Classification with Dynamic Reducts and Belief Functions

Salsabil Trabelsi[1], Zied Elouedi[1], and Pawan Lingras[2]

[1] Larodec, Institut Superier de Gestion de Tunis, Tunisia
[2] Saint Mary's University Halifax, Canada

Abstract. In this paper, we propose two approaches of classification namely, *Dynamic Belief Rough Set Classifier* (D-BRSC) and *Dynamic Belief Rough Set Classifier based on Generalization Distribution Table* (D-BRSC-GDT). Both the classifiers are induced from uncertain data to generate classification rules. The uncertainty appears only in decision attribute values and is handled by the Transferable Belief Model (TBM), one interpretation of the belief function theory. D-BRSC only uses the basic concepts of Rough Sets (RS). However, D-BRSC-GDT is based on GDT-RS which is a hybridization of Generalization Distribution Table (GDT) and Rough Sets (RS). The feature selection step relative to the construction of the two classifiers uses the approach of dynamic reduct which extracts more relevant and stable features. The reduction of uncertain and noisy decision table using dynamic approach generates more significant decision rules for the classification of unseen objects. To prove that, we carry experimentations on real databases according to three evaluation criteria including the classification accuracy. We also compare the results of D-BRSC and D-BRSC-GDT with those obtained from Static Belief Rough Set Classifier (S-BRSC) and Static Belief Rough Set Classifier based on Generalization Distribution Table (S-BRSC-GDT). To further evaluate our rough sets based classification systems, we compare our results with those obtained from the Belief Decision Tree (BDT).

Keywords: rough sets, generalization distribution table, belief function theory, uncertainty, dynamic reduct, classification.

1 Introduction

The idea of Rough Sets (RS) was introduced by Pawlak [13,14] to deal with imprecise and vague concepts. It has been successfully applied to many real-life problems. It constitutes a sound basis for decision support system and data mining. It is a mathematical approach to data analysis and classification problems. Learning from incomplete or uncertain data sets is usually more difficult than from complete data sets. Several extensions of induction systems based on rough sets have been developed to handle the problem of incomplete data sets which may be characterized by missing condition attribute values and not

J.F. Peters et al. (Eds.): Transactions on Rough Sets XIV, LNCS 6600, pp. 202–233, 2011.
© Springer-Verlag Berlin Heidelberg 2011

with partially uncertain decision attributes [8,9,10]. Previously, we have developed two approaches to classification based on rough sets called *Belief Rough Set Classifier* (BRSC) [27] and *Belief Rough Set Classifier based on Generalization Distribution Table* (BRSC-GDT) [28] which are able to learn decision rules from uncertain data. The uncertainty appears only in decision attributes. This type of uncertainty exists in many real-world applications such as marketing, finance, management and medicine. For example, some decision attribute values in a client's database, used by the bank to plan a loan policy, are partially uncertain. Such an uncertainty can be represented by belief functions as defined in the Transferable Belief Model (TBM), which is one interpretation of the belief function theory [18]. TBM is a useful tool for representing and managing partially or totally uncertain knowledge because of its relative flexibility [15]. The belief function theory is widely applied in machine learning and also in real life problems related to decision making and classification.

Feature selection is one of the important steps used to build BRSC and BRSC-GDT. This step is based on computing reduct, a minimal set of attributes that preserves the discernibiliy of objects from different decision classes, from uncertain data. In many cases, finding reducts from uncertain and noisy data leads to results which are unstable and sensitive to the sampling process. Dynamic reducts [2] can provide better performance in very large datasets. In fact, rules induced with dynamic reducts are more appropriate to classify new objects, since these reducts are more stable and appear more frequently in sub-decision systems created by random samples of a given decision system.

This paper takes advantage of dynamic reducts in two approaches to classification denoted as Dynamic Belief Rough Set Classifier (D-BRSC) and Dynamic Belief Rough Set Classifier based on Generalization Distribution Table (D-BRSC-GDT), which are able to learn uncertain decision rules from uncertain data. As in our previous work, the uncertainty appears only in decision attributes and is handled by the TBM. The first classification technique is based only on the basic concepts of rough sets [26]. However, the second more sophisticated technique is based on the hybrid system named GDT-RS which is a combination of Generalization Distribution Table (GDT) and Rough Sets (RS). The feature selection step used to build the new proposed classifiers is based on dynamic reducts. The dynamic reducts are redefined [25] in order to extract more stable reducts from data with an uncertain decision attribute. To prove that the uncertain decision rules induced with dynamic reducts are more appropriate to classify unseen objects than static reducts, we carry experimentations on real databases to evaluate our proposed classifiers based on three evaluation criteria: the classification accuracy, the time requirement and the size. To further validate the experimental results, we compare them with those obtained from the Static Belief Rough Set Classifier (S-BRSC) [27] and Static Belief Rough Set Classifier based on Generalization Distribution Table (S-BRSC-GDT) [28]. Finally, the quality of our classification systems based on rough sets and generalized distribution table is compared with a similar classifier called Belief Decision Tree (BDT).

This paper is organized as follows: Section 2 provides an overview of the Rough Sets (RS), Generalization Distribution Table (GDT) and the hybrid system GDT-RS. Section 3 introduces the belief function theory as understood in the TBM. In Section 4, we propose two approaches to classification namely, Dynamic Belief Rough Set Classifier (D-BSRC) and Dynamic Belief Rough Set Classifier based on Generalization Distribution Table (D-BRSC-GDT). Finally, we report the experimental results using real databases for D-BRSC, D-BRSC-GDT, S-BRSC, S-BRSC-GDT and BDT.

2 Rough Sets (RS), Generalization Distribution Table (GDT) and the Hybrid System GDT-RS

The Rough Sets (RS), proposed by Pawlak [13,14], is a mathematical tool for dealing with imperfect knowledge. It has been successfully applied to data classification problems. This theory is not competitive but complementary to other methods and can be often used jointly with other approaches [17]. Here, we introduce the basics of Rough Sets (RS), the Generalization Distribution Table (GDT) and the hybridization system GDT-RS that will be useful in this paper.

2.1 Rough Sets (RS)

Let us recall some basic notions regarding rough sets [14]. A decision table (DT) is defined as A= (U, C, {d}), where $U = \{o_1, o_2, ..., o_n\}$ is a nonempty finite set of n objects called *the universe*, $C = \{c_1, c_2, ..., c_k\}$ is a finite set of k *condition* attributes and $d \notin C$ is a distinguished attribute called *decision*. The value set of d is called $\Theta = \{d_1, d_2, ..., d_s\}$. In this paper, the notation $c_i(o_j)$ is used to represent the value of a condition attribute $c_i \in C$ for an object $o_j \in U$. Similarly, $d(o_j)$ represents the value of the decision attribute d for an object o_j. We further extend these notations for a set of attributes $B \subseteq C$, by defining $B(o_j)$ to be value tuple of attributes in B for an object o_j.

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 the subset B of C. Rough sets adopt the concept of indiscernibility relation [13,14] to partition the object set U into disjoint subsets, denoted by U/B or IND_B, and the partition that includes o_j is denoted $[o_j]_B$.

$$[o_j]_B = \{o_i | B(o_i) = B(o_j)\} \tag{1}$$

$$IND_B = U/B = \{[o_j]_B | o_j \in U\} \tag{2}$$

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

$$IND_{\{d\}} = U/\{d\} = \{[o_j]_{\{d\}} | o_j \in U\} \tag{3}$$

Set approximation. The concept of indiscernibility relation is a natural tool for 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. In such cases, the notion of rough sets is useful.

Let $B \subseteq C$ and $X \subseteq U$. We can approximate X using $B-lower$ and $B-upper$ *approximations* of X, denoted respectively as $\underline{\mathrm{B}}(\mathrm{X})$ and $\bar{\mathrm{B}}(\mathrm{X})$ where

$$\underline{\mathrm{B}}(X) = \{o_j | [o_j]_B \subseteq X\} \; and \; \bar{\mathrm{B}}(X) = \{o_j | [o_j]_B \cap X \neq \emptyset\} \tag{4}$$

Objects in $\underline{\mathrm{B}}(X)$ can be classified with certainty as members of X on the basis of knowledge in B, while objects in $\bar{\mathrm{B}}(X)$ can only be classified as possible members of X on the basis of knowledge in B.

The set $BN_B(X)$ is called the $B-boundary\ region$ of X, and thus consists of those objects that we cannot decisively classify into X on the basis of knowledge in B:

$$BN_B(X) = \bar{\mathrm{B}}(\mathrm{X}) - \underline{\mathrm{B}}(\mathrm{X}) \tag{5}$$

A set is said to be rough if the boundary region is non-empty.

Dependency of attributes and positive region. Another important issue in data analysis is discovering dependencies between attributes [14]. Intuitively, the decision attribute d depends totally on a set of condition attributes C, denoted $C \Rightarrow \{d\}$, if all values of attribute d are uniquely determined by values of attributes from C. Formally, a functional dependency can be defined in the following way. The attribute d depends on the set of attributes C in a degree $k\,(0 \leq k \leq 1)$, denoted $C \Rightarrow_k \{d\}$, if

$$k = \gamma_C(A, \{d\}) = \frac{|Pos_C(A, \{d\})|}{|U|} \tag{6}$$

Where

$$Pos_C(A, \{d\}) = \bigcup_{X \in U/\{d\}} \underline{\mathrm{C}}(X) \tag{7}$$

$Pos_C(A, \{d\})$, 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.

Reduct and core. With indiscernibility relation, 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 indisernibility and consequently set approximation. There are usually several such subsets of attributes and those which are minimal are called reducts [14]. In order to express the above idea more precisely, we need some auxiliary notions.

Let $c \in C$, the attribute c is dispensable in C with respect to d, if $Pos_C(A, \{d\}) = Pos_{C-c}(A, \{d\})$. Otherwise attribute c is indispensable in C with respect to d. If $\forall\ c \in C$ is indispensable in C with respect to d, then C is called independent.

1. **Reduct:** A subset $B \subseteq C$ is a reduct of C with respect to d, if B is independent and:
$$Pos_B(A, \{d\}) = Pos_C(A, \{d\}) \tag{8}$$
A reduct of knowledge is its essential part, which is sufficient to define all basic concepts occurring in the considered knowledge.
2. **Core:** The set of all the condition attributes indispensable in the decision system A with respect to d is denoted by $Core_C(A, \{d\})$.
$$Core_C(A, \{d\}) = \bigcap Red_C(A, \{d\}) \tag{9}$$
Where $Red_C(A, \{d\})$ is the set of all relative reducts of the decision system A.
Since the relative core is the intersection of all relative reducts, it is included in every reduct. Thus, in a sense, the core is the most important subset of attributes which cannot be removed when reducing the data.

Value reduct and value core. To further simplify a decision table, we can eliminate some values of an attribute from the table. We also need a concept of value reduct and value core.

We say that the value of attribute $c \in C$, is dispensable for o_j with respect to d, if $[o_j]_C \subseteq [o_j]_{\{d\}}$ implies $[o_j]_{C-\{c\}} \subseteq [o_j]_{\{d\}}$; otherwise the value of attribute c is indispensable for o_j with respect to d.

If for every attribute $c \in C$ the value of c is indispensable for o_j, then C is called *independent* for o_j with respect to d.

A subset $B \subseteq C$ is a value reduct of C for o_j, iff B is *independent* for o_j and $[o_j]_C \subseteq [o_j]_{\{d\}}$ implies $[o_j]_B \subseteq [o_j]_{\{d\}}$.

The set of all indispensable attributes values from C for o_j with respect to d is called value core and is denoted by $Core^j_C(A, \{d\})$.

We also have the following property:
$$Core^j_C(A, \{d\}) = \bigcap Red^j_C(A, \{d\}) \tag{10}$$
Where $Red^j_C(A, \{d\})$ is the family of all the reduct values of C for o_j with respect to d.

Dynamic reduct. If $A = (U, C \cup \{d\})$ is a decision table, then any system $A' = (U', C \cup \{d\})$ such that $U' \subseteq U$ is called a subtable of A. Let F be a family of subtables of A [2].
$$DR(A, F) = Red(A, \{d\}) \cap \bigcap_{A' \in F} Red(A', \{d\}) \tag{11}$$
Any element of $DR(A, F)$ is called an F-dynamic reduct of A. From the definition of dynamic reducts, it follows that a relative reduct of A is dynamic if it is also a reduct of all the subtables from a given family F. This notation can sometimes be too restrictive, so we apply a more general notion of dynamic reduct.

They are called (F, ε)-dynamic reducts, where $0 \leq \varepsilon \leq 1$. The set $DR_\varepsilon(A, F)$ of all (F, ε)-dynamic reducts is defined by:

$$DR_\varepsilon(A, F) = \left\{ R \in Red(A, \{d\}) : \frac{|\{A' \in F : R \in Red(A', \{d\})\}|}{|F|} \geq 1 - \varepsilon \right\} \tag{12}$$

If $R \in Red(B, \{d\})$, then the number $\frac{|\{A' \in F : R \in Red(A', \{d\})\}|}{|F|}$ is called the stability coefficient of R (relative to F).

2.2 Generalization Distribution Table (GDT)

The Generalization Distribution Table (GDT) proposed by Zhong et al. in [30] consists of three components:

1. Possible instances (PI), represented at the top row of GDT, are defined by all possible combinations of attribute values from a database.
2. Possible generalizations of instances (PG), represented by the left column of a GDT, are all possible cases of generalization for all possible instances. A wild card '*' denotes the generalization for instances.
3. Probabilistic relationships between possible instances (PI) and possible generalizations (PG), represented by entries G_{ij} 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 generalization. The prior distribution is assumed to be uniform if background knowledge is not available. Thus, it is defined by:

$$G_{ij} = p(PI_j|PG_i) = \begin{cases} \frac{1}{N_{PG_i}} & \text{if } PG_i \text{ is a generalization of } PI_j \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

where PI_j is the j-th possible instance, PG_i is the i-th possible generalization and N_{PG_i} is the number of the possible instances satisfying the i-th possible generalization. The latter is defined as follows:

$$N_{PG_i} = \prod_{k \in \{l | PG_i[l] = *\}} n_k \tag{14}$$

where $PG_i[l]$ is the value of the l-th attribute in the possible generalization PG_i and n_k is the number of values of the k-th attribute. From the probability theory, we have $\sum_j G_{ij} = 1$ for any i.

2.3 Generalization Distribution Table and Rough Set System (GDT-RS)

The GDT-RS, a soft hybrid induction system, is based on a hybridization of the *Generalization Distribution Table* (GDT) and the *Rough Set* methodology (RS). It is proposed for discovering classification rules from databases with noisy

data [3,29]. The GDT-RS system can generate a set of rules with the minimal description length, having large strength and covering all instances.

From the decision table (DT), we can generate decision rules expressed in the following form:

$$P \rightarrow Q \text{ with } S,$$

- P is a conjunction of descriptors over C,
- Q denotes a concept that the rule describes,
- S is a 'measure of the strength' of the rule.

According to the GDT-RS, the strength S of a given rule reflects incompleteness and noise and is equal to [3,29]:

$$S(P \rightarrow Q) = s(P) * (1 - r(P \rightarrow Q)) \tag{15}$$

where $s(P)$ is the strength of the generalization P and r is the noise rate function. On the assumption that the prior distribution is uniform, the strength of the generalization $P = PG$ is given by:

$$s(P) = \sum_l p(PI_l|P) = \frac{1}{N_P}|[P]_{DT}| \tag{16}$$

where $[P]_{DT}$ is the set of all the objects in DT satisfying the generalization P and N_P is the number of possible instances satisfying the generalization P which is computed using eqn. (14). The strength of the generalization P represents its ability of prediction for unseen instances.

On the other hand, the noise rate is given by:

$$r(P \rightarrow Q) = 1 - \frac{|[P]_{DT} \cap [Q]_{DT}|}{|[P]_{DT}|} \tag{17}$$

It shows the quality of classification measured by the number of the instances satisfying the generalization P which cannot be classified into class Q.

3 Belief Function Theory

In this Section, we briefly review the main concepts underlying the belief function theory as interpreted in the TBM [18,19]. The latter is a useful model to represent quantified belief functions.

3.1 Definitions

Let Θ be a finite set of elementary events to a given problem, called the frame of discernment. All the subsets of Θ belong to the power set of Θ, denoted by 2^{Θ}.

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

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

$$\sum_{E \subseteq \Theta} m(E) = 1 \tag{18}$$

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

Associated with m is a belief function, denoted *bel*, which assigns the sum of masses of belief committed to every subset of $E \in 2^{\Theta}$ by m [15].

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

$$bel(E) = \sum_{\emptyset \neq F \subseteq E} m(F) \tag{19}$$

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

The plausibility function *pl* is defined as follows:

$$pl(E) = \sum_{E \cap F \neq \emptyset} m(F), \forall E \subseteq \Theta \tag{20}$$

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 TBM, 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 [20]:

$$(m_1 \textcircled{\cap} m_2)(E) = \sum_{F,G \subseteq \Theta : F \cap G = E} m_1(F) m_2(G) \tag{21}$$

2. The disjunctive rule: When we only know that at least one of the sources of information is reliable but we do not know which is reliable, then the bba representing the combined evidence satisfies [20]:

$$(m_1 \textcircled{\cup} m_2)(E) = \sum_{F,G \subseteq \Theta : F \cup G = E} m_1(F) m_2(G) \tag{22}$$

3.3 Discounting

In the TBM, discounting takes into consideration the reliability of the information source that generates the bba m. For $\alpha \in$ [0,1], let (1-α) be the degree of 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^{α}:

$$m^{\alpha}(E) = (1 - \alpha) m(E), \qquad for\, E \subset \Theta \tag{23}$$

$$m^{\alpha}(\Theta) = \alpha + (1 - \alpha) m(\Theta) \tag{24}$$

3.4 Decision Making

In the TBM, holding beliefs and making decisions are distinct processes. Hence, it proposes two levels:

- *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 pignistic probabilities denoted $BetP$ and defined as follows [19]:

$$BetP(\{a\}) = \sum_{F \subseteq \Theta} \frac{\mid \{a\} \cap F \mid}{\mid F \mid} \frac{m(F)}{(1 - m(\emptyset))}, \text{ for all } \ a \in \Theta \tag{25}$$

4 Dynamic Belief Rough Set Classifier (D-BRSC) and Dynamic Belief Rough Set Classifier Based on Generalization Distribution Table (D-BRSC-GDT)

In this Section, we propose two approaches of classification called Dynamic Belief Rough Set Classifier (D-BRSC) and Dynamic Belief Rough Set Classifier based on Generalization Distribution Table (D-BRSC-GDT) which are based on dynamic approach of feature selection. These classifiers are built from uncertain data under the belief function framework. The uncertainty appears only in decision attributes and is handled by the TBM. Before describing the main steps of the construction procedure of D-BRSC and D-BRSC-GDT especially the feature selection, we need to first describe the modified basic concepts of rough sets under uncertainty [23] such as decision table, indiscernibility relation (tolerance relation), set approximation, positive region, reduct and core.

4.1 Basic Concepts of Rough Sets under Uncertainty

In order to extend the rough sets under the belief function framework, we need to redefine the basic concepts of rough sets in this new context. This subsection describes the modified definitions of decision table, indiscernibility relation (tolerance relation), set approximations, positive region and dependency of attributes. These adaptations were originally proposed in [23].

Uncertain Decision Table (UDT). Our UDT is given by $A = (U, C \cup \{ud\})$, where $U = \{o_j : 1 \leq j \leq n\}$ is characterized by a set of certain condition attributes C=$\{c_1, c_2,...,c_k\}$, and an uncertain decision attribute ud. The value set of ud, is called Θ =$\{ud_1, ud_2,...,ud_s\}$. We represent the uncertainty of each object o_j by a bba m_j expressing beliefs on decisions defined on the power set of Θ. These bba's are generally given by an expert (or several experts) to partial uncertainty, they can also present the two extreme cases of total knowledge and total ignorance.

Table 1. Uncertain decision table

U	Age	Property	Income	Unpaid -credit	Loan
o_1	Young	Middle	Average	No	$m_1(\{Yes\}) = 0.95 \quad m_1(\{No\}) = 0.05$
o_2	Old	Middle	High	No	$m_2(\{Yes\}) = 1$
o_3	Old	Middle	Average	No	$m_3(\{Yes\}) = 0.7 \quad m_3(\Theta) = 0.3$
o_4	Young	Less	High	Yes	$m_4(\{No\}) = 0.9 \quad m_4(\Theta) = 0.1$
o_5	Young	Less	High	Yes	$m_5(\{No\}) = 1$
o_6	Young	Middle	Average	No	$m_6(\{No\}) = 0.9 \quad m_6(\Theta) = 0.1$
o_7	Young	Middle	Low	No	$m_7(\{No\}) = 1$
o_8	Old	Less	High	No	$m_8(\{Yes\}) = 1$
o_9	Young	Greater	Average	No	$m_9(\{Yes\}) = 1$

Example: *Let us use Table 1 to describe our UDT. It contains nine objects (clients), four certain condition attributes* C=$\{Age, Property, Income, Unpaid-credit\}$ *and an uncertain decision attribute ud with two possible values* $\{Yes, No\}$ *representing* Θ*. The decision* Yes *means that a bank accepts to give the loan to the client and the decision* No *means that the bank refuses to give the loan. This table is useful for a bank to plan a loan policy. For example, for the object* o_3*, belief of 0.7 is exactly committed to the decision* Yes*, whereas belief of* 0.3 *is assigned to the whole of frame of discernment* Θ *(ignorance). Through a bba, we can also represent the certain case, like for the objects* o_2*,* o_5 *and* o_7*.*

Tolerance relation. The indiscernibility relation for the condition attributes 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. For this reason, the indiscernibility relation will be denoted by *tolerance relation*.

To redefine $U/\{ud\}$ in the new context, we need to assign each object o_j to the right tolerance class X_i representing the decision value ud_i. For each object o_j, we should check the distance between its bba m_j and the decision value ud_i. The idea is to use a distance measure between the bba m_j and a certain bba m (such that $m(\{ud_i\}) = 1$).

Many distance measures between two bba's were developed [1,6,7,11,21,31] which can be characterized into two kinds:

- Distance measures based on pignistic transformation [1,6,21,31]: For these distances, one unavoidable step is the pignistic transformation of the bba's. Since, there is no bijection between bba's and pignistic probabilities (transformations from the power set to the set). This kind of distance may lose information given by the initial bba's. Besides, we can obtain the same pignistic transformation on two different bba distributions. So, the distance between the two obtained results does not reflect the actual similarity between the starting bba distributions.

– Distance measures directly defined on bba's and on the power set Θ were developed in [7,11]. The first one developed by Fixen and Mahler [7] is a pseudo-metric, since the condition of non-degeneracy is not respected. In our case, we choose the distance measure proposed in [11] which satisfies more properties such as:

1. Non-negativity: dist(m_1,m_2)$\geq$0
2. Non-degeneracy: dist(m_1,m_2) = 0 $\Longleftrightarrow$ $m_1 = m_2$
3. Triangle inequality: dist(m_1,m_2) $\leq$ dist(m_1,m_3) + dist(m_3,m_2)
4. Symmetry: dist(m_1,m_2) = dist(m_2,m_1)

For every ud_i, we define a tolerance class as follows:

$$X_i = \{o_j | dist(m, m_j) < 1 - threshold\},\ such\, that\, m(\{ud_i\}) = 1 \tag{26}$$

Besides, we define a tolerance relation as follows:

$$IND_{\{ud\}} = U/\{ud\} = \{X_i | ud_i \in \Theta\} \tag{27}$$

Where *dist* is a distance measure between two bba's [11]:

$$dist(m_1, m_2) = \sqrt{\frac{1}{2}(\parallel m_1^{\rightarrow} \parallel^2 + \parallel m_2^{\rightarrow} \parallel^2 - 2 < m_1^{\rightarrow}, m_2^{\rightarrow} >)} \tag{28}$$

$$0 \leq dist(m_1, m_2) \leq 1 \tag{29}$$

Where $< m_1^{\rightarrow}, m_2^{\rightarrow} >$ is the scalar product defined by:

$$< m_1^{\rightarrow}, m_2^{\rightarrow} >= \sum_{i=1}^{|2^{\Theta}|} \sum_{j=1}^{|2^{\Theta}|} m_1(A_i) m_2(A_j) \frac{|A_i \cap A_j|}{|A_i \cup A_j|} \tag{30}$$

with A_i, $A_j \in 2^{\Theta}$ for $i, j = 1, 2, \cdots, |2^{\Theta}|$. $\parallel m_1^{\rightarrow} \parallel^2$ is then the square norm of $m_1^{\rightarrow}$.

It should be noted here that we have replaced the term equivalence class from the certain decision attribute case by tolerance class for the uncertain decision attribute, because the resulting classes may overlap.

Example: *Let us continue with the same example to compute the equivalence classes based on condition attributes in the same manner as in the certain case:* U/C= $\{\{o_1, o_6\}, \{o_2\}, \{o_3\}, \{o_4, o_5\}, \{o_7\}, \{o_8\}, \{o_9\}\ \}$ *and to compute the tolerance classes based on the uncertain decision attribute* $U/\{ud\}$. *If the user fixed the value of the threshold at* 0.1, *we would obtain the following results.*

For the uncertain decision value Yes *and* $m(\{Yes\}) = 1$*:*

$dist(m, m_1) = 0.07\ < 0.9$
$dist(m, m_2) = 0.00\ < 0.9$

$dist(m, m_3) = 0.34 < 0.9$
$dist(m, m_4) = 0.93 > 0.9$
$dist(m, m_5) = 1.00 > 0.9$
$dist(m, m_6) = 0.93 > 0.9$
$dist(m, m_7) = 1.00 > 0.9$
$dist(m, m_8) = 0.00 < 0.9$
$dist(m, m_9) = 0.00 < 0.9$
$So,\ X_1 = \{o_1,\ o_2,\ o_3,\ o_8,\ o_9\}$

For the uncertain decision value No and $m(\{No\}) = 1$*:*

$dist(m, m_1) = 0.92 > 0.9$
$dist(m, m_2) = 1.00 > 0.9$
$dist(m, m_3) = 0.66 < 0.9$
$dist(m, m_4) = 0.06 < 0.9$
$dist(m, m_5) = 0.00 < 0.9$
$dist(m, m_6) = 0.06 < 0.9$
$dist(m, m_7) = 0.00 < 0.9$
$dist(m, m_8) = 1.00 > 0.9$
$dist(m, m_9) = 1.00 > 0.9$
$So,\ X_2 = \{\ o_3,\ o_4,\ o_5,\ o_6,\ o_7\}$

$U/\{ud\} = \{\{o_1,\ o_2,\ o_3,\ o_8,\ o_9\},\ \{o_3,\ o_4,\ o_5,\ o_6,\ o_7\ \}\}$

Set approximation. Like the tolerance relation $U/\{ud\}$, the set approximation concept should be redefined in the new uncertain context. Hence, to compute the new *lower* and *upper* approximations for our UDT, we should follow two steps:

1. We combine the bba's for each equivalence class from U/C using the operator mean [12]. The latter is more suitable than the rule of combination in eqn. (21) which is proposed especially to combine different beliefs on decision for one object and not different bba's relative to different objects. The combination using operator mean [12] can reduce the conflict between pieces of evidence. The combined bba's for the equivalence class $[o_j]_C$ containing the object o_j is computed as follows:

$$\bar{\mathrm{m}}_{[o_j]_C}(E) = \frac{1}{|[o_j]_C|} \sum_{o_i \in [o_j]_C} m_i(E),\ \ for\, all\, E \subseteq \Theta \tag{31}$$

2. We compute the new *lower* and *upper* approximations for each tolerance class X_i from $U/\{ud\}$ based on uncertain decision attribute ud_i as follows:

$$\underline{\mathrm{C}}(X_i) = \{o_j | [o_j]_C \subseteq X_i and\ dist(m, \bar{\mathrm{m}}_{[o_j]_C}) \leq threshold\} \tag{32}$$

such that $m(\{ud_i\}) = 1$

$$\bar{\mathrm{C}}(X_i) = \{o_j | [o_j]_C \cap X_i \neq \emptyset\} \tag{33}$$

In the new *lower* approximation, all equivalence classes from U/C are included in X_i where the distance between the combined bba $\bar{\mathrm{m}}_{[o_j]_C}$ and the certain bba m (such that $m(\{ud_i\}) = 1$) is less than a *threshold*. However, the *upper* approximation is computed in the same manner as in the certain case.

As in conventional rough set theory, the boundary region of X_i is defined as:

$$BN_C(X_i) = \bar{\mathrm{C}}(X_i) - \underline{\mathrm{C}}(X_i) \tag{34}$$

Note that in the case of uncertainty the *threshold value* gives more flexibility to the tolerance relation and the set approximations.

Example: *We continue with the same example to compute the new lower and upper approximations. After the first step, we obtain the combined bba for each equivalence class from U/C using the operator mean. Table 2 represents the combined bba for the equivalence classes $[o_1]_C$ and $[o_4]_C$. Note that to simplify the notation, we have used $m_{1,6}$ to mean $\bar{m}_{[o_1]_C}$.*

Table 2. The combined bba for the subsets $\{o_1, o_6\}$ and $\{o_4, o_5\}$

Object	$m(\{Yes\})$	$m(\{No\})$	$m(\Theta)$
o_1	0.95	0.05	0
o_6	0	0.9	0.1
$m_{1,6}$	0.47	0.47	0.05
o_4	0	0.9	0.1
o_5	0	1	0
$m_{4,5}$	0	0.95	0.05

Next, we compute the *lower* and *upper* approximations for each tolerance class $U/\{ud\}$. We will use $threshold = 0.1$.

For the uncertain decision value Yes, let $X_1 = \{o_1, o_2, o_3, o_8, o_9\}$. The subsets $\{o_2\}$, $\{o_3\}$, $\{o_8\}$ and $\{o_9\}$ are included in X_1. We should check the distance between their bba and the certain bba m (such that $m(\{Yes\}) = 1$).

$dist(m, m_2) = 0.00 < 0.1$
$dist(m, m_3) = 0.34 > 0.1$
$dist(m, m_8) = 0.00 < 0.1$
$dist(m, m_9) = 0.00 < 0.1$

$\underline{\mathrm{C}}(X_1)=\{o_2, o_8, o_9\}$, $\bar{\mathrm{C}}(X_1)=\{o_1, o_2, o_3, o_6, o_8, o_9\}$ and $BN_C(X_1)=\{o_1, o_3, o_6\}$

For uncertain decision value No, let $X_2 = \{o_3, o_4, o_5, o_6, o_7\}$. The subsets $\{o_3\}$, $\{o_4, o_5\}$ and $\{o_7\}$ are included in X_2. We should check the distance between their bba and the certain bba m (such that $m(\{No\}) = 1$).

$dist(m, m_3) \quad = 0.66 > 0.1$
$dist(m, m_{4,5}) = 0.06 < 0.1$
$dist(m, m_7) \quad = 0.00 < 0.1$

$\underline{C}(X_2)=\{o_4, o_5, o_7\}$, $\bar{C}(X_2)=\{o_1, o_3, o_4, o_5, o_6, o_7\}$ and $BN_C(X_2)=\{o_1, o_3, o_6\}$

Dependency of attributes and positive region. With this new *lower* approximation, we can define the new positive region denoted $UPos_C(\{ud\})$:

$$UPos_C(A, \{ud\}) = \bigcup_{X_i \in U/\{ud\}} \underline{C}(X_i) \tag{35}$$

The attribute ud depends on the set of attributes C to a degree k, if

$$k = \gamma_C(A, \{ud\}) = \frac{|UPos_C(A, \{ud\})|}{|U|} \tag{36}$$

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

$$UPos_C(A, \{ud\})=\{o_2, o_4, o_5, o_7, o_8, o_9\}, \gamma_C(A, \{ud\}) = \tfrac{6}{9}$$

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:

$$UPos_B(A, \{ud\}) = UPos_C(A, \{ud\}) \tag{37}$$

The core is the most important subset of attributes. It is the intersection of all reducts.

$$UCore_C(A, \{ud\}) = \bigcap URed_C(A, \{ud\}) \tag{38}$$

Where $URed_C(A, \{ud)$) is the set of all reducts of A relative to ud.

Example: *Let us continue with the same example in Table 1 to compute the possible reducts using the new formalism of positive region. We find that only the subset* $\{Property, Income, Unpaid-credit\}$ *has the same positive region as the whole set of condition attributes.*

$UPos_{\{Property,Income,Unpaid-credit\}}(A, \{ud\})= UPos_C(A, \{ud\})$

There is only one reduct: $B=\{Property, Income, Unpaid-credit\}$.

Belief decision rules. The decision rules induced from the uncertain decision table are called belief decision rules where the decision is represented by a bba.

Example: *The belief decision rule relative the object* o_3 *from the Table 1 is as follows: If* $Age = Young$ *and* $Property = Middle$ *and* $Income = Average$ *and* $Unpaid-credit = No$ *Then* $m_3(\{Yes\}) = 0.7 \quad m_3(\Theta) = 0.3$.

4.2 Dynamic Belief Rough Set Classifier (D-BRSC)

Our first technique of classification under the belief function framework called the Dynamic Belief Rough Set Classifier (D-BRSC) was originally proposed in [26]. This technique is based only on the basic concepts of Rough Sets (RS). The construction procedure of the D-BRSC is described by means of the following steps:

Step 1. Feature selection based on dynamic reduct: It consists of removing the superfluous condition attributes that are not in reduct. This leaves us with a minimal set of attributes that preserve the ability to perform same classification as the original set of attributes. However, our uncertain decision table (UDT) shown in subsection 4.1 is characterized by a high level of uncertain and noisy data. One of the issues with such data is that the resulting reducts are not stable, and are sensitive to sampling. The belief decision rules generated are not suitable for classification. The solution to this problem is to redefine the concept of dynamic reduct [2] in the new context as we have done in [25]. The rules calculated by means of dynamic reducts are better predisposed to classify unseen objects, because they are the most frequently appearing reducts in sub-decision systems created by random samples of a given decision system.

According to the uncertain context, we can redefine the concept of dynamic reduct as follows:

$$UDR(A,\, F) = URed_C(A,\, \{ud\}) \cap \bigcap_{A' \in F} URed_C(A',\, \{ud\}) \tag{39}$$

Where F is a family of subtables of A. Any element of $UDR(A,\, F)$ is called an F-dynamic reduct of A.

In some cases in the uncertain context, we need to apply the general notion called $(F,\, \varepsilon)$-dynamic reducts which is more flexible, where $0 \leq \varepsilon \leq 1$. The set $UDR_\varepsilon(A,\, F)$ of all $(F,\, \varepsilon)$-dynamic reducts is defined by:

$$UDR_\varepsilon(A,\, F) =$$

$$\left\{ R \in URed_C(A,\, \{ud\}) : \frac{|\{A' \in F : R \in URed_C(A',\, \{ud\})\}|}{|F|} \geq 1 - \varepsilon \right\} \tag{40}$$

If $R \in URed(A',\, ud)$, then the number $\frac{|\{A' \in F : R \in URed_C(A',\, \{ud\})\}|}{|F|}$ is called the stability coefficient of R (relative to F).

Example: *Let us continue with the same example in Table 1. Let us remind that the set B=$\{Property,\, Income,\, Unpaid-credit\}$ is the only reduct of A. The two subtables A'=$\{o_1,\, o_2,\, o_3,\, o_4,\, o_9\}$ and A''=$\{o_5,\, o_6,\, o_7,\, o_8,\, o_9\}$ form a family F of sub-decision tables. Both of them have the same relative reduct as the whole uncertain decision table A. So, the set B=$\{Property,\, Income,\, Unpaid-credit\}$ is also the dynamic reduct relative to the chosen family F which is described in Table 3.*

Table 3. The dynamic reduct

U	Property	Income	Unpaid-credit	Loan
o_1	Middle	Average	No	$m_1(\{Yes\}) = 0.95 \quad m_1(\{No\}) = 0.05$
o_2	Middle	High	No	$m_2(\{Yes\}) = 1$
o_3	Middle	Average	No	$m_3(\{Yes\}) = 0.7 \quad m_3(\Theta) = 0.3$
o_4	Less	High	Yes	$m_4(\{No\}) = 0.9 \quad m_4(\Theta) = 0.1$
o_5	Less	High	Yes	$m_5(\{No\}) = 1$
o_6	Middle	Average	No	$m_6(\{No\}) = 0.9 \quad m_6(\Theta) = 0.1$
o_7	Middle	Low	No	$m_7(\{No\}) = 1$
o_8	Less	High	No	$m_8(\{Yes\}) = 1$
o_9	Greater	Average	No	$m_9(\{Yes\}) = 1$

Step 2. Elimination of the redundant objects: After removing the superfluous condition attributes, we can find some redundant objects having the same condition attribute values. They may not have the same bba on decision attribute. So, we combine their bba's using the operator mean as a rule of combination as follows:

$$\bar{\mathrm{m}}_{[o_j]_B}(E) = \frac{1}{|[o_j]_B|} \sum_{o_i \in [o_j]_B} m_i(E), \ for\, all\, E \subseteq \Theta \tag{41}$$

Where B is the relative reduct of C with respect to ud and $[o_j]_B$ is the equivalence class containing the object o_j with respect to the condition attribute subset B.

Please note that the operator mean [12] is more suitable to combine these bba's than the rule of combination in eqn. (21) which is proposed especially to combine different beliefs on decision for one object and not different bba's relative to different objects.

Example: *After removing the superfluous condition attributes and the redundant objects for the uncertain decision table, we obtain Table 4. Let us recall that we have used the notation* $m_{1,3,6}$ *to mean* $\bar{m}_{[o_1]_B}$ *with* B *is equal to* $\{Property, Income, Unpaid-credit\}$.

Step 3. Elimination of the superfluous condition attribute values: To further simplify the uncertain decision table (UDT), we can eliminate some attribute values. We also need to redefine the concept of value reduct and value core as follows:

For each belief decision rule of the form: **If** $C(o_j)$ **then** m_j.
For all $B \subset C$, Let $X = \{o_k | B(o_j) = B(o_k)\, and\, j \neq k\}$
If $X = \emptyset$ **then** $B(o_j)$ is a value reduct of o_j.
Else If $Max(dist(m_j, \bar{\mathrm{m}}_{[o_k]_C})) \leq$ threshold **then** $B(o_j)$ is a value reduct of o_j.

Table 4. Combined bba's of redundant objects

U	Property	Income	Unpaid-credit	Loan
o_1, o_3, o_6	Middle	Average	No	$m_{1,3,6}(\{Yes\}) = 0.55$ $m_{1,3,6}(\{No\}) = 0.32$ $m_{1,3,6}(\Theta) = 0.13$
o_2	Middle	High	No	$m_2(\{Yes\}) = 1$
o_4, o_5	Less	High	Yes	$m_{4,5}(\{No\}) = 0.95$ $m_{4,5}(\Theta) = 0.05$
o_7	Middle	Low	No	$m_7(\{No\}) = 1$
o_8	Less	High	No	$m_8(\{Yes\}) = 1$
o_9	Greater	Average	No	$m_9(\{Yes\}) = 1$

The set of all indispensable attribute values from C for o_j with respect to ud will be called value core and will be denoted by $UCore^j_C(A, \{ud\})$.

$$UCore^j_C(A, \{ud\}) = \bigcap URed^j_C(A, \{ud\}) \tag{42}$$

Where $URed^j_C(A, \{ud\})$ is the family of all reduct value of C for o_j with respect to ud.

Example: *We compute the value reduct of each decision rule from Table 4 with threshold* $= 0.1$. *We take one decision rule from the table: If* $Property = Middle$ *and* $Income = Low$ *and* $Unpaid - credit = No$ *then* m_7.

Let us split the rule into:

1. *If* $Property = Middle$ *then* $m_7 : X = \{o_1, o_2, o_3, o_6\}$ *and dist(*m_7, $m_{1,3,6}$*)* $>$ *0.1. So,* $Property = Middle$ *is not a reduct value.*
2. *If* $Income = Low$ *then* $m_7 : X = \emptyset$. *So,* $Income = Low$ *a reduct value.*
3. *If* $Unpaid - credit = No$ *then* $m_7 : X = \{o_1, o_2, o_3, o_6, o_8, o_9\}$ *and dist(*m_7, $m_{1,3,6}$*)* $>$ *0.1. So,* $Unpaid - credit = No$ *is not a reduct value.*

If we compute the value reduct for all belief decision rules, we obtain Table 5. Note that the object o_2 *has two reduct values.*

Step 4. Generation of belief decision rules: After the simplification of the uncertain decision table, we can generate shorter and significant belief decision rules. With simplification, we can improve the time and the performance of classification of unseen objects. From one example of UDT, we can find many solutions of simplification which have many reducts and value reducts. The number of solutions, denoted nb, is computed as follows:

Table 5. Simplified uncertain decision table

U	Property	Income	Unpaid-credit	Loan
o_1,o_3,o_6	Middle	Average	No	$m_{1,3,6}(\{Yes\}) = 0.55$ $m_{1,3,6}(\{No\}) = 0.32$ $m_{1,3,6}(\Theta) = 0.13$
o_2	-	High	No	$m_2(\{Yes\}) = 1$
$o_{2'}$	Middle	High	-	$m_{2'}(\{Yes\}) = 1$
o_4, o_5	-	-	Yes	$m_{4,5}(\{No\}) = 0.95$ $m_{4,5}(\Theta) = 0.05$
o_7	-	Low	-	$m_7(\{No\}) = 1$
o_8	-	High	No	$m_8(\{Yes\}) = 1$
o_9	Greater	-	-	$m_9(\{Yes\}) = 1$

$$nb = |URed_C(A, \{ud\})| * \prod_{j=1}^{n} |URed_C^j(A, \{ud\})| \tag{43}$$

Example: *Table 5 gives one solution of rule classification described as follows:*
If Property=Middle and Income=Average and Unpaid-credit=No then $m_{1,3,6}$ $(\{Yes\}) = 0.55$ $m_{1,3,6}(\{No\}) = 0.32$ $m_{1,3,6}(\Theta) = 0.13$
If Income=High and Unpaid-credit=No then $m_{2,8}(\{Yes\}) = 1$
If Unpaid-credit=Yes then $m_{4,5}(\{No\}) = 0.95$ $m_{4,5}(\Theta) = 0.05$
If Income=Low then $m_7(\{No\}) = 1$
If Property=Greater then $m_9(\{Yes\}) = 1$

The generated belief decision rules are used to classify unseen objects. To simplify this task, we can compute the pignistic probability to m_j relative to each decision rule using eqn. (25).

4.3 Dynamic Belief Rough Set Classifier Based on Generalization Distribution Table (D-BRSC-GDT)

In this subsection, we propose our second classification approach called Dynamic Belief Rough Set Classifier based on Generalization Distribution Table (D-BRSC-GDT) which can generate a set of rules from uncertain databases with the minimal description length, having large strength and covering all instances. This method is based on the hybrid system GDT-RS [3,29]. In this subsection, we detail the main steps to construct our D-BRSC-GDT method allowing the discovery of classification rules from UDT under the belief function framework based on dynamic approach for feature selection.

Step 1. Creation of the GDT: This step can be omitted because the prior distribution of a generalization can be calculated using eqns. (13) and (14). Note that the standard GDT depends only on condition attributes and not on decision attribute values. In the uncertain context, our GDT still has the same structure as in [30].

Step 2. Feature selection based on dynamic reduct: It is similar to the first step in the construction procedure relative to the D-BRSC. Let us recall that this step consists of removing the superfluous condition attributes that are not in dynamic reducts. This leaves us with a minimal set of attributes that preserve the ability to perform the same classification as the original set of attributes.

Example: *Let us continue with Table 1 to compute the dynamic reduct of our uncertain decision table using eqn. (39). We find that the subset* B=$\{Property, Income, Unpaid-credit\}$ *is also the dynamic reduct using a family* F *formed by two other subtables* A'=$\{o_1, o_2, o_3, o_8, o_9\}$ *and* A''=$\{o_4, o_5, o_6, o_7, o_9\}$. *The two subtables have the same reduct as the whole decision table* A. *Table 3 shows the dynamic reduct relative to the chosen family* F.

Step 3. Definition of the compound object: Consider the indiscernibily classes with respect to the condition attribute subset B as one object, called the compound object o'_j and defined as follows:

$$o'_j = \{o_i | B(o_i) = B(o_j)\} \tag{44}$$

U' is the set of all compound objects obtained from U over the subset of condition attributes B. For objects that compose each compound object, we combine their bba's using the operator mean as rule of combination as follows:

$$m'_j(E) = \frac{1}{|o'_j|} \sum_{o_j \in o'_j} m_j(E), \forall E \subseteq \Theta \tag{45}$$

Example: *Let us continue with the same example. By applying the step 3, we obtain the following tables:*

Step 4. Elimination of the contradictory compound objects: For any compound object o'_j from U', compute $r_{ud_i}(o'_j)$ which represents a noise rate for each decision value ud_i. If there exists a ud_i such that $r_{ud_i}(o'_j) = min\{r_{ud_{i'}}(o'_j) | ud_{i'} \in \Theta\} < T_{noise}$ (threshold value), then we assign the decision value corresponding to ud_i to the object o'_j. If there is no $ud_i \in \Theta$ such that $r_{ud_i}(o'_j) < T_{noise}$, we treat the compound object o'_j as a contradictory one, and set the decision value of o'_j to $\perp$(uncertain).

The noise rate is originally calculated using eqn. (17). This is not appropriate in our uncertain context since the decision value is represented by a bba. In this new situation, the noise rate will be computed as follows:

$$r_{ud_i}(o'_j) = dist(m'_j, m), \text{such that } m(\{ud_i\}) = 1 \tag{46}$$

Table 6. The compound objects

U	U'	Property	Income	Unpaid-credit
o_1, o_3, o_6	o'_1	Middle	Average	No
o_2	o'_2	Middle	High	No
o_4, o_5	o'_4	Less	High	Yes
o_7	o'_7	Middle	Low	No
o_8	o'_8	Less	High	No
o_9	o'_9	Greater	Average	No

Table 7. The combined bba's of each compound object

U'	Property	Income	Unpaid-credit	Loan
o'_1	Middle	Average	No	$m'_1(\{Yes\}) = 0.55$ $m'_1(\{No\}) = 0.32$ $m'_1(\Theta) = 0.13$
o'_2	Middle	High	No	$m'_2(\{Yes\}) = 1$
o'_4	Less	High	Yes	$m'_4(\{No\}) = 0.95$ $m'_4(\Theta) = 0.05$
o'_7	Middle	Low	No	$m'_7(\{No\}) = 1$
o'_8	Less	High	No	$m'_8(\{Yes\}) = 1$
o'_9	Greater	Average	No	$m'_9(\{Yes\}) = 1$

The idea is to use the distance between two bba's m'_j and a certain bba m (such that $m(\{ud_i\}) = 1$). In this manner, we can check that the decisions of all instances belonging to the compound object are close to a certain case. So, it is considered as a non-contradictory compound object.

Example: *By applying the step* 4 *to Table 7, we obtain Table 8.*

Step 5. Minimal description length of decision rule: Let U'' be the set of all non-contradictory compound objects. From U'', select one compound object o'_j. Create a discernibility vector (the row or the column with respect to o'_j in the discernibility matrix) for o'_j.

The discernibility matrix [16] of A is a symmetric $n \times n$ matrix. Each entries a_{ij} in A consists of the set of attributes upon which objects o_i and o_j differ.

$$a_{ij} = \{c \in C | c(o_i) \neq c(o_j)\} \; and \; ud(o_i) \neq ud(o_i) \; for \; i, j = 1, ..., n \qquad (47)$$

Table 8. Contradictory and non-contradictory compound objects

U'	Property	Income	Unpaid-credit	Loan
o'_1	Middle	Average	No	$\perp$
o'_2	Middle	High	No	Yes
o'_4	Less	High	Yes	No
o'_7	Middle	Low	No	No
o'_8	Less	High	No	Yes
o'_9	Greater	Average	No	Yes

Next, we compute all the so-called local relative reducts for the compound object o'_j by using the discernibility function $f_A(o'_j)$. It is a boolean function of k boolean variables corresponding to the k condition attributes defined as below:

$$f_A(o'_j) = \bigwedge\{\bigvee a_{ij} | 1 \leq i \leq n, a_{ij} \neq \emptyset\} \tag{48}$$

The set of all prime implicants of $f_A(o'_j)$ determines the sets of all reduct values of the compound object o'_j.

Example: *According to Table 8, the discernibility vector for the compound object o'_2 is as follows.*

Table 9. Discernibility vector for o'_2

U'	$o'_1(\perp)$	$o'_2(Yes)$	$o'_4(No)$	$o'_7(No)$	$o'_8(Yes)$	$o'_9(Yes)$
$o'_2(Yes)$	Income	$\emptyset$	Property,Unpaid-credit	Income	$\emptyset$	$\emptyset$

We obtain two reducts: {Property, Income} and {Income, Unpaid-credit} by applying the indiscernibility function : $f_A(o'_2)$=$(Income) \wedge (Property \vee Unpaid-credit) \wedge (Income)$= $(Property \wedge Income) \vee (Income \wedge Unpaid-credit)$

Step 6. Selection of the best rules: We start by constructing the decision rules obtained from the local reducts for the object o'_j and revising their strength using eqn. (15). Then, we select only the best rule for o'_j having the best strength.

Example: *According to the same example, the following rules are acquired for object o'_2:*

1. $If\ Property = Middle\ and\ Income = High\ then\ Yes\ with\ S = (\frac{1}{2}*1)*(1)=$ *0.5*
2. $If\ Income = High\ and\ Unpaid-credit = No\ then\ Yes\ with\ S = (\frac{1}{3}*2)*(1)=$ *0.66*

The rule $If\ Income = High\ and\ Unpaid - credit = No\ then\ Yes$ *is selected for the compound object* o'_2 *due to its strength.*

Step 7. Stopping criterion: Let $U'' = U'' - \{o'_j\}$. $If\ U'' \neq \emptyset$, then go back to *Step* 5. Otherwise, STOP.

Example: *As a result, we obtain a set of decision rules that are able to classify unseen objects shown in the Table 10.*

Table 10. Decision rules

U	rules	strengths
o'_2, o'_8	If $Income = High$ and $Unpaid - credit = No$ then Yes	0.66
o'_4	If $Unpaid - credit = Yes$ then No	0.11
o'_7	If $Income = Low$ then No	0.16
o'_9	If $Property = Greater$ then Yes	0.16

4.4 Computational Complexity

Our two classification systems D-BRSC and D-BRSC-GDT are based on the reduct set computation. The time requirement of the decision rule generation can be too expensive when the uncertain decision table has too many attributes or different values of attributes or objects. According to the size of the UDT, the reduct set computation can be exponential and the problem of calculating a minimal reduct is NP-hard [16]. For these reasons, it would be better to apply some heuristic algorithms to avoid the costly computation of a minimal reduct [24].

Note that the time complexity of the D-BRSC and the D-BRSC-GDT is $O(n \times m^2 \times Nr_{max})$, where n is the number of instances in a given database, m stands for the number of attributes, Nr_{max} is the maximal number of reducts.

5 Experimentations

In our experiments, we have performed several tests on real databases obtained from the U.C.I. repository[1] to evaluate our two classifiers: D-BRSC and D-BRSC-GDT. A brief description of these databases is presented in Table 11.

5.1 Constructing Uncertainty in Databases

These databases were artificially modified in order to include uncertainty in the decision attribute. We took different degrees of uncertainty (Low, Middle and High) based on increasing values of probabilities P used to transform the actual decision value d_i of each object o_j to a bba $m_j(\{d_i\}) = 1 - P$ and $m_j(\Theta) = P$.

[1] http://www.ics.uci.edu/ mlearn/MLRepository.html

Table 11. Description of databases

Database	#instances	#attributes	#decision values
W. Breast Cancer	690	8	2
Balance Scale	625	4	3
C. Voting records	497	16	2
Zoo	101	17	7
Nursery	12960	8	3
Solar Flares	1389	10	2
Lung Cancer	32	56	3
Hyes-Roth	160	5	3
Car Evaluation	1728	6	4
Lymphography	148	18	4
Spect Heart	267	22	2
Tic-Tac-Toe Endgame	958	9	2

- Low degree of uncertainty: we take $0 < P <= 0.3$
- Middle degree of uncertainty: we take $0.3 < P <= 0.6$
- High degree of uncertainty: we take $0.6 < P <= 1$

A larger P gives a larger degree of uncertainty.

5.2 Evaluation Criteria and Technique of Sampling

The relevant criteria used to judge the performance and the quality of our proposed classifiers: D-BRSC and D-BRSC-GDT are as follows:

1. *The classification accuracy* represents the percent of correct classification (PCC) of the objects belonging to testing set. For each testing instance, we determine its corresponding rule and we look for the most probable decision corresponding to this rule using the pignistic probability computed from its bba.
2. *The time requirement* represents the number of seconds needed for the construction procedure.
3. *The size* represents the number of belief decision rules generated from the classifier.

Each database is divided into ten parts. Nine parts are used as the training set, the last is used as the testing set. The procedure is repeated ten times, each time another part is chosen as the testing set. This method, called a cross-validation, permits an efficient estimation of the evaluation criterion. In this paper, we perform ten-fold cross-validation tests with different data splits and we report the average of the evaluation criteria.

The technique for sampling of uncertain decision table and dynamic reduct computation in the uncertain context is similar to the technique employed in the certain case [2] which consists of the following steps.

– In the first step, some of samples of the uncertain decision table are computed. Next, reducts for all these subtables are calculated.
– In the second step, reducts with the stability coefficients higher than a fixed threshold are extracted. They are considered as dynamic reducts.

5.3 Experimental Results

In this subsection, we detail different results obtained from our two classifiers (D-BRSC and D-BRSC-GDT) relative to the three evaluation criteria. To evaluate them, we compare the experimental results with those obtained from Static Belief Rough Set Classifier (S-BRSC) [27] and Static Belief Rough Set Classifier based on Generalization Distribution Table (S-BRSC-GDT) [28]. The latter two are based on static reduct in the feature selection step.

To further evaluate the performance of our two proposed dynamic classifiers (D-BRSC, D-BRSC-GDT) with their static versions (S-BRSC, S-BRSC-GDT), we compare the results with those given by another classifier not based on rough set methodology. We choose the Belief Decision Tree (BDT) approach developed in [4]. The latter is a classification technique in an uncertain environment where the uncertainty is represented through the TBM and appears in the decision attribute of training objects. The model generated by the BDT also consists of belief decision rules. Hence, the comparison between all the classifiers is possible. From the procedures of building BDT, we mention the averaging and the conjunctive approaches.

Note that the pruning step is indispensable to improve the size and the accuracy of the BDT. In [22], we have concluded that post-pruning is better than pre-pruning [5] in term of size and accuracy. For this reason, we focus in our experimentations only on the post-pruned BDT in averaging and conjunctive approaches.

The classification accuracy criterion. We start with the more important evaluation criterion, i.e. the classification accuracy. Tables 12, 13 and 14 summarize the experimental results for the PCC relative to the proposed classifiers for all databases and for different degrees of uncertainty. Table 12 reports the results relative to the D-BRSC and the S-BRSC. On the other hand, Table 13 shows the results relative to the D-BRSC-GDT and the S-BRSC-GDT. Table 14 summarizes the results relative to the averaging and the conjunctive pruned BDT.

From these Tables, we conclude that reducing uncertain and noisy databases using dynamic feature selection approach is more suitable for classification of the unseen cases than the static approach. The mean PCC (low, middle, high) obtained from D-BRSC and D-BRSC-GDT is better than the mean PCC obtained from S-BRSC and S-BRSC-GDT. It is true for all databases and for all degrees of uncertainty. For example, the mean PCC for *Balance Scale* database is 87.33% with D-BRSC and 83.57% with S-BRSC. The mean PCC for *Lung Cancer* database is 80.21% with D-BRSC-GDT and 75.49% with S-BRSC-GDT.

Table 12. The PCC relative to D-BRSC and S-BRSC

Database	D-BRSC PCC (%)					S-BRSC PCC (%)				
	Low	Middle	High	Mean	S. deviation	Low	Middle	High	Mean	S. deviation
W. Breast Cancer	89.82	89.63	89.21	89.55	0.31	84.43	84.15	83.65	84.07	0.39
Balance Scale	87.59	87.33	87.07	87.33	0.26	82.72	82.58	82.41	82.57	0.15
C. Voting records	98.91	98.77	98.48	98.72	0.21	98.54	98.31	98.09	98.31	0.22
Zoo	96.56	96.42	95.85	96.27	0.37	94.16	94.03	93.54	93.91	0.32
Nursery	97.64	97.18	97.03	97.28	0.31	96.37	95.94	95.71	96.03	0.33
Solar Flares	92.66	92.58	92.42	92.55	0.12	89.69	89.61	89.57	89.62	0.06
Lung Cancer	81.79	81.54	81.19	81.50	0.30	76.61	76.32	76.18	76.37	0.21
Hyes-Roth	98.83	98.32	97.62	98.27	0.60	97.58	97.24	97.08	97.28	0.25
Car Evaluation	85.71	85.22	85.19	85.37	0.29	83.94	83.68	83.14	83.58	0.40
Lymphography	86.49	86.31	85.75	86.18	0.38	84.73	84.51	84.02	84.45	0.36
Spect Heart	89.47	89.25	89.11	89.27	0.18	87.44	87.33	87.14	87.30	0.15
Tic-Tac-Toe	88.72	88.45	88.23	88.46	0.24	86.87	86.71	86.33	86.63	0.27

Additionally, we conclude that generating uncertain decision rules using D-BRSC yields better PCC than using D-BRSC-GDT. For example, the mean PCC for *W. Breast Cancer* database is 89.55% with D-BRSC and 86.49% with D-BRSC-GDT. Besides, we also conclude that our classification systems based on rough sets are more accurate than the pruned BDT in averaging and conjunctive approaches. We further note that the PCC slightly increases when the uncertainty decreases for all the approaches.

Finally, we also conclude that the standard deviation of the PCC (low, middle, high) obtained from our classification systems based on rough sets: D-BRSC, S-BRSC, D-BRSC-GDT and S-BRSC-GDT is less than the standard deviation obtained from the pruned BDT (averaging and conjunctive). For example, the standard deviation of the PCC for *Spect Heart* database is 0.65 with conjunctive BDT and 0.18 with D-BRSC.

The time requirement criterion. We now move to the second evaluation criterion representing the time requirement needed for the construction procedure to our classification systems. Tables 15, 16 and 17 give the different results obtained from all databases for the time requirement relative to the BRSC (static and dynamic), the BRSC-GDT (static and dynamic) and the pruned BDT (averaging and conjunctive) respectively. Note that the time requirement is almost the same for the different degrees of uncertainty. Note also that from the ten-fold cross-validation tests, we report the minimum, the maximum and the average of the time requirement. These tables show that the two approaches of classification based on static rough sets namely the S-BRSC and the S-BRSC-GDT are faster than the two approaches of classification based on dynamic rough sets: D-BRSC and D-BRSC-GDT. Especially, with large databases like the *Nursery* database where the mean time requirement goes from 380 seconds with S-BRSC

Table 13. The PCC relative to D-BRSC-GDT and S-BRSC-GDT

Database	D-BRSC-GDT PCC (%)					S-BRSC-GDT PCC (%)				
	Low	Middle	High	Mean	S. deviation	Low	Middle	High	Mean	S. deviation
W. Breast Cancer	86.79	86.52	86.18	86.49	0.30	83.76	83.55	83.17	83.48	0.29
Balance Scale	83.66	83.34	83.16	83.39	0.25	80.34	80.21	80.03	80.19	0.15
C. Voting records	98.83	98.55	98.26	98.54	0.28	98.46	98.22	98.03	98.23	0.21
Zoo	96.58	96.31	96.08	96.32	0.25	93.61	93.46	93.02	93.36	0.30
Nursery	97.08	97.01	96.98	97.02	0.05	96.84	96.27	95.87	96.32	0.48
Solar Flares	89.86	89.75	89.59	89.73	0.13	89.67	89.61	89.56	89.61	0.05
Lung Cancer	80.39	80.15	80.08	80.21	0.16	75.78	75.45	75.24	75.49	0.27
Hyes-Roth	98.70	98.50	97.31	98.17	0.75	97.77	97.43	97.13	97.44	0.32
Car Evaluation	84.69	84.31	84.25	84.41	0.23	81.59	81.23	81.20	81.34	0.21
Lymphography	86.16	86.11	85.96	86.08	0.10	84.41	84.17	84.05	84.21	0.18
Spect Heart	88.59	88.36	88.19	88.38	0.20	87.34	87.28	87.07	87.23	0.14
Tic-Tac-Toe	87.54	87.41	87.19	87.38	0.17	86.49	86.36	86.21	86.35	0.14

Table 14. The PCC relative to the averaging and conjunctive BDT

Database	Averaging pruned BDT PCC (%)					Conjunctive pruned BDT PCC (%)				
	Low	Middle	High	Mean	S. deviation	Low	Middle	High	Mean	S. deviation
W. Breast Cancer	83.41	83.01	82.07	82.83	0.68	83.46	83.01	82.17	82.88	0.65
Balance Scale	78.05	77.72	77.67	77.82	0.20	78.15	77.83	77.76	77.91	0.21
C. Voting records	98.21	97.70	97.74	97.84	0.28	98.28	97.76	97.71	97.91	0.32
Zoo	91.90	93.32	91.36	91.53	0.32	91.94	93.36	91.41	91.57	0.32
Nursery	95.79	95.08	95.07	95.31	0.41	95.84	95.13	95.11	95.36	0.42
Solar Flares	85.69	85.49	85.35	85.52	0.19	85.78	85.61	85.46	85.61	0.16
Lung Cancer	74.91	74.70	74.38	74.65	0.21	74.63	74.36	74.07	74.35	0.28
Hyes-Roth	83.64	83.17	82.30	83.02	0.67	83.66	83.31	82.14	83.03	0.80
Car Evaluation	73.38	73.01	72.87	73.11	0.26	73.49	73.11	72.97	73.19	0.27
Lymphography	79.21	78.92	78.91	79.01	0.17	79.25	78.97	78.94	79.05	0.17
Spect Heart	83.25	82.79	82.01	82.67	0.62	83.46	83.01	82.17	82.88	0.65
Tic-Tac-Toe	84.83	84.68	84.35	84.60	0.24	83.91	83.75	83.42	83.69	0.25

to 1106 seconds with D-BRSC. These tables also show that the S-BRSC and the S-BRSC-GDT are slightly faster than the pruned BDT in both averaging and conjunctive approaches. Besides, we also conclude from these tables that S-BRSC-GDT and D-BRSC-GDT approaches, which can avoid many iterations, are slightly faster than S-BRSC and D-BRSC, respectively. For example, the mean time requirement for *Balance Scale* database is 297 seconds with D-BRSC and 282 seconds with D-BRSC-GDT.

Table 15. The time requirement (seconds) relative to D-BRSC and S-BRSC

Database	D-BRSC			S-BRSC		
	Min	Max	Mean	Min	Max	Mean
W. Breast Cancer	279	421	393	149	159	154
Balance Scale	282	312	297	121	134	129
C. Voting records	217	267	235	104	115	110
Zoo	204	256	221	96	107	101
Nursery	1079	1128	1106	271	387	380
Solar Flares	398	447	412	152	163	157
Lung Cancer	181	207	187	48	48	48
Hyes-Roth	199	246	205	85	101	91
Car Evaluation	402	487	426	171	181	178
Lymphography	208	274	217	100	109	102
Spect Heart	229	268	233	106	111	109
Tic-Tac-Toe	303	356	312	137	144	139

Table 16. The time requirement (seconds) relative to D-BRSC-GDT and S-BRSC-GDT

Database	D-BRSC-GDT			S-BRSC-GDT		
	Min	Max	Mean	Min	Max	Mean
W. Breast Cancer	367	423	381	133	139	136
Balance Scale	268	312	282	122	128	126
C. Voting records	210	256	224	102	110	106
Zoo	203	236	207	95	96	96
Nursery	859	895	867	349	378	356
Solar Flares	384	412	391	143	151	146
Lung Cancer	168	195	174	42	42	42
Hyes-Roth	185	221	198	79	87	82
Car Evaluation	395	437	418	155	156	156
Lymphography	202	231	209	93	103	97
Spect Heart	211	244	226	86	96	89
Tic-Tac-Toe	385	405	293	126	129	127

The size criterion. Our third evaluation criterion is the number of the learned uncertain decision rules generated from our classifiers: BRSC (static and dynamic), BRSC-GDT (static and dynamic), and pruned BDT (averaging and conjunctive). Tables 18, 19 and 20 summarize the different results for the size relative to the chosen databases relative to our classification systems. Note that the size is almost the same for the different degrees of uncertainty. Note also that from the ten-fold cross-validation tests, we report the minimum, the maximum and the average of the size. From these tables, we conclude that that the classification based on static rough sets namely S-BRSC and S-BRSC-GDT give a few

Table 17. The time requirement (seconds) relative to averaging and conjunctive BDT

Database	Averaging BDT			Conjunctive BDT		
	Min	Max	Mean	Min	Max	Mean
W. Breast Cancer	151	159	156	152	163	158
Balance Scale	136	143	139	139	143	141
C. Voting records	114	123	117	118	129	122
Zoo	100	107	103	102	109	107
Nursery	378	406	386	383	411	397
Solar Flares	158	166	160	159	166	161
Lung Cancer	54	56	56	54	56	56
Hyes-Roth	91	97	93	91	97	94
Car Evaluation	183	197	189	189	197	193
Lymphography	105	110	108	108	113	112
Spect Heart	109	116	111	112	119	115
Tic-Tac-Toe	146	155	149	148	155	152

Table 18. The size relative to D-BRSC and S-BRSC

Database	D-BRSC			S-BRSC		
	Min	Max	Mean	Min	Max	Mean
W. Breast Cancer	50	55	52	48	53	50
Balance Scale	59	66	63	58	61	59
C. Voting records	52	57	54	49	53	51
Zoo	40	49	44	36	40	38
Nursery	223	239	233	221	232	227
Solar Flares	125	132	129	118	125	121
Lung Cancer	28	28	28	27	27	27
Hyes-Roth	41	47	44	32	41	39
Car Evaluation	174	183	179	171	181	174
Lymphography	56	61	59	53	59	57
Spect Heart	52	55	54	50	54	53
Tic-Tac-Toe	120	131	125	117	127	123

more combined decision rules than the classification based on dynamic rough sets, i.e. D-BRSC and D-BRSC-GDT. For example, the number of the uncertain decision rules induced from *Balance Scale* database goes from 59 using S-BRSC to 63 using D-BRSC. These tables also show that BRSC-GDT (static and dynamic) gives smaller models than BRSC (static and dynamic) as well as pruned BDT (averaging and conjunctive) for all the databases and for all the degrees of uncertainty. For example, the number of the uncertain decision rules induced from *W. Breast Cancer* database goes from 38 using S-BRSC-GDT to 50 using S-BRSC and to 45 using averaging BDT. Finally, we conclude that pruned BDT (averaging and conjunctive) gives a few more combined decision rules for all the databases than BRSC (static and dynamic).

Table 19. The size relative to D-BRSC-GDT and S-BRSC-GDT

Database	D-BRSC-GDT			S-BRSC-GDT		
	Min	Max	Mean	Min	Max	Mean
W. Breast Cancer	40	41	41	36	39	38
Balance Scale	49	52	51	47	51	48
C. Voting records	42	45	40	39	40	39
Zoo	30	31	30	26	26	26
Nursery	191	203	194	186	192	188
Solar Flares	108	117	110	99	106	101
Lung Cancer	27	27	27	21	21	21
Hyes-Roth	30	34	32	29	30	29
Car Evaluation	158	163	161	151	156	154
Lymphography	43	45	44	41	42	41
Spect Heart	40	44	42	39	42	40
Tic-Tac-Toe	108	115	110	102	110	106

Table 20. The size relative to averaging and conjunctive BDT

Database	Averaging BDT			Conjunctive BDT		
	Min	Max	Mean	Min	Max	Mean
W. Breast Cancer	43	47	45	45	49	47
Balance Scale	49	52	51	50	53	52
C. Voting records	40	45	42	42	46	44
Zoo	32	32	32	35	36	35
Nursery	199	203	202	205	210	207
Solar Flares	109	118	112	115	120	117
Lung Cancer	24	24	24	25	26	25
Hyes-Roth	33	34	33	34	35	35
Car Evaluation	167	173	170	169	176	173
Lymphography	47	51	49	50	52	51
Spect Heart	44	45	45	47	48	47
Tic-Tac-Toe	112	117	115	114	120	118

6 Conclusion and Future Work

In this paper, we have presented two new approaches of classification called Dynamic Belief Rough Set Classifier (D-BRSC) and Dynamic Belief Rough Set Classifier based on Generalization Distribution Table (D-BRSC-GDT). Both are based on rough sets and induced from uncertain data under the belief function framework. The feature selection step relative to their construction procedure is based on dynamic approach.

We have done experimentations on real databases to evaluate our proposed classifiers based on three evaluation criteria. To evaluate our approaches, we compare the results with those obtained from Static Belief Rough Set Classifier(S-BRSC)

and Static Belief Rough Set Classifier based on Generalization Distribution Table (S-BRSC-GDT). To further evaluate the performance of our two proposed dynamic classifiers (D-BRSC, D-BRSC-GDT) with their static versions (S-BRSC, S-BRSC-GDT), we compare the results with those given by another similar classifier called the Belief Decision Tree (BDT), which is not based on rough set methodology.

According to the experimental results, we find that generating decision rules based on dynamic approach of feature selection is more suitable for classification process than the static one. In addition, the experimental results show that all our classification systems based on rough sets (dynamic and static versions) are more accurate than the pruned BDT in averaging and conjunctive approaches. As a future work, we plan to work with uncertainty in condition attribute values.

Acknowledgements

The authors would like to thank the anonymous reviewers for their detailed comments, criticisms, and suggestions that have helped greatly in improving the quality of this paper.

References

1. Bauer, M.: Approximation algorithms and decision making in the Dempster-Shafer theory of evidence - An empirical study. IJAR 17(2-3), 217–237 (1997)
2. Bazan, J., Skowron, A., Synak, P.: Dynamic reducts as a tool for extracting laws from decision tables. In: Raś, Z.W., Zemankova, M. (eds.) ISMIS 1994. LNCS (LNAI), vol. 869, pp. 346–355. Springer, Heidelberg (1994)
3. Dong, J.Z., Zhong, N., Ohsuga, S.: Probabilistic rough induction: The GDT-RS methodology and algorithms. In: Raś, Z.W., Skowron, A. (eds.) ISMIS 1999. LNCS (LNAI), vol. 1609, pp. 621–629. Springer, Heidelberg (1999)
4. Elouedi, Z., Mellouli, K., Smets, P.: Belief decision trees: Theoretical foundations. International Journal of Approximate Reasoning 28(2-3), 91–124 (2001)
5. Elouedi, Z., Mellouli, K., Smets, P.: A Pre-Pruning Method in Belief Decision Trees. In: The Ninth International Conference on Information Processing and Management of uncertainty in Knowledge-Based Systems, IPMU 2002, Annecy, France, vol. 1, pp. 579–586 (2002)
6. Elouedi, Z., Mellouli, K., Smets, P.: Assessing sensor reliability for multisensor data fusion within the transferable belief model. IEEE Trans. Syst. Man Cyben. 34(1), 782–787 (2004)
7. Fixen, D., Mahler, R.P.S.: The modified Dempster-Shafer approach to classification. IEEE Trans Syst Man Cybern. 27(1), 96–104 (1997)
8. Grzymala-Busse, J.W.: Rough set strategies to data with missing attribute values. In: Workshop Notes, Foundations and New Directions of Data Mining, the 3rd International Conference on Data Mining, Melbourne, Florida, pp. 56–63 (2003)
9. Grzymala-Busse, J.W., Siddhaye, S.: Rough Set Approaches to Rule Induction from Incomplete Data. In: Proceedings of the IPMU 2004, Perugia, Itally, July 4-9, vol. 2, pp. 923–930 (2004)

10. Hong, T.P., Tseng, L.H., Chien, B.C.: Learning fuzzy rules from incomplete quantitative data by rough sets, pp. 1438–1442. IEEE, Los Alamitos (2002)
11. Jousseleme, A.L., Grenier, D., Bosse, E.: A new distance between two bodies of evidence. Information Fusion 2(2), 91–101 (2001)
12. Murphy, C.K.: Combining belief functions when evidence conflits. Decision Support Systems 29, 1–9 (2000)
13. Pawlak, Z.: Rough Sets. International Journal of Computer and Information Sciences 11(5), 341–356 (1982)
14. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishing, Dordrecht (1991)
15. Shafer, G.: A mathematical theory of evidence. Princeton University Press, Princeton (1976)
16. Skowron, A., Rauszer, C.: The discernibility matrices and functions in information systems. In: Slowinski, R. (ed.) Intelligent Decision Support, pp. 331–362. Kluwer Academic Publishers, Boston (1992)
17. Skowron, A., Grzymala-Busse, J.W.: From rough set theory to evidence theory. In: Advances in the Dempster-Shafer Theory of Evidence, New York, pp. 193–236 (1994)
18. Smets, P., Kennes, R.: The transferable belief model. Artificial Intelligence 66(2), 191–234 (1994)
19. Smets, P.: The transferable belief model for quantified belief representation. In: Gabbay, D.M., Smets, P. (eds.) Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol. 1, pp. 267–301. Kluwer, Doordrecht (1998)
20. Smets, P.: Application of the transferable belief model to diagnostic problems. International Journal of Intelligent Systems 13(2-3), 127–157 (1998)
21. Tessem, B.: Approximations for efficient computation in the theory of evidence. Artif. Intell. 61(2), 315–329 (1993)
22. Trabelsi, S., Elouedi, Z., Mellouli, K.: Pruning belief decision tree methods in averaging and conjunctive approaches. International Journal of Approximate Reasoning 46, 91–124 (2007)
23. Trabelsi, S., Elouedi, Z.: Learning decision rules from uncertain data using rough sets. In: The 8th International FLINS Conference on Computational Intelligence in Decision and Control, Madrid, Spain, September 21-24, pp. 114–119. World Scientific, Singapore (2008)
24. Trabelsi, S., Elouedi, Z.: Heuristic method for attribute selection from partially uncertain data using rough sets. International Journal of General Systems 39(3), 271–290 (2010)
25. Trabelsi, S., Elouedi, Z., Lingras, P.: Dynamic reduct from partially uncertain data using rough sets. In: Sakai, H., Chakraborty, M.K., Hassanien, A.E., Ślęzak, D., Zhu, W. (eds.) RSFDGrC 2009. LNCS (LNAI), vol. 5908, pp. 160–167. Springer, Heidelberg (2009)
26. Trabelsi, S., Elouedi, Z., Lingras, P.: A comparison of dynamic and static belief rough set classifier. In: Szczuka, M., Kryszkiewicz, M., Ramanna, S., Jensen, R., Hu, Q. (eds.) RSCTC 2010. LNCS (LNAI), vol. 6086, pp. 366–375. Springer, Heidelberg (2010)
27. Trabelsi, S., Elouedi, Z., Lingras, P.: Belief rough set classifier. In: Gao, Y., Japkowicz, N. (eds.) AI 2009. LNCS (LNAI), vol. 5549, pp. 257–261. Springer, Heidelberg (2009)

28. Trabelsi, S., Elouedi, Z., Lingras, P.: Rule discovery process based on rough sets under the belief function framework. In: Hüllermeier, E., Kruse, R., Hoffmann, F. (eds.) IPMU 2010. LNCS (LNAI), vol. 6178, pp. 726–736. Springer, Heidelberg (2010)
29. Zhong, N., Dong, J.Z., Ohsuga, S.: Data mining: A probabilistic rough set approach. In: Rough Sets in Knowledge Discovery, vol. 2, pp. 127–146 (1998)
30. Zhong, N., Ohsuga, S.: Using generalization distribution tables as a hypotheses search space for generalization. In: Proceedings of Fourth International Workshop on Rough Sets, Fuzzy Sets, and Machine Discovery, RSFD 1996, pp. 396–403 (1996)
31. Zouhal, L.M., Denoeux, T.: An evidence-theory k-NN rule with parameter optimization. IEEE Trans. Syst. Man Cybern. 28(2), 263–271 (1998)