*Proceedings of the Third International Conference on Modeling, Simulation and Applied Optimization*
*Sharjah, U.A.E. January 20-22, 2009*

# ATTRIBUTE SELECTION FROM PARTIALLY UNCERTAIN DATA USING ROUGH SETS

*Salsabil TRABELSI*

LARODEC, Institut Supérieur de Gestion de Tunis
41 Rue de la Liberté, 2000 Bardo, Tunisie
salsabilt@yahoo.fr

*Zied ELOUEDI*

LARODEC, Institut Supérieur de Gestion de Tunis
41 Rue de la Liberté, 2000 Bardo, Tunisie
zied.elouedi@gmx.fr

**ABSTRACT**

In this paper, we deal with the problem of attribute selection from partially uncertain data based on rough sets. We handle uncertainty in decision attributes (classes) under the belief function framework. The simplification of the decision table yields to generate more significant rules in quick time.

**keywords:** Uncertainty, belief function theory, rough sets, attribute selection, classification.

## 1. INTRODUCTION

A real world database always contains a lot of attributes that are redundant and not necessary for rule discovery. If these redundant attributes are not removed, not only the time complexity of rule discovery increases, but also the quality of the discovered rules may be significantly depleted. A problem of relevant feature selection is one of the important problems in pre-processing stage of the whole modeling process studied in machine learning. There are also several attempts to solve this problem based on rough set theory [8] [19].

This theory constitutes a sound basis for data mining proposed as a tool to discover patterns hidden in data. The rough set approach offers solutions to the problem of decision rule generation, decision making and to solve the problem of attribute selection. One of the ideas was to consider as relevant features those in reduct of the information system [12] [13]. In fact, a reduct is a minimal set of attributes that preserves the ability to perform classifications as the whole attribute set does.

Another issue in real world database is the uncertainty, impression or incompleteness. This kind of data exists in many real-world applications like in medicine where symptoms or diseases of some patients may be totally or partially uncertain. Many researches have been done to adapt rough sets to this kind of environment [6] [7] [9] [15]. These extensions do not deal with partially uncertain decision attribute values in decision system. This kind of uncertainty can be represented by the theory of belief functions. It is introduced by Dempster [3] and Shafer [14]. The belief functions has been proposed for modeling someone's degrees of belief result from uncertainty. It is considered as a useful theory for representing and managing total or partial uncertain knowledge because of its relative flexibility. The belief function theory is widely applied in artificial intelligence and to real life problems for decision making and classification. In this paper, we use the Transferable Belief Model (TBM) one interpretation of belief function theory [17].

In this paper, we deal with the problem of attribute selection from partially uncertain data based on rough sets. We handle uncertainty in decision attributes (classes) under the belief function framework. The uncertainty exists in the decision attribute and not in condition attribute values of the decision system. This uncertainty is represented by the belief function theory. We will adapt the concept of the relative reduct and core in this context in order to eliminate the superfluous attributes in the objective to generate more significant decision rules from our uncertain table using rough sets. These decision rules learned from partially uncertain table constitute a rough set classifier able to predict the class of new objects. The new formulation of reduct and core in this new context is based on redefined basic concept of rough sets such as indiscernibility relation, set approximation and positive region.

This paper is organized as follows: Section 2 provides an overview about the rough set theory. Section 3 introduces the belief function theory as understood in the transferable belief model (TBM). Section 4 describes the basic concepts of rough sets under uncertainty in order to redefine the new core and reduct needed for simplification and feature selection.

## 2. ROUGH SETS

The idea of rough sets was proposed by Pawlak [10] [11] as a new mathematical tool to deal with vague concepts. The main goal of the rough set analysis is induction of approximations of concepts. Rough set theory constitutes a sound basis for knowledge discovery database (KDD) as a tool to discover patterns hidden in data. It can be used for feature selection, discretization, data reduction and decision rule generation, etc.

### 2.1. Information and decision system

Information systems are the basic vehicles for data representation in inductive learning algorithms. One can define an information system [10] in terms of a pair $A = (U, C)$, where $U$ is a non-empty, finite set of objects (cases) called $the\ universe$ and $C$ is a non-empty, finite set of $condition$ attributes, i.e., c:U $\rightarrow Vc$ for c $\in C$ ($Vc$ is called the value set of attribute c).

In supervised learning, a special case of information systems is considered, called decision systems (decision tables). A decision table is any information system of the form $A$ = (U, C $\cup$ {d}), where $d \notin C$ is a distinguished attribute called $decision$.

### 2.2. Indiscernibility relation

A decision system expresses all the knowledge about the model. This table may be unnecessarily large. The same or indiscernible objects may be represented several times. The objects $O_i$ and $O_j$ are indiscernible on a subset of attributes B $\subseteq$ C, if they have the same values for each attribute in subset B of C.

The concepts of indiscernibility relation to partition training instances according to some criteria. The equivalence classes thus partition the object set $U$ into disjoint subsets, denoted by $U/B$ or $IND_B$ is called the B-indiscernibility relation. The partition that includes $O_j$ is denoted $[O_j]_B$.

$$IND_B = U/B = \{[o_j]_B | o_j \in U)\} \quad (1)$$

Where

$$[O_j]_B = \{O_i | \forall c \in B\, c(O_i) = c(O_j)\} \quad (2)$$

The equivalence classes based on the decision attribute is denoted by $U/\{d\}$

$$IND_{\{d\}} = U/\{d\} = \{[O_j]_{\{d\}} | O_j \in U)\} \quad (3)$$

### 2.3. Set Approximation

The concepts of indiscernibility relation is a natural dimension of reducing data. Since only one element of the equivalence class is needed to represent the entire class. Subsets that are most often of interest have the same value of the outcome attribute. It may happen that a target concept cannot be defined in a crisp manner. In other words, it is not possible to induce a crisp description of such objects from table. It is here that the notion of rough sets emerges. It is possible to delineate the objects that certainly have a positive outcome, the objects that certainly do not have a positive outcome and finally the objects that belong to a boundary between the certain cases. If this boundary is non-empty, the set is rough. These notions are formally expressed as follows:

Let $A = (U, C)$ be an information system and let $B \subseteq C$ and $X \subseteq U$. We can approximate X using only the information contained by constructing the B-lower and B-upper approximations of X, denoted $\underline{B}$X and $\bar{B}$X respectively where

$$\underline{B}X = \{x \,|[x]_B \subseteq X\} \quad \text{and} \quad \bar{B}X = \{x \,|[x]_B \cap X \neq \emptyset\}$$

The objects in $\underline{B}$X can be certainty classified as members of $X$ on the basis of Knowledge in $B$, while the objects in $\bar{B}$X can be only classified as possible members of $X$ on the basis of knowledge in $B$.

### 2.4. Decision rules

The rule presentation induced from a decision table is shown as below:

$$\alpha \longrightarrow \beta \text{ with S}$$

-$\alpha$ denotes the conjunction of the conditions that a concept must satisfy.
-$\beta$ denotes a concept that the rule describes.
-S is a measure of support of which the rule holds.

A decision rule $\alpha \longrightarrow \beta$ may only reveal a part of the overall of the decision system from which it was derived. The quantity S gives a measure of how trustworthy the rule in drawing conclusion $\beta$ on the basics of evidence $\alpha$, and is a frequency based estimate of the conditional probability Pr( $\beta/\alpha$).

$$S(\alpha \longrightarrow \beta) = \frac{|\alpha - \beta|}{|\alpha|}$$

After the lower and the upper approximations have been found, the rough set theory can be then used to derive certain and possible rules from them. Rules induced from the lower approximation of the concept certainly describe it, so they are called certain.

*"if Age=16-30 and LEMS=50 then walk=yes" S=1*

On the other hand, rules induced from the upper approximation of the concept it only possibly, so they are called possible.

*"if Age=31-45 and LEMS=1-25 then walk=yes" S=1/2*

### 2.5. positive region

$POS_C(\{d\})$ is called a positive region of the partition $U/d$ with respect to $C$, is the set of all elements of $U$ that can be uniquely classified to blocks of the partition $U/\{d\}$, by means of $C$. where

$$POS_C(\{d\}) = \bigcup_{X \in U/\{d\}} \underline{C}X, \quad (4)$$

### 2.6. Dependency degree

Another important issue in data analysis is discovering dependencies between attributes. Intuitively, a set of attributes depends totally on a set of attributes $C$, denoted $C \Rightarrow D$, if all values of attributes from D are uniquely determined by values of attributes from C.

Formally, a functional dependency can be defined in the following way. Let $D$ and $C$ be subsets of attributes. $D$ depends on $C$ in a degree $k(0 \leq k \leq 1)$, denoted $C \Rightarrow_k D$, if

$$k = \gamma(C, D) = \frac{|POS_C D|}{|U|} \quad (5)$$

If k=1 we say that $D$ depends totally on $C$, and if k $\prec$ 1, we say that $D$ depends partially (in a degree k) on $C$.

The coefficient k expresses the ratio of all elements of the universe, which can be definable classified to blocks of partition $U/D$, employing attributes $C$ and will be called the $degree\ of\ dependency$.

### 2.7. Core and reducts

In the previous section, we have investigated one natural dimension of reducing data which is to identify equivalence classes. The other dimension in reduction is to keep only those attributes that preserve the indiscernibility relation and consequently set approximation. The remaining attributes are redundant since their removal does not worsen the classification. There are usually several such subsets of attributes and those which are minimal are called reducts.

In order to express the above idea more precisely, we need some auxiliary notions.

**Dispensable and Indispensable Attributes**

Let c $\in C$, attribute c is dispensable in $A$ if $Pos_C(\{d\})$ = $Pos_{C-c}(\{d\})$. Otherwise attribute c is indispensable in $A$. Where, the C-positive region of $D$: $Pos_C(\{d\})$ = $\bigcup_{X \in U/\{d\}}$ CX and $U/(\{d\})$ is the partitioning of the universe decision attribute.

Let $A = (U, C, \{d\})$, $A$ is independent if all c $\in C$ are indispensable in $A$.

**Reducts**

Given an information system $A = (U, C)$ the definitions of these notions are as follows. A reduct of $C$ is a minimal set of attributes $B \subseteq C$ such that $IND_B = IND_C$. In other words,

$$Pos_B(\{d\}) = Pos_C(\{d\}). \quad (6)$$

So, A reduct is a minimal set of attributes from $C$ that preserves the partitioning of the universe and the positive region, and hence the ability to perform classifications as the whole attributes set $C$ does. In other words, attributes that do not belong to a reduct are superfluous with regard to classification of elements of the universe.

**Remark:** Finding a minimal reduct (reduct with a minimal number of attributes) among all reducts is NP-hard. This means that computing reducts is not a trivial task. Fortunately, there exist good heuristics based on genetic algorithms that compute sufficiently many reducts in often acceptable time.

**Core**

The set of all the condition attributes indispensable in $A$ is denoted by CORE(C). It is the intersection of all reducts in $A$.

$$CORE(C) = \bigcap RED(C) \quad (7)$$

Since the core is the intersection of all reducts, it is included in every reduct. Thus, in a sense, the core is the most important subset of attributes, for none of its elements can be removed without affecting the classification power of attributes.

## 3. BELIEF FUNCTION THEORY

In this section, we briefly review the main concepts underlying the belief function theory as interpreted by the Transferable Belief Model (TBM). This theory is also appropriate to handle uncertainty in classification problems [4] [5].

### 3.1. Definitions

The TBM is a model to represent quantified belief functions [16]. Let $\Theta$ be a finite set of elementary events to a given problem, called the frame of discernment. All the subsets of $\Theta$ belong to the power set of $\Theta$, denoted by $2^\Theta$.

The impact of a piece of evidence on the different subsets of the frame of discernment $\Theta$ is represented by a basic belief assignment (bba).

The bba is a function $m : 2^\Theta \to [0, 1]$ such that:

$$\sum_{A \subseteq \Theta} m(A) = 1 \quad (8)$$

The value *m(A)*, named a basic belief mass (bbm), represents the portion of belief committed exactly to the event $A$.

Associated with $m$ is the belief function, denoted *bel*, corresponding to a specific bba *m*, assigns to every subset $A$ of $\Theta$ the sum of masses of belief committed to every subset of $A$ by $m$ [14]. Contrary to the bba which expresses only the part of belief that one commits to $A$ without being also committed to $\bar{A}$.

The belief function *bel* is defined for $A \subseteq \Theta$, $A \neq \emptyset$ as:

$$bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B) \quad (9)$$

The plausibility function *pl* quantifies the maximum amount of belief that could be given to a subset $A$ of the frame of discernment. It is equal to the sum of the bbm's relative to subsets $B$ compatible with $A$.

The plausibility function *pl* is defined as follows:

$$pl(A) = \sum_{A \cap B \neq \emptyset} m(B), \forall A \subseteq \Theta \quad (10)$$

The basic belief assignment (*m*), the belief function (*bel*) and the plausibility function (*pl*) are considered as different expressions of the same information.

### 3.2. Combination

Handling information induced from different experts (information sources) requires an evidence gathering process in order to get the fused information. In the transferable belief model, the basic belief assignments induced from distinct pieces of evidence are combined by either the conjunctive rule or the disjunctive rule of combination.

1. The conjunctive rule : When we know that both sources of information are fully reliable then the bba representing the combined evidence satisfies [18]:

$$(m_1 \bigcirc\!\!\!\!\cap\, m_2)(A) = \sum_{B,C \subseteq \Theta : B \cap C = A} m_1(B)m_2(C) \quad (11)$$

2. The disjunctive rule : When we only know that at least one of sources of information is reliable but we do not know which is reliable, then the bba representing the combined evidence satisfies [18]:

$$(m_1 \bigcirc\!\!\!\!\cup\, m_2)(A) = \sum_{B,C \subseteq \Theta : B \cup C = A} m_1(B) m_2(C) \quad (12)$$

### 3.3. Discounting

In the transferable belief model, discounting allows to take in consideration the reliability of the information source that generates the bba $m$. For $\alpha \in$ [0,1], let (1-$\alpha$) be the degree of confidence ('reliability'), we assign to the source of information. If the source is not fully reliable, the bba it generates is 'discounted' into a new less informative bba denoted $m^{\alpha}$:

$$m^{\alpha}(A) = (1-\alpha)m(A), \qquad for\, A \subset \Theta \quad (13)$$

$$m^{\alpha}(\Theta) = \alpha + (1-\alpha)m(\Theta) \quad (14)$$

### 3.4. Decision making

In the transferable belief model, holding beliefs and making decisions are distinct processes. Hence, it proposes two level models:

- *The credal level* where beliefs are represented by belief functions.
- *The pignistic level* where beliefs are used to make decisions and represented by probability functions called the pigninstic probabilities denoted $BetP$ and is defined as [17]:

$$BetP(A) = \sum_{B \subseteq \Theta} \frac{\mid A \cap B \mid}{\mid B \mid} \frac{m(B)}{(1-m(\emptyset))} \text{ for all } \; A \in \Theta \quad (15)$$

## 4. CORE AND REDUCT UNDER UNCERTAINTY

In this Section, a new approach for simplification partially uncertain decision system is proposed. We will remove the superfluous and not necessary attributes for rules discovery. We will keep only the features in reduct. It is a minimal set of attributes that preserve the ability of classification as the whole of condition attriutes does. We will redefine the concept of core and reduct induced from uncertain decision table in the objective of attribute selection.The uncertainty is represented by the TBM and exists only in decision attribute values. Hence, we will adapt the basic concepts of rough sets such as indiscernibility relation, set approximation and positive region in order to redefine new core and reduct. First, we will give an overview about this new situation : Our uncertain decision system denoted $A$ contains $n$ objects $O_j$, characterized by a set of certain condition attributes $C$={$c_1$, $c_2$,...,$c_k$} and uncertain decision attribute $ud$. We propose to represent the uncertainty of each object by a bba $m_j$ expressing belief on decision defined on the frame of discernment $\Theta$={$ud_1$, $ud_2$,...,$ud_s$} representing the possible values of $ud$.

**Example:** Let us take Table 1 to describe our uncertain decision system. This latter contains eight objects, three certain condition attributes $C$={Headache, Muscle-pain, Temperature} and an uncertain decision attribute $ud$=Flu with possible value {yes, no} representing $\Theta$.

Table 1: Uncertain decision table.

| Obj | H | M | T | Flu |
|---|---|---|---|---|
| $O_1$ | yes | yes | very high | $m_1(yes) = 1$ |
| $O_2$ | yes | no | high | $m_2(yes) = 1$ |
| $O_3$ | yes | yes | high | $m_3(yes) = 0.5 \quad m_3(\Theta) = 0.5$ |
| $O_4$ | no | yes | normal | $m_4(no) = 0.6 \quad m_4(\Theta) = 0.4$ |
| $O_5$ | no | yes | normal | $m_5(no) = 1$ |
| $O_6$ | yes | no | high | $m_6(no) = 1$ |
| $O_7$ | no | yes | very high | $m_7(yes) = 1$ |
| $O_8$ | no | no | normal | $m_8(no) = 1$ |

For the patient $O_3$, 0.5 of beliefs are exactly committed to the decision $ud_1$=yes, whereas 0.5 of beliefs is assigned to the whole of frame of discernment $\Theta$ (ignorance).

### 4.1. Indiscernibility relation

For the condition attributes, the indiscernibility relation $U/C$ is the same as in the certain case because their values are certain however, the indiscernibility relation for the decision attribute $U/\{ud\}$ is not the same as in the certain case. The decision value is represented by a bba. So, we need for optimal decision making to assign each object to the right equivalence classes. The idea is to use the pignistic transformation. It is a function which can transform the belief function to probability function in order to make decisions from beliefs. We suggest, for each object $O_j$ in the decision system $U$, compute the pignistic probability, denoted $BetP_j$, by applying the pignistic transformation to $m_j$.

For every $ud_i$, a decision value, we define:

$$X_i = \{O_j | BetP_j(ud_i) >= 0\} \quad (16)$$

$$IND_{\{ud\}} = U/\{ud\} = \{X_i | ud_i \in \Theta\} \quad (17)$$

**Example:** Let us continue with the same example, to compute the equivalence classes based on condition attributes as the same manner as the certain case as follows:
U/C= $\{\{O_1\}, \{O_2, O_6\}, \{O_3\}, \{O_4, O_5\}, \{O_7\}, \{O_8\}\}$ and to compute the equivalence classes based on uncertain decision attribute $U/\{ud\}$ as follows: Table 2 shows the pignistic probability applying to each $m_j$.

Table 2: Pignistic transformation to $m_j$ for $O_j$.

| $m_j$ | $BetP_j$ |
|---|---|
| $m_1$ | $BetP_1(yes) = 1 \; BetP_1(no) = 0$ |
| $m_2$ | $BetP_2(yes) = 1 \; BetP_2(no) = 0$ |
| $m_3$ | $BetP_3(yes) = 0.75 \; BetP_3(no) = 0.25$ |
| $m_4$ | $BetP_4(yes) = 0.2 \; BetP_4(no) = 0.8$ |
| $m_5$ | $BetP_5(yes) = 0 \; BetP_5(no) = 1$ |
| $m_6$ | $BetP_6(yes) = 0 \; BetP_6(no) = 1$ |
| $m_7$ | $BetP_7(yes) = 1 \; BetP_7(no) = 0$ |
| $m_8$ | $BetP_8(yes) = 0 \; BetP_8(no) = 1$ |

The objects $O_1$, $O_2$ and $O_7$ are assigned to the equivalence class $ud_1$=yes. The objects $O_5$, $O_6$ and $O_8$ are assigned to the equivalence class $ud_2$=no. The objects $O_3$ and $O_4$ are included in the two equivalence classes. So, the two equivalence classes $ud_1$=yes and $ud_2$=no based on uncertain decision attribute are as follows: $U/\{ud\}$=$\{\{O_1$,$O_2$,$O_3$,$O_4$,$O_7\}$, $\{O_3$,$O_4$,$O_5$,$O_6$,$O_8\}\}$.

### 4.2. Set approximation

To compute the new lower and upper approximation for our uncertain decision table, we follow two steps:

1. For each equivalence classes based on condition attributes $C$, combine their bba using the operator mean. In order to check which of them has certain bba.
2. For each equivalence classes $X_i$ based on uncertain decision attribute, we compute the new lower and upper approximation, as follows:

   $$\underline{C}X_i=\{O_j \,|\, [O_j]_C \subseteq X_i \text{ and } m_j(ud_i)=1\}$$

   In the lower approximation, we find all equivalence classes included to $X_i$ and have a certain bba.

   $$\overline{C}X_i=\{O_j \,|\, [O_j]_C \cap X_i \neq \emptyset\ \}$$

   We compute the lower as the same manner as the certain case.

**Example:** We continue with the same example to compute the new lower and upper approximation. After the first step, we obtain the combined bba for each equivalence classes $U/C$ using operator mean. Table 3 and Table 4 represent the combined bba for the subsets $\{O_4, O_5\}$ and $\{O_2, O_6\}$.

Table 3: The combined bba for the subset $\{O_2, O_6\}$.

| Patient | $m(yes)$ | $m(no)$ | $m(\Theta)$ |
|---|---|---|---|
| $O_2$ | 1 | 0 | 0 |
| $O_6$ | 0 | 1 | 0 |
| $m$ | 0.5 | 0.5 | 0 |

Table 4: The combined bba for the subset $\{O_4, O_5\}$.

| Patient | $m(yes)$ | $m(no)$ | $m(\Theta)$ |
|---|---|---|---|
| $O_4$ | 0 | 0.4 | 0.6 |
| $O_5$ | 0 | 1 | 0 |
| $m$ | 0 | 0.7 | 0.3 |

Next, we compute the lower and upper approximation for each equivalence classes $U/\{ud\}$.

For $ud_1$=yes, let $X_1$= $\{O_1$,$O_2$,$O_3$,$O_4$,$O_7\}$

The subsets $\{O_1\}$ and $\{O_7\}$ are included to $X_1$ and have a certain bba. Hence, we put them in the lower $\underline{C}$X. The subset $\{O_3\}$ is included to $X_1$. But, it has uncertain bba. So, we put it in the upper $\overline{C}X_1$. The subsets $\{O_2, O_6\}$ and $\{O_4, O_5\}$ are partially included to $X_1$. So, we let them in the upper $\overline{C}X_1$.

$\underline{C}X_1$=$\{\{O_1\}, \{O_7\}\}$ and $\overline{C}X_1$=$\{\{O_1\}, \{O_2, O_6\}, \{O_3\}, \{O_4, O_5\}, \{O_7\}\}$

For $ud_2$=no, let $X_2$= $\{O_3$,$O_4$,$O_5$,$O_6$,$O_8\}$

The subset $\{O_8\}$ is included to $X_2$ and has a certain bba. Hence, we put it in the lower $\underline{C}X_2$. The subsets $\{O_4, O_5\}$ and $\{O_3\}$ are included to $\underline{X}_2$. But, they have uncertain bba. So, we put them in the upper $\overline{C}X_2$. The subset $\{O_2, O_6\}$ is partially included to $X_2$. So, we let it in the upper $\overline{C}X_2$.

$\underline{C}X_2$=$\{\{O_8\}\}$ and $\overline{C}X_2$=$\{\{O_3\}, \{O_2, O_6\}, \{O_4, O_5\}, \{O_8\}\}$

### 4.3. Belief decision rules

The decision rules induced from our new partially uncertain decision system are denoted belief decision rules where the decision is represented by a bba.

**Example** : One of the belief decision rules induced from our decision table for the object $O_3$ is as follows :

*If Headache=yes and Muscle-pain=yes and Temperature=High Then* $m_3(yes) = 0.5 \quad m_3(\Theta) = 0.5$

Hence, these rules could be simplified by removing superfluous attributes.

### 4.4. Positive region and dependency degree

With this new lower approximations, we can define the new positive region denoted $UPos_B(\{ud\})$:

$$UPOS_C(\{ud\}) = \bigcup_{X_i \in U/ud} \underline{C}X, \tag{18}$$

So, we can compute the new dependency degree as follows:

$$\gamma(C, \{ud\}) = \frac{|POS_C(\{ud\})|}{|U|} \tag{19}$$

**Example:** Let us continue with the same example, to compute the positive region and dependency degree of $A$.

$$UPOS_C(\{\text{ud}\})=\{O_1, O_7, O_8\}$$

$$\gamma(C, \{ud\}) = \tfrac{3}{8}$$

### 4.5. Reduct and core

Using the new formalism of positive region, we can find the reduct of $C$ as a minimal set of attributes $B \subseteq C$ such that:

$$UPos_B(\{ud\}) = UPos_C(\{ud\}). \tag{20}$$

The relative core is intersection of all reducts or is the set of all indispensable attributes form $C$.

**Example:** Let us continue with the same example, to compute the relative reduct of $A$.

$UPOS_{\{Headache\}}(\{ud\})$=$\emptyset$
$UPOS_{\{Muscle-pain\}}(\{ud\})$=$\emptyset$
$UPOS_{\{Temperature\}}(\{ud\})$=$\{O_1, O_7\}$

$UPOS_{\{Headache,Muscle-pain\}}(\{ud\})$=$\{O_8\}$
$UPOS_{\{Headache,Temperature\}}(\{ud\})$=$\{O_1, O_7\}$
$UPOS_{\{Muscule-pain,Temperature\}}(\{ud\})$=$\{O_1, O_7, O_8\}$

So, {Muscle-pain, Temperature} is the only relative reduct to the decision $Flu$ in our decision system $A$. Hence, it is the core. We can compute the reduct by starting find the relative core as follows:
Remove the attribute $Headache$ from the condition attributes:
$UPOS_{\{M,T\}}(\{ud\})$=$\{O_1, O_7, O_8\}$= $Pos_C(\{ud\})$
So, the attribute $Headache$ is not indispensable.
Remove the attribute $Temperature$ from the condition attributes:
$UPOS_{\{H,M\}}(\{ud\})$=$\{O_8\} \neq Pos_C(\{ud\})$
So, the attribute $Temperature$ is indispensable.
Remove the attribute Muscle-pain from the condition attributes:
$UPOS_{\{H,T\}}(\{ud\})$=$\{O_1, O_7\} \neq Pos_C(\{ud\})$
So, the attribute $Muscle-pain$ is indispensable.
Hence, the set {Muscle-pain, Temperature} is the relative core. It is a minimal set of attributes from $C$ that preserves the positive region. So, it is the relative reduct.

Our uncertain decision system can be simplified as follows:

Table 5: Simplified uncertain decision table.

| Obj | M | T | Flu |
|---|---|---|---|
| $O_1$ | yes | very high | $m_1(yes) = 1$ |
| $O_2$ | no | high | $m_2(yes) = 1$ |
| $O_3$ | yes | high | $m_3(yes) = 0.5 \quad m_3(\Theta) = 0.5$ |
| $O_4$ | yes | normal | $m_4(no) = 0.6 \quad m_4(\Theta) = 0.4$ |
| $O_5$ | yes | normal | $m_5(no) = 1$ |
| $O_6$ | no | high | $m_6(no) = 1$ |
| $O_7$ | yes | very high | $m_7(yes) = 1$ |
| $O_8$ | no | normal | $m_8(no) = 1$ |

The belief decision rules become more simple and shorter such as :

*If Muscle-pain=yes and Temperature=High Then* $m_3(yes) = 0.5 \quad m_3(\Theta) = 0.5$.

## 5. CONCLUSION AND FUTURE WORK

In this paper, we have redefined the concept of relative core and reduct for feature selection from uncertain decision system using rough sets. We handle uncertainty in decision attributes (classes) under the belief function theory as understood by the TBM. With this simplification, the decision rules generated will be more significant, however finding optimal reduct is NP-hard problem. Hence, we suggest as a future work improve our new approach by proposing an heuristic method to finding reduct.

## 6. REFERENCES


[1] Barnett, J.A.: Calculating Dempster-Shafer plausibility. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 13(6), (1991) 599-602.

[2] Ben hariz, S., Elouedi, Z., Mellouli, K.:Clustering Approach Using Belief Function Theory, AIMSA 2006, LNAI 4183, (2006) 162171.

[3] Dempster, A.P.: A generalization of Bayesian inference. Journal of the Royal Statistical Society, Series B, Vol 30, (1968) 205-247. Canada (1993) 289-293.

[4] Denoeux, T.: A neural network classifier based on Dempster-Shafer theory. IEEE transactions on Systems, Man and Cybernetics A, Vol 30 (2), (2000) 131-150.

[5] Elouedi, Z., Mellouli, K., Smets, P.: Belief decision trees: Theoretical foundations. In International Journal of Approximate Journal IJAR 28, (2001) 91-124.

[6] Grzymala-Busse, J.: Modified algorithms LEM1 and LEM2 for rule induction from data with missing attribute values. Proc of the fifth international workshop on Rough sets and soft computing at the third joint conference on Information Sciences, (1997) 69-72.

[7] Kryszkiewiez, M.: Rough set approach to incomplete information systems, Proceedings of the second annual joint conference on Information Science, Wrightsville, NC (1995) 194-197.

[8] Modrzejewski, M.: Feature selection using rough sets theory. In Proceedings of the 11th International Conference on Machine Learning, (1993) 213-226.

[9] Orlowska, E.(ed.) Incomplete Information: Rough Set Analysis, Physical-Verlag (1997).

[10] Pawlak, Z.: Rough Sets, International Journal of Computer and Information Sciences, Vol 11, (1982) 341-356.

[11] Pawlak, Z., Zdzislaw, A.: Rough Sets: Theoretical Aspects of Reasoning About Data. Dordrecht: Kluwer Academic Publishing. ISBN 0-7923-1472-7(1991).

[12] Pawlak, Z. and Rauszer, C. M.: Dependency of attributes in Information systems. Bull. Polish Acad. Sci., Math., 33, (1985) 551-559

[13] Rauszer, C. M.: Reducts in Information systems. (1990) Fundamenta Informaticae.

[14] Shafer, G.: A mathematical theory of evidence. Princeton University Press. Princeton, NJ (1976).

[15] Slowinski, R., Stefanowski, J.: Rough classification in incomplete information systems, Mathematical and Computer Modeling, Vol 12, (1989) 1347-1357.

[16] Smets, P., Kennes, R.: The transferable belief model. Artificial Intelligence, Vol 66, (1994) 191-236.

[17] Smets, P.: The transferable belief model for quantified belief representation. In D.M. Gabbay & P. Smets (Eds), Handbook of defeasible reasoning and uncertainty management systems, Vol 1, Doordrecht, The Netherlands: Kluwer (1998) 207-301.

[18] Smets, P.: The application of belief transferable belief model to diagnostic problems. International Journal of Intelligent Systems, Vol 13, (1998) 127-157.

[19] Tanaka, H., Ishibuchi, H. and Matuda, N.: Reduction of information system based on rough sets and its application to fuzzy expert system. Advancement of Fuzzy Theory and Systems in china and Japan. (1990) International Academic Publishers.