{(x, V_a) \mid x \text{ and } V_a \text{ have the same Age}\}$$

R_1 divided into three portions,

$$Y_1 = \{X1, X3\}, \quad Y_2 = \{X2, X5\}, \quad Y_3 = \{X4, X6\}.$$

We can write that as: $R_1^* = \{Y_1, Y_2, Y_3\}$.

2.2. Indiscernibility Relation

One of the fundamental ideas of rough set theory is an Indiscernibility relation. For $B \subseteq A$ and $x, y \in U$, the Indiscernibility relation $\text{IND}(B)$ is a relation on U defined as follows:

$(x, y) \in \text{IND}(B)$ if and only if $V(x, a) = V(y, a)$ for all $a \in B$. The Indiscernibility relation $\text{IND}(B)$ is an equivalence relation. Equivalence classes of $\text{IND}(B)$ are called elementary sets and are denoted by $[x]_B$. Elementary sets may be computed by using attribute-value pair blocks.

Let $a \in A$ and let v be a value of a , for some case. For complete decision tables if $t = (a, v)$ is an attribute value pair, then a block of t , denoted $[t]$, is a set of all cases from U that for attribute a have value. [4]

For example if $t = (\text{Gender, Male})$ then $[t] = \{X1, X5, X6\}$.

2.3. Authors

For $B \subseteq A$, B is called reduct if and only if:

$$B^* = A^* \quad \text{and}$$

B is minimal this means that

$$(B - \{a\})^* \neq A^* \quad \forall a \in B$$

For example

$$A^* = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$$

$$\text{Let } B = \{\text{Age, Height, Gender}\}, \quad C = \{\text{Age, Gender}\}$$

We see that:

$$B^* = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\} = A^*$$

$$C^* = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\} = A^*$$

Also $(B^* - \{\text{Height}\}) = A^*$ this means that B is not minimal

Therefore, C is reduct of A because

$$C^* = \dots \quad \text{and} \quad C \text{ is minimal } (C - \{a\})^* \neq A^* \text{ for all } a \in C$$

2.4. Approximation space

For completely specified decision tables lower and upper approximations are defined on the basis of the Indiscernibility relation. Any finite union of elementary sets, associated with B, will be called a B-definable set. Let X be any subset of the set U of all cases. The set X is called a concept and is usually defined as the set of all cases defined by a specific value of the decision. In general, X is not a B-definable set. However, set X may be approximated by two B-definable sets; the first one is called a B-lower approximation of X, denoted by \underline{BX} and a defined as follows

$$\{x \in U \mid [x]_B \subseteq X\}$$

The second set is called a B-upper approximation of X, denoted by \overline{BX} and defined as follows

$$\{x \in U \mid [x]_B \cap X \neq \emptyset\}$$

The above shown way of computing lower and upper approximations, by constructing these approximations from singletons x will be called the first method. The B-lower approximation of X is the greatest B-definable set, contained in X. The B-upper approximation of X is the smallest B-definable set containing X."[4]

In our example $U = \{X1, X2, X3, X4, X5, X6\}$ and

$A = \{\text{Age, Height, Gender, Accepted}\}$ If $B \subseteq A$ and $B = \{\text{Height}\}$ then $B^* = \{\{X1, X2, X3\}, \{X5\}, \{X4, X6\}\}$

Suppose we have $X = \{X2, X3, X5\}$ In this case:

Lower approximation $\underline{BX} = \{X5\}$

Upper approximation $\overline{BX} = \{X1, X2, X3, X5\}$

According to using lower and upper approximations discussed above, we can distinguish three regions in approximation space:

- The positive region $\text{POS}(BX) = \underline{BX}$
- The boundary region $\text{BND}(BX) = \overline{BX} - \underline{BX}$

- The negative region $NEG(BX) = U - \overline{BX}$

2.5. Rule Induction

For the inconsistent input data, Rules induced from the lower approximation of the concept certainly describe the concept, so they are called certain. On the other hand, rules induced from the upper approximation of the concept describe the concept only possibly (or plausibly), so they are called possible [4].

For example: As a certain we can say

If (Age, old) and (Height, medium) then (accepted, no)

As a possible we can say:

If (Gender, Female) then (accepted, yes) with $\alpha = 0.67$

α is called a confidence factor and can be defined as the percentage of the number of elements that are in the elementary set and satisfy the concept for the rule from the total number of elements in the elementary set (upper approximation) in this example

$B = \{\text{Gender}\}$ then $B^* = \{\{X1, X5, X6\}, \{X2, X3, X4\}\}$

$X = \{X2, X3\}$ $P|X| = 2$

$\overline{BX} = \{X2, X3, X4\}$ $P|\overline{BX}| = 3$

$$\alpha = \frac{P|X|}{P|\overline{BX}|} \quad \alpha = \frac{2}{3} = 0.67$$

3. MODULAR ROUGH DECISION MODEL (MRDM)

Modular design has been used in various areas, e.g., robotics and neural networks [5]. Simplicity can be used for the design of each of the smaller subsystems. Since the subsystems are smaller than the original system, the design effort and computation needed in each subsystem design are typically lower. In many cases, appropriate decomposition of modules is related to variety of users.

Procedure in modular design starts with decomposing the given system into modules for simpler design, followed by aggregating the modular designs and applying to the overall system. Figure (1) represents the structure of proposed model.

A primary application of fuzzy rough sets is to reduce the number of attributes in databases thereby improving the performance of applications in a number of aspects including speed, storage, and accuracy. For a decision table with real valued conditional attributes, this can be done by reducing the number of redundant conditional attributes and find a subset of the original conditional attributes that are the most informative. [8]

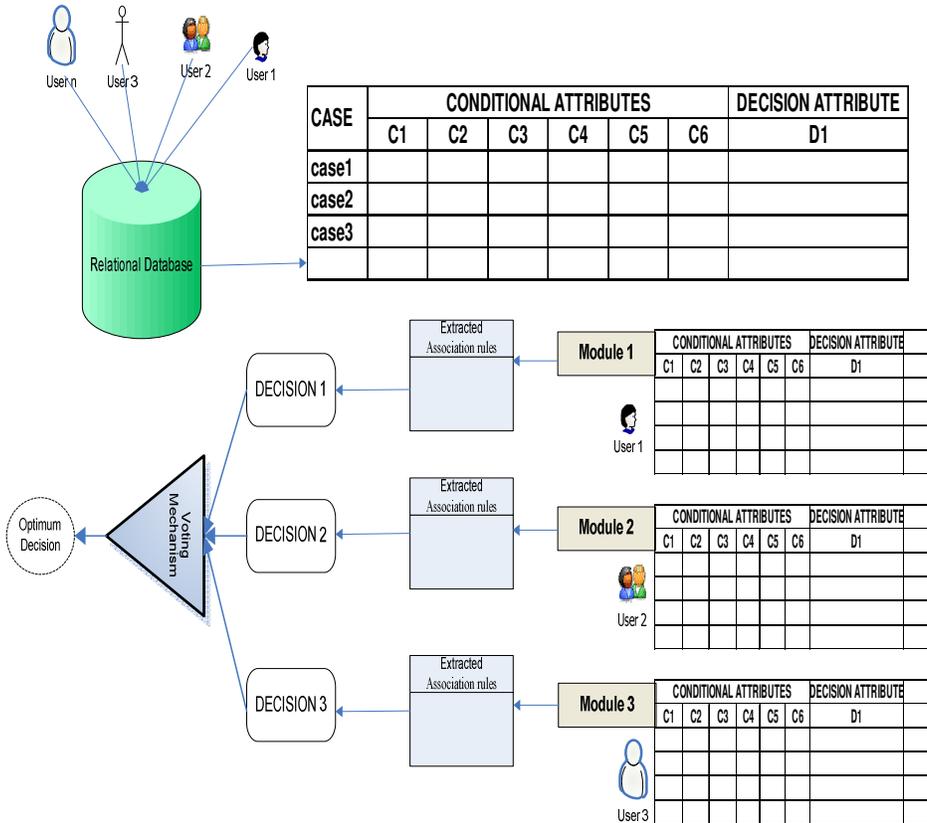


Figure 1. MRDM structure

Previous structure performed through Graphical User Interface (GUI). MRDM proposed model allows its user to build work space, we called schema, to implement Modular Rough Decision model

3.1. Data preparation

This stage concerned with using selected relational to collect the data for building the information table which, consists off column for the problem case, columns for conditions attributes and columns for decisions attributes. Information collected in excel file as in table (2) and throw MRDM proposed model the data in this file have been cached and arranged into some sort of attributes.

The data represented in rough information table in two types of attributes:

1. Conditional attributes contains data selected from the database according to the problem case, and used to take one or more decisions about this problem.
2. Decision attributes contains available decisions from the database according to transactional data.

The main objective of MRDM proposed module, is taking decision through given rough information system, by creating some modules of the main information system, then we can take decision through each module, after this step we use gating technique for taking the final decision among decisions of deferent modules. In MRDM proposed module voting is used as a gating technique for chose optimum decision.

Table 2. Excel file represents data as information system

Case	Temperatur e	Hypertension	Headache	Cough	Flue
1	39	120	Yes	Yes	Yes
2	42	180	Yes	No	Yes
3	39	130	No	No	No
4	38	200	Yes	Yes	Yes
5	37	170	Yes	No	No
6	37	180	No	Yes	No
7	40	190	Yes	No	No
8	40	200	Yes	Yes	Yes
9	38	200	Yes	Yes	Yes
10	37	170	Yes	No	No
11	37	180	No	Yes	Yes
12	37	120	No	No	No
13	42	130	Yes	Yes	Yes
14	37	220	Yes	No	No
15	41	180	Yes	No	No
16	39	130	No	Yes	Yes
17	40	200	Yes	Yes	Yes
18	38	130	No	No	No
19	42	220	Yes	Yes	Yes
20	37	120	Yes	Yes	Yes

3.2. MRDM modular approaches

In MRDM proposed model there are two approaches for rough information system modular:

1. Grid modular: Through this approach, some modules can be created by splitting the main rough information system to smaller information systems, each one considered as a module. Through MRDM proposed model, we can control the number of module created from the main information system.
2. User modular: Through this approach, any Number of modules can be created from the main rough information system, each one reflex user point of view, this modular approach depends on applying fuzzy rules on some conditional attributes which represent continuous data and can be transformed to fuzzy sets. (MRDM) proposed model allows users to create module for each by choosing the membership function for each fuzzy set according their points of view.

These approaches will be illustrated later, but now we represent how to use (MRDM) proposed model for take a decision from rough information system represented in table (2). The first step is to define inference rules which are given to take a decision, this step is one by the user of

International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.1, January 2012
(MRDM) proposed model. The other step which is done by (MRDM) proposed model is determining upper and lower approximation for all possible cases to choose the optimum decision and calculating α degree for the decision
In our example given rules is

Headache = 'Yes' and Temperature ≥ 38
The possible decisions Flue = {'Yes', 'No'}
 $B = \{\text{Headache, Temperature}\}$
 $B^* = \{\{1, 2, 4, 7, 8, 9, 13, 15, 17, 19\}, \{3, 16\}, \{5, 10, 14, 20\}, \{6, 11, 12\}\}$
 $X = \{x \mid \text{if headache} = \text{'Yes'} \text{ and temperature} \geq 38 \text{ then flue} = \text{'Yes'}\}$
 $X = \{1, 2, 4, 8, 9, 13, 17, 19\}$
 $\overline{B}X = \{1, 2, 4, 7, 8, 9, 13, 15, 17, 19\}$
 $P|\overline{B}X| = 10$
 $P|X| = 8$

$$\alpha = \frac{P|X|}{P|\overline{B}X|} \qquad \alpha = \frac{8}{10} \qquad \alpha = 0.8 \qquad (1)$$

$Y = \{x \mid \text{if headache} = \text{'Yes'} \text{ and temperature} \geq 38 \text{ then flue} = \text{'No'}\}$
 $Y = \{7, 15\}$
 $P|Y| = 2$

$$\alpha = \frac{P|Y|}{P|\overline{B}X|} \qquad \alpha = \frac{2}{10} \qquad \alpha = 0.2 \qquad (2)$$

From 1, 2 $\alpha(X) > \alpha(Y)$
The optimum decision is Flue = 'Yes' with $\alpha = 0.8$

3.2.1. Grid Modular

This approach is a modularity technique; the main objective of using is to overcome the problem of complexity as it splits the rough decision table to smaller decision tables, which simplify reduction process by decreasing the number of attributes.

This approach depends on one of three mechanisms for splitting the main information system to sub information systems (modules). These mechanisms are:

1. Serial: according to this mechanism, the main information system splits to a given number of modules, according to the order of its attributes.
2. Random: according to this mechanism, each module has some attributes, each one is randomly chosen from the main information system.
3. Formula: This mechanism looks like the above, but each attribute of the module is chosen from the main information system, as a result of implementing a given formula, instead of random chooses.

Using of MRDM proposed model to implement grid modular approach with the same example above, the first step is to determine the number of modules, the user need to create from the main information system. According this given number of modules, through MRDM proposed model, the number of attributes for each module will be determined. After that the user chooses the mechanism to split the main information system.

International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.1, January 2012
 Figure (2) represents splitting the main rough system to four modules, created through MRDM proposed model.

After creating modules from the main rough information system, MRDM proposed model allows user to define rules, which are needed to get a decision. Note that the same example used in taking decision from main rough information system represented in table (2), is used to represent using MRDM proposed model to take a decision through modularity, and also given rules are used.

Headache = 'Yes' and Temperature \geq 38

After defining rules, one decision is taken from each module with α degree. Final step in taking decision through MRDM proposed model using modularity approach is gating process. In MRDM proposed model voting technique is used as a gating process, this done by making vote between the decisions taken by the modules. The voting process is taking into account two factors which are α degree and possibility degree. Possibility degree is calculated in MRDM proposed model as percentage between the numbers of cases achieve the given rule to number of cases achieve chosen decision

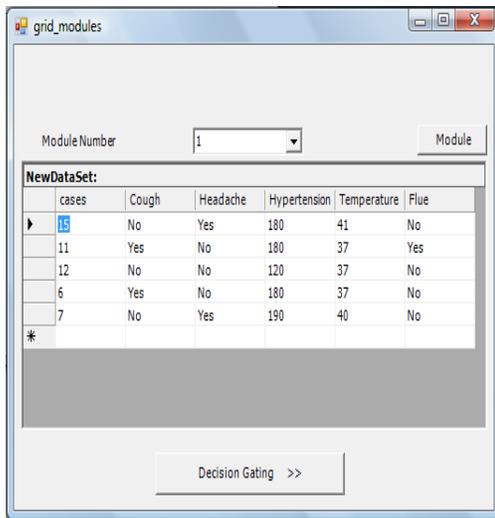


Figure 2. a. Represents Module number 1

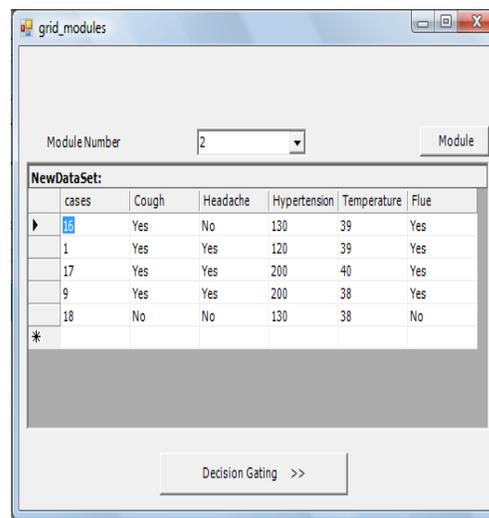


Figure 2. b. Represents Module number 2

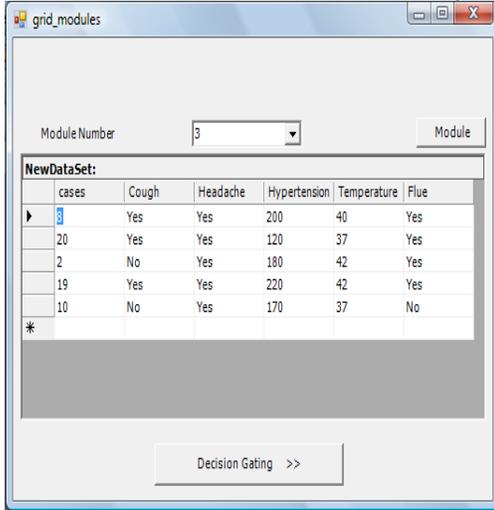


Figure. 2.c. Represents Module number 3

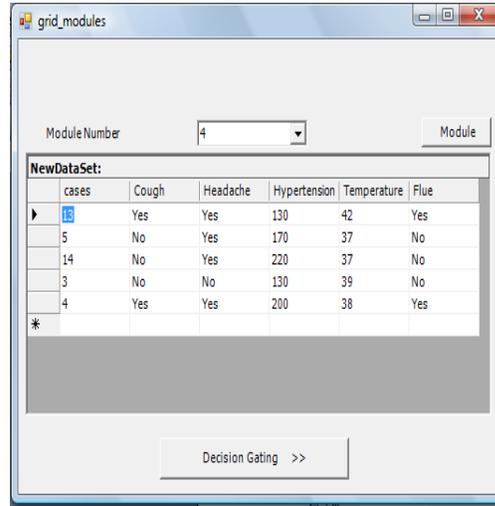


Figure. 2.d. Represents Module number 4

In our example

$X = \{x \mid \text{if headache} = \text{'Yes'} \text{ and temperature} \geq 38 \text{ then flue} = \text{'Yes'}\}$

$D = \{x \mid \text{if flue} = \text{'Yes'}\}$

$$\text{Possibility} = \frac{P \mid X \mid}{P \mid D \mid}$$

The vote of each decision is calculated as a summation of [α * possibility] for each module achieve the decision, figure (3) explains how to implement the voting process.

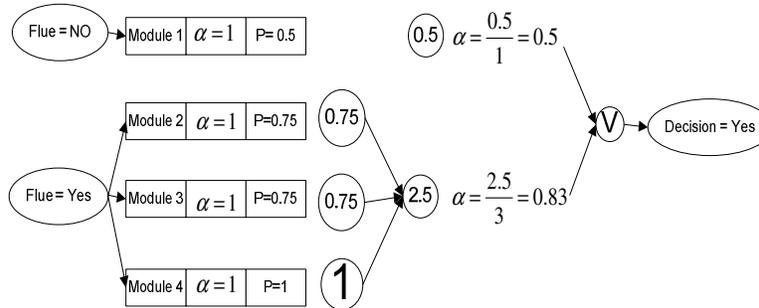


Figure 3. Voting process implementation

Advantages of grid modular that computation needed in each subsystem design are typically lower, as we need $N P_K$ tries to search in rough information system as:

$$N = K * L$$

K = number of attributes in the given rule

L = number of cases

Proof

If we have main rough information system for L cases and we need to take a decision according to rule contains K attributes

Then we have $T = N P_K$ trials

$$T = N * (N - 1) \dots (N - K-1) \quad (1)$$

If we need to split the main rough information system to M number of modules
Then we have $T = \frac{N}{M} P_K$ trials for each model

$$T = \frac{N}{M} * (\frac{N}{M} - 1) \dots (\frac{N}{M} - k-1) \quad (2)$$

If we multiply equation (2) by M

$$MF = M * \frac{N}{M} * (\frac{N}{M} - 1) \dots (\frac{N}{M} - k-1)$$

$$T = N * (N - M) \dots (N - M(k-1)) \quad (3)$$

From (1), (2) and (3)

$$MF < T$$

This means that total number trials to search in all modules smaller than number of trials to search in main rough model.

In our example

$$K = 2 \quad L = 20$$

$$N = 2 * 20 = 40$$

$$T = 40 P_2 = 40 * 39 = 1560 \text{ trails}$$

The main rough information system is split into 4 modules each module has 5 cases

$$T = 10 P_2 = 10 * 9 = 90 \text{ trails}$$

Then total number of trails = $90 * 4 = 360$ trails.

So, computations needed to take a decision through splitting the main rough information system, are typically lower than those needed to take a decision directly from the main rough information system.

3.2.2. User Modular

Applying fuzzy rules can be used to split the interval of continuous data to some fuzzy sets represented by linguistic value, the data within each information table attribute has a membership in the fuzzy set that represents this attribute with membership degree. These attributes determined using (MRDM) proposed model for allowing users to choose the membership function for each fuzzy set according their points of view. (MRDM) proposed model performs the data preparation stage using:

- Relational database
- Fuzzy data dictionary for fuzzy relational databases
- Rough data dictionary
- Stored function used with fuzzy data

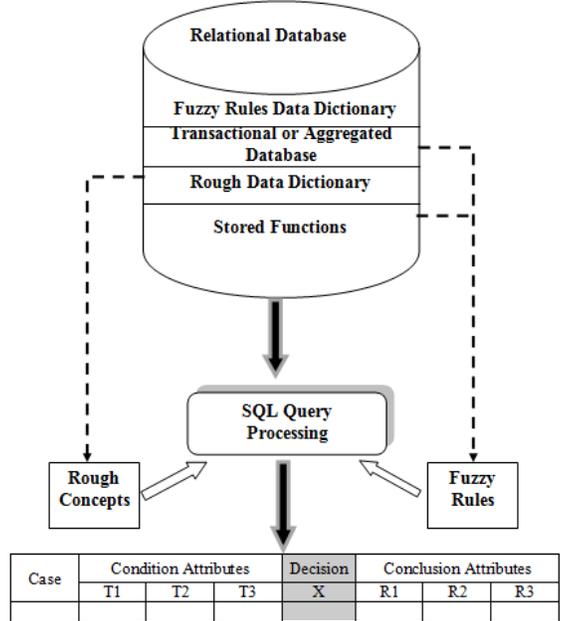


Figure 4. MRDM proposed model structure

Rough set modularity overcomes the problem of complexity, AS previous (MRDM) proposed model allows users to choose the membership function for each fuzzy set according their points of view, used for conditional attributes. So, optimization of decision is contrasted as a reason of deference between users' points of view. So, we can build more than one decision table. Figure (6) represents two examples each one is a fuzzy, but using deferent membership functions in fuzzification process for conditional attributes. Where figure (5.a) represents fuzzification process for the attribute hypertension defining fuzzy set "Height" for user "Ahmed", and figure (5.b) represents the same process for user "Ashraf". Figures rough decision table after applying fuzzification process, according to user profile which contains membership functions, for each fuzzy set for each fuzzy attribute in the main rough decision table in our case MRDM proposed model allows each user to make his profile this profile contains membership functions for each fuzzy set according to his point of view, and according to these profiles the proposed model transforms the main rough decision table to many fuzzy rough decision tables. (MRDM) proposed model allows users to design modules of the decision table, this allows to get one decision, taking in mind differences of users behaviours, instead of more than one different decision. So, MRDM proposed model used to determine different fuzzy sets to cover the interval domain for each fuzzy attribute. So, optimization of decision is contrasted as a reason of deference between users' points of view.



Figure 5. Represents fuzzy sets according to user's profile (for two deferent users)

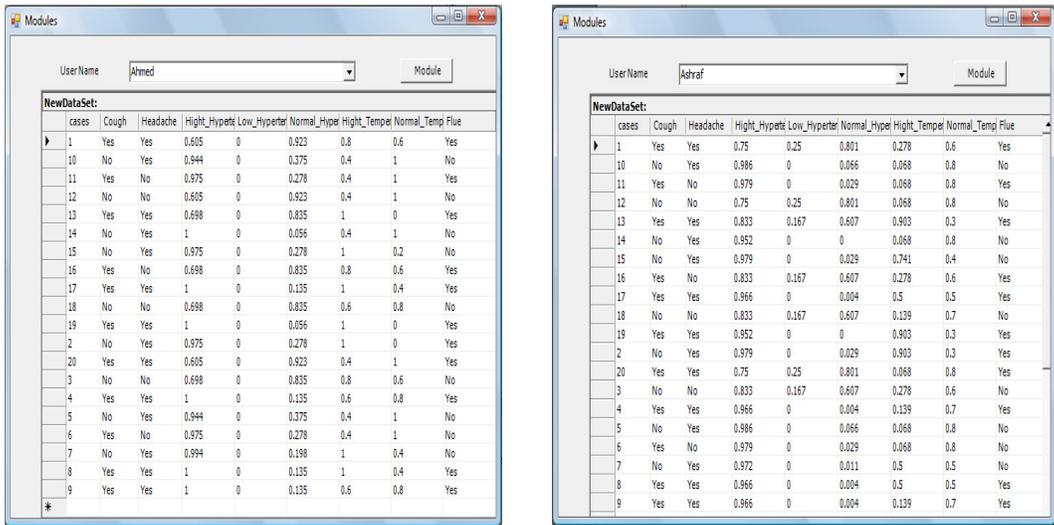


Figure 6. Represents two modules of fuzzy rough decision table according to previous two user's profiles

3.2.3. MRDM Proposed Model Reduct

Through the proposed model decision table contains both fuzzy and crisp attributes. So, in this research, two methods will be discussed for doing system reduction.

1) Method 1 Reduction using fuzzy attributes

U is a non-empty, finite set of objects and is called the universe and A is a non-empty, finite set of attributes. V is the set of all attribute values such as $v_a : U \times A \rightarrow V$ for each $x \in U$. The set of attributes A divided into two sets of attributes AF and AC, where AF is a set of fuzzy attributes and AC is a set of crisp attributes.

For any attribute $a \in A$, $t = (a, v)$ where $v \in V$, and [t] represents the set of cases in U match this condition.

International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.1, January 2012
 Let $a \in AF$, $v \in F$ with membership degree, where F is a non-empty set of fuzzy values Fuzzy sets for the attribute a . Number fuzzy values denoted by k

$$t = (a, \bigvee_{i=1}^k \mu_{F_i}(a)) \quad (1)$$

For $B \subseteq A$ and $x, y \in U$, the Indiscernibility relation $IND(B)$ denoted by $[X]_B$ is a relation on U defined as follows:

$(x, y) \in IND(B)$ if and only if $V(x, a) = V(y, a)$ for all $a \subseteq B$.

If $B = \{a, b\}$ where $a, b \in A$ then

$$[X]_B = [X]_a \otimes [X]_b \quad (2)$$

Let $a \in AF$ and $b \in AC$ then

$$[X]_a = [(a, F_i) \mid \bigvee_{i=1}^k \mu_{XF_i}(a)] \quad (3)$$

$$[X]_b = [(b, v)] \quad (4)$$

From (3), (4)

$$[X]_B = [(a, F_i) \mid \bigvee_{i=1}^k \mu_{F_i}(a)] \otimes [(b, v)] \quad (5)$$

Thus, In general for any number of fuzzy attributes $B1 \subseteq B$, $B1 \subseteq AF$ and any number of crisp attributes $B2 \subseteq B$, $B2 \subseteq AC$

$[X]_B = [X]_{B1} \otimes [X]_{B2}$ where

$$[X]_{B1} = \otimes (\forall a \in B, a \in AF \rightarrow [(a, \bigvee_{i=1}^k \mu_{F_i}(a))]) \quad (6)$$

$$[X]_{B2} = \otimes (\forall b \in B, b \in AC \rightarrow [(b, v)]) \quad (7)$$

Example (1): from last case

$U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20\}$

$A = \{\text{Headache, Cough, Temperature, Hypertension}\}$

Where $AF = \{\text{Temperature, Hypertension}\}$ and $AC = \{\text{Headache, Cough}\}$

Let $B = \{\text{Headache, Temperature}\}$

Let $X = \{V(x, flu) = \text{yes} \mid V(x, headache) = \text{Yes}\}$, $V(x, Tempretaur e) = \{\text{Hight_Temp}\}$

Then $V_{Headache} = \{\text{yes, no}\}$ and $F_i(\text{Tempretaur e}) = \{\text{Hight_Temp, Normal_Temp}\}$

From module 1 "Ahmed"

$X = \{1, 2, 8, 13, 17, 19\}$

$[X]_{Headache} = \{\{1, 2, 4, 5, 7, 8, 9, 10, 13, 14, 15, 17, 19, 20\}, \{3, 6, 11, 12, 16, 18\}\}$

$[X]_{Temperature} = \{\{1, 2, 3, 7, 8, 13, 15, 16, 17, 19\}, \{4, 5, 6, 9, 10, 11, 12, 14, 18, 20\}\}$

$[X]_B = [X]_{Headache} \otimes [X]_{Temperature}$

$[X]_B = \{\{1, 2, 7, 8, 13, 15, 16, 17, 19\}, \{4, 5, 9, 10, 14, 20\}, \{3, 16\}, \{6, 11, 12, 18\}\}$

$\underline{BX} = \{x \in U \mid [x]_B \subseteq X\}$

$\underline{BX} = \{\emptyset\}$

$\overline{BX} = \{1, 2, 7, 8, 13, 15, 16, 17, 19\}$

$$\alpha = \frac{|PX|}{|BX|} = \frac{6}{9} = 0.66$$

From module 2 "Ashraf"

$$X = \{2, 8, 13, 17, 19\}$$

$$[X]_{Headache} = \{\{1, 2, 4, 5, 7, 8, 9, 10, 13, 14, 15, 17, 19, 20\}, \{3, 6, 11, 12, 16, 18\}\}$$

$$[X]_{Temperature} = \{\{1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 16, 17, 18, 20\}, \{2, 7, 8, 13, 15, 17, 19\}\}$$

$$[X]_B = [X]_{Headache} \otimes [X]_{Temperature}$$

$$[X]_B = \{\{1, 4, 5, 7, 8, 9, 10, 14, 17, 20\}, \{2, 7, 8, 13, 15, 17, 19\}, \{3, 6, 11, 12, 16, 18\}\}$$

$$\underline{BX} = \{x \in U \mid [x]_B \subseteq X\}$$

$$\underline{BX} = \{\emptyset\}$$

$$\overline{BX} = \{1, 2, 4, 5, 7, 8, 9, 10, 13, 14, 15, 17, 19, 20\}$$

$$\alpha = \frac{|PX|}{|BX|} = \frac{5}{14} = 0.36$$

2) Method 2 Reduction using fuzzy attributes with certain threshold

As in method 1 the classification operation done using fuzzy attributes. In method 1 the value of attribute is determined according to the membership degree of its fuzzy values as in (1). But in method 2 classification operation done using fuzzy attributes according to fuzzy value but with the condition that its membership degree is more than certain threshold λ

$$t = ((a, \bigvee_{i=1}^k \mu_{F_i}(a)) \mid \mu_{F_i}(a) \geq \lambda) \quad (8)$$

Let $a \in AF$ and $b \in AC$ then

$$[X]_a = [(a, F_i) \mid \bigvee_{i=1}^k \mu_{X_{F_i}}(a), \mu_{F_i}(a) \geq \lambda] \quad (9)$$

For any number of fuzzy attributes $B1 \subseteq B$, $B1 \subseteq AF$, from (8), (9)

$$[X]_{B1} = \otimes (\forall a \in B, a \in AF \rightarrow [(a, \bigvee_{i=1}^k \mu_{F_i}(a)) \mid \mu_{F_i}(a) \geq \lambda]) \quad (10)$$

And from (7), (10) IND (B) denoted as $[X]_B$ calculated using

$$[X]_B = [X]_{B1} \otimes [X]_{B2}$$

Example (2)

Using the information system in example (1) with threshold $\lambda = 0.85$

From module 1 "Ahmed"

$$\text{Let } X = \{V(x, flu) = \text{yes} \mid V(x, headache) = \text{Yes}\} \wedge \{V(x, Temperature) = \text{Hight_Temp} \mid \lambda \geq 0.85\}$$

Then $V_{Headache} = \{\text{yes, no}\}$ and $F_i(Temperature) = \{\text{Hight_Temp } \lambda \geq 0.85, \text{Hight_Temp } \lambda < 0.85, \text{Normal_Temp}\}$

$X = \{2, 8, 13, 17, 19\}$

$[X]_{Headache} = \{\{1, 2, 4, 5, 7, 8, 9, 10, 13, 14, 15, 17, 19, 20\}, \{3, 6, 11, 12, 16, 18\}\}$

$[X]_{Temperature} = \{\{2, 7, 8, 13, 15, 17, 19\}, \{4, 5, 6, 9, 10, 11, 12, 14, 18\}, \{1, 3, 16\}\}$

$[X]_B = [X]_{Headache} \otimes [X]_{Temperature}$

$[X]_B = \{\{2, 7, 8, 13, 15, 17, 19\}, \{4, 5, 9, 10, 14\}, \{6, 11, 12, 18\}, \{1\}, \{3, 16\}\}$

$\underline{BX} = \{x \in U \mid [x]_B \subseteq X\}$

$\underline{BX} = \{\emptyset\}$

$\overline{BX} = \{2, 7, 8, 13, 15, 17, 19\}$

$\alpha = \frac{P|\underline{X}|}{|\underline{BX}|} = \frac{5}{6} = 0.83$

From module 2 "Ashraf"

$X = \{2, 13, 19\}$

$[X]_{Headache} = \{\{1, 2, 4, 5, 7, 8, 9, 10, 13, 14, 15, 17, 19, 20\}, \{3, 6, 11, 12, 16, 18\}\}$

$[X]_{Temperature} = \{\{2, 13, 19\}, \{7, 8, 15, 17\}, \{1, 3, 4, 5, 6, 9, 10, 11, 12, 14, 16, 18, 20\}\}$

$[X]_B = [X]_{Headache} \otimes [X]_{Temperature}$

$[X]_B = \{\{2, 13, 19\}, \{7, 8, 15, 17\}, \{1, 4, 5, 9, 10, 14, 20\}, \{3, 6, 11, 12, 16, 18\}\}$

$\underline{BX} = \{x \in U \mid [x]_B \subseteq X\}$

$\underline{BX} = \{2, 13, 19\}$

$\overline{BX} = \{2, 13, 19\}$

$\alpha = \frac{P|\underline{X}|}{|\underline{BX}|} = \frac{3}{3} = 1$

4. CONCLUSION

Modular approach reduces computation complexity. In many cases, appropriate decomposition of modules is related to variety of users. The idea is to ignore interconnection among subsystems in the design stage. Design effort and computation needed in each subsystem design are typically lower. The system will also be easier to debug and maintain.

5. FUTURE WORK

Adding another level of modularity in MRDM proposed model to use the first approach (Grid modular), by fuzzifying the data in modules, each information table attribute is assigned to a membership function to transform the data to the form of fuzzy set with membership degree.

REFERENCES

- [1] Zadeh L., (1965) "Fuzzy Sets", Information and Control 8:338-353.
- [2] Lefteri H. Tsoukalas and Robert E. Uhrig, (1997) "Fuzzy and Neural Approaches In Engineering ", A Wiley-Interscience Publication.
- [3] Pedrycz W. and Gomide F., (1998) "An Introduction To Fuzzy Sets Analysis and Design", Massachusetts Institute of Technology.
- [4] J. W. Grzymala-Busse, S. Siddhaye, (2004)" Rough Set Approaches to Rule Induction from Incomplete Data", the 10th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Perugia, Italy, July 4–9, 2004, vol. 2, 923–930
- [5] H. C. Tseng, B. Almogahed, (2008) "Modular neural networks with applications to pattern profiling problems", Neurocomputing, 2008.10.020.
- [6] Y. Qian , J. Liang, D. Li , H. Zhang, C. Dang, (2008)"Measures for evaluating the decision performance of a decision table in rough set theory", Information Sciences 178 (2008) 181–202.
- [7] K. Thangavel , A. Pethalakshmi, (2009) " Dimensionality reduction based on rough set theory " , Applied Soft Computing 9 (2009) 1–12
- [8] C. Degang, Z. Suyun, (2010) "Local reduction of decision system with fuzzy rough sets " , Fuzzy Sets and Systems 161 (2010) 1871–1883

Authors

Mr. Ahmed Taisser Shawky is IT spcialist in Health Insurance Organization in Cairo, and a Ph.D. student in Computer Sciences Department at the Institute of Statistical Studies and Research, Cairo University. His Ph.D. in the filed of Computational Intelligence. His research interests include Rough set theory, Fuzzy set theory , Modularity, Rough decision models



Dr. Hesham Ahmed Hefny is an Associate Professor and the head of Computer & Information Sciences Department at the Institute of Statistical Studies and research, Cairo University. His research interests include Artificial Neural Networks, Fuzzy Systems, Genetic Algorithms, Swarm Intelligence, Pattern Recognition, and Data Mining. Dr. Hesham has published over 35 peer refereed papers in academic journals and conferences on topics within Artificial Intelligence and related areas.



Prof. Ashraf Hassan Abdelwahab is a professor of computer engineering, Electronics Research Institute, Cairo, Egypt. He received his M. Sc. in 1988, Faculty of Engineering, Cairo University in the area of Artificial Intelligence, and in 1992 he received his Ph.D. in Machine Learning and Evolutionary Algorithms. He has published over 60 technical papers in National, and International journals and conferences in the areas of Evolutionary Algorithms, Machine Learning, and Data Mining.

