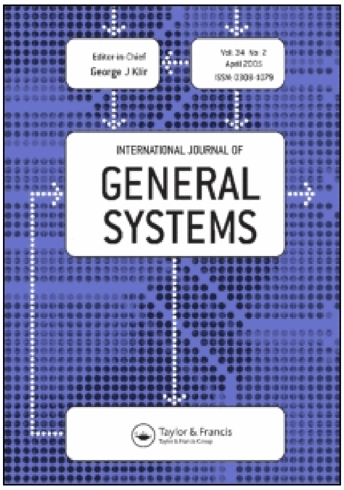

International Journal of General Systems
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713642931

Heuristic method for attribute selection from partially uncertain data using rough sets

Salsabil Trabelsi [a];Zied Elouedi [a]
[a] Larodec Laboratory, Tunis Higher Institute of Management, Le Bardo, Tunisia

Online publication date: 05 March 2010

To cite this Article Trabelsi, Salsabil andElouedi, Zied(2010) 'Heuristic method for attribute selection from partially uncertain data using rough sets', International Journal of General Systems, 39: 3, 271 — 290
To link to this Article: DOI: 10.1080/03081071003625305
URL: http://dx.doi.org/10.1080/03081071003625305

International Journal of General Systems
Vol. 39, No. 3, April 2010, 271–290

Heuristic method for attribute selection from partially uncertain data using rough sets

Salsabil Trabelsi* and Zied Elouedi

Larodec Laboratory, Tunis Higher Institute of Management, 41 rue de la liberté, 2000 Le Bardo, Tunisia

(*Received 31 January 2009; final version received 29 May 2009*)

In this paper, we deal with the problem of attribute selection from partially uncertain data based on rough sets without costly calculation. The uncertainty exists in decision attributes and is represented by the transferable belief model, one interpretation of the belief function theory. To solve this problem, we propose a heuristic method for attribute selection able to extract the more relevant features needed in the classification process. The simplification of the uncertain decision table using this heuristic method yields to learn simplified and more significant belief decision rules in a quick time. The experiments show interesting results based on two evaluation criteria such as the accuracy classification and the time complexity.

Keywords: uncertainty; belief function theory; rough sets; heuristic; attribute selection; classification

1. Introduction

The knowledge discovery databases (KDD) process usually involves multi-steps like attribute selection, discretisation of continuous attributes and decision-rule generation. In fact, one of the important phases of the modelling process studied in machine learning is feature reduction. It is defined as a process of finding an optimal subset of features from the original set according to some criteria. Features can be redundant and irrelevant (having no effect on processing performance), especially in real-world databases characterised by a large number of attributes. 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.

The rough set theory constitutes a sound basis for data mining proposed as a tool to discover hidden patterns in data. The rough set approach offers solutions to the problem of decision-rule generation, attribute selection and discretisation. Using rough sets for attribute selection was proposed in several contributions (Pawlak 1991, Skowron and Rauszer 1992, Modrzejewski 1993). The simplest approach is based on the calculation of a core for a discrete attribute dataset, containing the more relevant attributes, and reducts, containing a core plus additional weakly relevant features. In fact, a reduct is a minimal set of attributes, which preserves the ability to perform classifications as the whole attribute set does. To find the optimal reduct from the decision system is a non-deterministic polynominal time (NP)-hard problem. Many researchers proposed several heuristic methods for feature selection using rough sets in order to avoid the NP-hard complexity

*Corresponding author. Email: salsabilt@yahoo.fr

ISSN 0308-1079 print/ISSN 1563-5104 online

DOI: 10.1080/03081071003625305
http://www.informaworld.com

algorithm to find the optimal reduct from all possibilities (Wroblewski 1995, Chouchoulas and Shen 2001, Zhong *et al.* 2001, Jensen and Shen 2003).

Another issue in real-world databases is the uncertainty. This kind of data exists in many real-world applications like medicine, where symptoms (condition attributes) or diseases (decision attributes) of some patients (objects) may be totally or partially uncertain. It is not efficient to eliminate these objects from the classification process because it will result in a loss of important information. Much research has been done to adapt rough sets to this kind of environment (Slowinski and Stefanowski 1989, Orlowska 1994, Kryszkiewiez 1995, Grzymala-Busse 2003). These extensions do not deal with partially uncertain decision attribute values in a decision system. This kind of uncertainty can be represented by the theory of belief functions introduced by Shafer (1976). The latter has been proposed for modelling someone's degrees of belief resulting from uncertainty. It is considered to be 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 real-life problems for decision making and classification.

In this paper, we deal with the problem of attribute selection using a heuristic method from partially uncertain data based on rough sets. Actually, uncertainty is handled in decision attributes rather than in condition attribute values of the decision system (we deal only with symbolic attributes). The uncertainty is represented by the transferable belief model (TBM), one interpretation of the belief function theory (Smets 1994, 1998a). To solve this problem, we have chosen the heuristic method proposed by Zhong *et al.* (2001) and we will adapt it to extract the more relevant subset of attributes in a quick time from partially uncertain data under the belief function framework. The new formulation of the heuristic method in this new context is based on the redefined basic concepts of rough sets such as the indiscernibility relation, set approximation and the positive region. We will also need the new formalism of the concept of reduct and core in this context in order to eliminate the superfluous attributes. The simplification of the decision system using the heuristic method yields to learn and generate simplified and relevant belief decision rules without costly calculation.

This paper is organised as follows: Section 2 provides an overview of the rough set theory. Section 3 introduces the belief function theory as understood in the TBM. In Section 4, we adapt an attribute selection heuristic using rough sets to simplify partially uncertain data. Finally, in Section 5, we carry out experiments to evaluate our heuristic feature selection method and the generated belief decision rules on different databases based on two evaluation criteria: accuracy classification and time complexity.

2. Rough sets

The idea of rough sets has been introduced by Pawlak (1982, 1991) to deal with imprecise and vague concepts. In recent years, we have witnessed a rapid growth of interest in rough set theory and its applications. It constitutes a sound basis for KDD. It can be used for feature selection, discretisation, data reduction, decision-rule generation, etc. Here, we introduce only the basic notations from rough set approach used in this paper.

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 (Pawlak 1982) in terms of a pair $A = (U, C)$, where U is a finite set of objects (cases) called *the universe*

$U = \{o_1, o_2, \ldots, o_n\}$ and C is a non-empty, finite set of *condition* attributes, $C = \{c_1, c_2, \ldots, c_k\}$.

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*. The value set of d, called $\Theta = \{d_1, d_2, \ldots, d_s\}$. 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 $o_j \in U$, similar to the notation $d(o_j)$, which represents the value of the decision attribute d for an object o_j.

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 rough sets adopt the concept of indiscernibility relation (Pawlak 1982, 1991) 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 $[o_j]_B$,

$$\mathrm{IND}_B = U/B = \{[o_j]_B | o_j \in U)\}, \tag{1}$$

where

$$[o_j]_B = \{o_i | \forall\, c \in B c(o_i) = c(o_j)\}. \tag{2}$$

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

$$\mathrm{IND}_{\{d\}} = U/\{d\} = \{[o_j]_{\{d\}} | o_j \in U)\}. \tag{3}$$

2.3 *Set approximation*

The concept 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 (Pawlak 1982).

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 by $\underline{B}(X)$ and $\bar{B}(X)$, respectively, where

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

The objects in $\underline{B}(X)$ can be with certainty and are 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.

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:

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

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

2.4 *Decision rules*

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

$$\alpha \rightarrow \beta \text{ with S},$$

where α denotes the conjunction of the conditions that a concept must satisfy, β a concept that the rule describes and S a *measure of strength* of which the rule holds.

The support S gives is a measure of how trustworthy the rule is in drawing the conclusion β on the basis of evidence α and is a frequency-based estimate of conditional Probability $\Pr(\beta/\alpha)$.

2.5 *Dependency degree*

Another important issue in data analysis is discovering dependencies between attributes (Pawlak 1991). Intuitively, the decision attribute d depends totally on a set of condition attributes C, denoted as $C \Rightarrow \{d\}$, 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

$$k = \gamma(C, \{d\}) = \frac{|\mathrm{Pos}_C(\{d\})|}{|U|}, \tag{6}$$

where

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

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

If $k = 1$, we say that the attribute d depends totally on C, and if $k < 1$, we say that the attribute d depends partially (in a degree k) on the set of attributes C. The coefficient k expresses the ratio of all elements of the universe, which can be property classified to blocks of partition $U/\{d\}$, employing the set of attributes C and will be called the *degree of dependency*.

2.6 *Core and reduct*

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. There are usually several such subsets of attributes and those which are minimal are called reducts (Pawlak 1991). In order to express the above idea more precisely, we need some auxiliary notions.

2.6.1 *Dispensable and indispensable attributes*

Let $c \in C$, the attribute c is dispensable in C with respect to d, iff $\mathrm{Pos}_C(\{d\}) = \mathrm{Pos}_{C-c}(\{d\})$. 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 will be called independent.

2.6.2 *Reducts*

A subset $B \subseteq C$ is a reduct of C with respect to d, iff B is independent and

$$\mathrm{Pos}_B(\{d\}) = \mathrm{Pos}_C(\{d\}). \tag{8}$$

A reduct is a minimal subset 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. Computing reducts is not a trivial task that can be solved by a simple increase of time complexity. It is, in fact, one of the drawbacks of rough set methodology. Fortunately, there exist good heuristics (Wroblewski 1995, Chouchoulas and Shen 2001, Zhong *et al.* 2001, Jensen and Shen 2003) that compute sufficiently reducts in often acceptable time.

2.6.3 *Core*

The set of all the condition attributes indispensable in C with respect to d is denoted by $\mathrm{Core}(C)$,

$$\mathrm{Core}(C) = \bigcap \mathrm{Red}(C), \tag{9}$$

where $\mathrm{Red}(C)$ is the set of all reducts of C.

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, 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 in the TBM (Smets 1994, 1998a). This theory is also appropriate to handle uncertainty in classification problems (Denoeux 1995, 2000, Elouedi *et al.* 2001, Ben Hariz *et al.* 2006).

3.1 *Definitions*

The TBM is a model to represent quantified belief functions (Smets 1994). 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,

$$\sum_{E \subseteq \Theta} m(E) = 1. \tag{10}$$

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

Associated with m is the belief function, denoted bel, corresponding to a specific bba, m assigns to every subset E of Θ, the sum of masses of belief committed to every subset of E by m (Shafer 1976). Contrary to the bba which expresses only the part of belief that one commits to E without also being committed to $\overline{E}$.

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

$$\text{bel}(E) = \sum_{\emptyset \neq F \subseteq E} m(F). \tag{11}$$

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 bbm relative to subsets F compatible with E.

The pl is defined as follows:

$$\text{pl}(E) = \sum_{E \cap F \neq \emptyset} m(F), \quad \forall E \subseteq \Theta. \tag{12}$$

The bba (m), the bel and the 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 TBM, the bba 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 (Smets 1998b):

$$(m_1 \cap m_2)(E) = \sum_{F,G \subseteq \Theta : F \cap G = E} m_1(F) m_2(G). \tag{13}$$

The conjunctive rule is considered as an unnormalised Dempster's rule of combination dealing with the closed world assumptions, defined as follows (Shafer 1976):

$$(m_1 \oplus m_2)(A) = K(m_1 \cap m_2)(A), \tag{14}$$

where

$$K^{-1} = 1 - (m_1 \cap m_2)(\emptyset) \tag{15}$$

and

$$(m_1 \oplus m_2)(\emptyset) = 0. \tag{16}$$

where K is called *the normalisation factor*.

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 (Smets 1998b):

$$(m_1 \cup m_2)(E) = \sum_{F,G \subseteq \Theta : F \cup G = E} m_1(F) m_2(G). \tag{17}$$

3.3 *Discounting*

In the TBM, discounting allows us to take into consideration the reliability of the information source that generates the bba m. For $\alpha \in [0, 1]$, let $(1 - \alpha)$ 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}:

$$m^{\alpha}(E) = (1 - \alpha) m(E), \quad \text{for } E \subset \Theta, \tag{18}$$

$$m^{\alpha}(\Theta) = \alpha + (1 - \alpha) m(\Theta). \tag{19}$$

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 the pignistic probabilities denoted $\mathrm{Bet}P$ and is defined as follows (Smets 1998a):

$$\mathrm{Bet}P(\{a\}) = \sum_{F \subseteq \Theta} \frac{|\{a\} \cap F|}{|F|} \frac{m(F)}{(1 - m(\emptyset))}, \quad \text{for all } a \in \Theta. \tag{20}$$

4. Heuristic of simplification uncertain decision system using rough sets

Finding all possible reducts from an uncertain decision table is an NP-hard complexity problem like in the certain case, especially with datasets containing large numbers of features. In such a case, it would be impossible to process further. Another problem of finding all possible reducts using rough sets: what is the best reduct for the classification process? Which one should we select? The solution to these problems is to apply a heuristic attribute selection method able to select the relevant features from our partially uncertain data. In this section, we start by presenting an overview among heuristic attribute selection methods based on rough sets and choose one of them to adapt it to the uncertain context. Then, we describe the redefined basic concepts of rough sets in our new situation. Finally, we propose the adaptation of the chosen heuristic method of feature selection.

4.1 *Overview of heuristic attribute selection methods using rough sets*

Many researchers proposed several heuristic methods for feature selection using rough sets. In this subsection, we present an overview of these heuristic methods and then we choose one of them to adapt to extract the more relevant subset of attributes in a quick time from partially uncertain data under the belief function framework.

In Zhong *et al.* (2001), a heuristic filter-based approach is presented based on rough set theory. The algorithm proposed starts with the core of the dataset and incrementally adds attributes based on a heuristic measure. Additionally, a threshold value is required as a stopping criterion to determine when a reduct candidate is 'near enough' to being a reduct.

In Wroblewski (1995), another attribute selection heuristic was proposed which uses genetic algorithms to discover optimal or close-to-optimal reducts. Reduct candidates are encoded as bit strings, with the value in position i set if the ith attribute is present. The fitness function depends on two parameters. The first parameter is the number of bits set. The function penalises those strings which have larger numbers of bits set, driving the process to find smaller reducts. The second is the number of classifiable objects given this candidate. The reduct should discern between as many objects as possible (ideally all of them). This heuristic appears to be fast, but sometimes fails to find the global optimum.

The QuickReduct algorithm (adapted from Chouchoulas and Shen 2001) attempts to calculate a reduct without exhaustively generating all possible subsets. It starts off with an empty set and adds in turn, one at a time, those attributes that result in the greatest increase in the rough set dependency metric. This process continues until the dependency of the reduct equals the consistency of the dataset (1 if the dataset is consistent). Determining the consistency of the entire dataset is reasonable for most datasets. However, it may be unfeasible for very large data, so alternative stopping criteria may have to be used. One such criterion could be to terminate the search when there is no further increase in the dependency measure. Other developments include ReverseReduct where the strategy is backward elimination of attributes as opposed to the current forward selection process. Initially, all attributes appear in the reduct candidate; the least informative ones are incrementally removed until no further attribute can be eliminated without introducing inconsistencies. However, QuickReduct and ReverseReduct are not guaranteed to find a minimal subset. Using the dependency function to discriminate between candidates may lead the search down a non-minimal path. It is impossible to predict which combinations of attributes will lead to an optimal reduct based on changes in dependency with the addition or deletion of single attributes. They do result in a close-to-minimal subset.

Another heuristic is described in Jensen and Shen (2003). It is an ant-based framework applied to rough set-based selection. The precomputed heuristic desirability of edge traversal is the entropy measure, with the subset evaluation performed using the rough set dependency heuristic (to guarantee that true rough set reducts are found). The number of ants used is set to the number of features, with each ant starting on a different feature. Ants construct possible solutions until they reach a rough set reduct. To avoid fruitless searches, the size of the current best reduct is used to reject those subsets whose cardinality exceeds this value.

Among all these heuristic methods for feature selection, we will choose the heuristic proposed in Zhong *et al.* (2001) to adapt it to our context. The advantages of this heuristic is that it is fast and generates only one reduct. It does not guarantee to find a global optimum, however, it is better than the other heuristics at avoiding local optimum. It holds more flexibility with threshold.

4.2 *Basic concepts of rough sets under uncertainty*

In order to adapt the heuristic feature selection method based on rough sets proposed in Zhong *et al.* (2001) in our new context, we need to redefine the basic concepts of rough sets under uncertainty. This subsection describes the modified definitions of decision system, indiscernibility relation, set approximation, positive region, dependency degree and especially the core and reduct. These adaptations were partially proposed originally in Trabelsi and Elouedi (2008, 2009).

Table 1. Uncertainty decision table.

U	*Headache*	*Muscle-pain*	*Temperature*	*Flu*
o_1	Yes	Yes	Very high	$m_1(\{\text{yes}\}) = 1$
o_2	Yes	No	High	$m_2(\{\text{no}\}) = 1$
o_3	Yes	Yes	Normal	$m_3(\{\text{yes}\}) = 0.5$, $m_3(\Theta) = 0.5$
o_4	No	Yes	Normal	$m_4(\{\text{no}\}) = 0.6$, $m_4(\Theta) = 0.4$
o_5	No	Yes	Normal	$m_5(\{\text{no}\}) = 1$
o_6	Yes	No	High	$m_6(\{\text{no}\}) = 1$
o_7	No	Yes	Very high	$m_7(\{\text{yes}\}) = 1$
o_8	No	Yes	High	$m_8(\{\text{yes}\}) = 1$

4.2.1 *Decision system under uncertainty*

Our uncertain decision system denoted A contains n objects o_j, characterised by a set of certain condition attributes $C = \{c_1, c_2, \ldots, c_k\}$ 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 $\Theta = \{\text{ud}_1, \text{ud}_2, \ldots, \text{ud}_s\}$ representing the possible values of ud.

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

For the patient o_3, 0.5 of beliefs are exactly committed to the decision $\text{ud}_1 = \text{yes}$, whereas 0.5 of beliefs is assigned to the whole of frame of discernment Θ (ignorance). With bba, we can represent the certain case (with certain decision) like for the objects o_1, o_2, o_5, o_6, o_7 and o_8.

4.2.2 *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 to assign each object to the right equivalence classes for optimal decision making. 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, to compute the pignistic probability, denoted $\text{Bet}P_j$, by applying the pignistic transformation to m_j.

For every ud_i, a decision value, we define:

$$X_i = \{o_j | \text{Bet}P_j(\{\text{ud}_i\}) > 0\}, \tag{21}$$

$$\text{IND}_{\{\text{ud}\}} = U/\{\text{ud}\} = \{X_i | \text{ud}_i \in \Theta\}. \tag{22}$$

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: $U/C = \{\{o_1\}, \{o_2, o_6\}, \{o_3\}, \{o_4, o_5\}, \{o_7\}, \{o_8\}\}$ and to compute the equivalence classes based on uncertain decision attribute *U*/{ud} as shown in Table 2.

According to Table 2, the objects o_1, o_7 and o_8 are assigned to the equivalence class $\text{ud}_1 = \text{yes}$. The objects o_2, o_5 and o_6 are assigned to the equivalence class $\text{ud}_2 = \text{no}$. The objects o_3 and o_4 are included in the two equivalence classes. So, the two equivalence

Table 2. Pignistic transformation to m_j for o_j.

m_j	$BetP_j$
m_1	$\text{Bet}P_1(\{\text{yes}\}) = 1, \ \text{Bet}P_1(\{\text{no}\}) = 0$
m_2	$\text{Bet}P_2(\{\text{yes}\}) = 0, \ \text{Bet}P_2(\{\text{no}\}) = 1$
m_3	$\text{Bet}P_3(\{\text{yes}\}) = 0.75, \ \text{Bet}P_3(\{\text{no}\}) = 0.25$
m_4	$\text{Bet}P_4(\{\text{yes}\}) = 0.2, \ \text{Bet}P_4(\{\text{no}\}) = 0.8$
m_5	$\text{Bet}P_5(\{\text{yes}\}) = 0, \ \text{Bet}P_5(\{\text{no}\}) = 1$
m_6	$\text{Bet}P_6(\{\text{yes}\}) = 0, \ \text{Bet}P_6(\{\text{no}\}) = 1$
m_7	$\text{Bet}P_7(\{\text{yes}\}) = 1, \ \text{Bet}P_7(\{\text{no}\}) = 0$
m_8	$\text{Bet}P_8(\{\text{yes}\}) = 1, \ \text{Bet}P_8(\{\text{no}\}) = 0$

classes $\text{ud}_1 = \text{yes}$ and $\text{ud}_2 = \text{no}$ based on the uncertain decision attribute are as follows: $U/\{\text{ud}\} = \{\{o_1, o_3, o_4, o_7, o_8\}, \{o_2, o_3, o_4, o_5, o_6\}\}$.

4.2.3 *Set approximation*

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

(1) For each equivalence class based on condition attributes C, combine their bba using the operator mean. In order to check which of them has a certain bba, the operator mean is more suitable in our case to combine these bba's than the rule of combination in Equation (13) which is proposed especially to combine different beliefs on decision for one object and not different beliefs for different objects.
(2) For each equivalence class X_i based on uncertain decision attribute, we compute the new lower and upper approximations, as follows:

$$\underline{C}X_i = \{o_j | [o_j]_C \subseteq X_i \text{ and } m_j(\{\text{ud}_i\}) = 1\}.$$

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

$$\bar{C}X_i = \{o_j | [o_j]_C \cap X_i \neq \emptyset\}.$$

We compute the upper as the same manner as in the certain case.

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 of the equivalence classes *U*/*C* using the operator mean. Tables 3 and 4 represent the combined bba for the subsets $\{o_2, o_6\}$ and $\{o_4, o_5\}$.

Next, we compute the lower and upper approximations for each of the equivalence class $U/\{\text{ud}\}$.

For $\text{ud}_1 = \text{yes}$, let $X_1 = \{o_1, o_3, o_4, o_7, o_8\}$.

The subsets $\{o_1\}$, $\{o_7\}$ and $\{o_8\}$ are included to X_1 and have a certain bba. Hence, we put them in the lower $\underline{C}X_1$. The subset $\{o_3\}$ is included to X_1, but it has an uncertain bba. So, we put it in the upper $\bar{C}X_1$. The subset $\{o_4, o_5\}$ is partially included to X_1. So, we put

Table 3. The combined bba for the subset $\{o_2, o_6\}$.

Patient	$m(\{yes\})$	$m(\{no\})$	$m(\Theta)$
o_2	0	1	0
o_6	0	1	0
m	0	1	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

it in the upper $\bar{C}X_1$,

$$\underline{C}X_1 = \{o_1, o_7, o_8\} \quad \text{and} \quad \bar{C}X_1 = \{o_1, o_3, o_4, o_5, o_7, o_8\}.$$

For $\mathrm{ud}_2 = \mathrm{no}$, let $X_2 = \{o_2, o_3, o_4, o_5, o_6\}$.

The subset $\{o_2, o_6\}$ is included to X_2 and has a certain bba. Hence, we let it in the lower $\underline{C}X_2$. The subsets $\{o_4, o_5\}$ and $\{o_3\}$ are included to X_2. However, they have an uncertain bba. So, we put them in the upper $\bar{C}X_2$.

$$\underline{C}X_2 = \{o_2, o_6\} \quad \text{and} \quad \bar{C}X_2 = \{o_2, o_3, o_4, o_5, o_6\}.$$

4.2.4 *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. Some of the belief decision rules induced from our decision table for the object o_3 and o_8 are as follows:

If Headache = yes and Muscle-pain = yes and Temperature = normal Then $m_3(\{\mathrm{yes}\}) = 0.5$ $m_3(\Theta) = 0.5$.
If Headache = no and Muscle-pain = yes and Temperature = high Then $m_8(\{\mathrm{yes}\}) = 1$.

Hence, these belief decision rules could be simplified by removing superfluous attributes. With simplification, we can improve the time and the performance of classification of unseen objects.

4.2.5 *Positive region*

With this new lower approximation, we can define the new positive region denoted $U\,\mathrm{Pos}_C(\{\mathrm{ud}\})$:

$$U\,\mathrm{Pos}_C(\{\mathrm{ud}\}) = \bigcup_{X_i \in U/\{\mathrm{ud}\}} \underline{C}X_i. \tag{23}$$

4.2.6 *Dependency degree*

We can compute the new dependency degree as follows:

$$\gamma(C, \{\mathrm{ud}\}) = \frac{|U\,\mathrm{Pos}_C(\{\mathrm{ud}\})|}{|U|}. \tag{24}$$

Example. Let us continue with the same example, to compute the positive region and dependency degree of the uncertain decision system A,

$$U\,\mathrm{Pos}_C(\{\mathrm{ud}\}) = \{o_1, o_2, o_6, o_7, o_8\}, \quad \gamma(C, \{\mathrm{ud}\}) = \frac{5}{8}.$$

4.2.7 *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,

$$U\,\mathrm{Pos}_B(\{\mathrm{ud}\}) = U\,\mathrm{Pos}_C(\{\mathrm{ud}\}). \tag{25}$$

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,

$U\,\mathrm{Pos}_{\{\mathrm{Headache}\}}(\{\mathrm{ud}\}) = \emptyset$,
$U\,\mathrm{Pos}_{\{\mathrm{Muscle\text{-}pain}\}}(\{\mathrm{ud}\}) = \emptyset$,
$U\,\mathrm{Pos}_{\{\mathrm{Temperature}\}}(\{\mathrm{ud}\}) = \{o_1, o_7\}$,
$U\,\mathrm{Pos}_{\{\mathrm{Headache,Muscle\text{-}pain}\}}(\{\mathrm{ud}\}) = \{o_2, o_6\}$,
$U\,\mathrm{Pos}_{\{\mathrm{Headache,Temperature}\}}(\{\mathrm{ud}\}) = \{o_1, o_2, o_6, o_7, o_8\}$,
$U\,\mathrm{Pos}_{\{\mathrm{Muscule\text{-}pain,Temperature}\}}(\{\mathrm{ud}\}) = \{o_1, o_2, o_6, o_7, o_8\}$.

We find that only the subsets {Muscle-pain, Temperature} and {Headache, Temperature} have the same positive region that the whole subset of condition attributes C. So, {Muscle-pain, Temperature} and {Headache, Temperature} are two relative reducts to the decision Flu in our uncertain decision system A. It can be simplified in Table 5 or Table 6. The relative core is the attribute *Temperature*. It is the intersection of the two relative reducts.

4.2.8 *Belief decision rules after simplification*

The simplification of the uncertain decision system leads to extract the more generalised belief decision rules.

Table 5. The first reduct.

U	*Muscle-pain*	*Temperature*	*Flu*
o_1	Yes	Very high	$m_1(\{\mathrm{yes}\}) = 1$
o_2	No	High	$m_2(\{\mathrm{no}\}) = 1$
o_3	Yes	Normal	$m_3(\{\mathrm{yes}\}) = 0.5$, $m_3(\Theta) = 0.5$
o_4	Yes	Normal	$m_4(\{\mathrm{no}\}) = 0.6$, $m_4(\Theta) = 0.4$
o_5	Yes	Normal	$m_5(\{\mathrm{no}\}) = 1$
o_6	No	High	$m_6(\{\mathrm{no}\}) = 1$
o_7	Yes	Very high	$m_7(\{\mathrm{yes}\}) = 1$
o_8	Yes	High	$m_8(\{\mathrm{yes}\}) = 1$

Table 6. The second reduct.

U	*Headache*	*Temperature*	*Flu*
o_1	Yes	Very high	$m_1(\{\text{yes}\}) = 1$
o_2	Yes	High	$m_2(\{\text{no}\}) = 1$
o_3	Yes	Normal	$m_3(\{\text{yes}\}) = 0.5$, $m_3(\Theta) = 0.5$
o_4	No	Normal	$m_4(\{\text{no}\}) = 0.6$, $m_4(\Theta) = 0.4$
o_5	No	Normal	$m_5(\{\text{no}\}) = 1$
o_6	Yes	High	$m_6(\{\text{no}\}) = 1$
o_7	No	Very high	$m_7(\{\text{yes}\}) = 1$
o_8	No	High	$m_8(\{\text{yes}\}) = 1$

Example. We take the first solution in Table 5, the belief decision rules generated become more simple and shorter:

If Muscle-pain = yes and Temperature = very high Then $m_{1,7}(\{\text{yes}\}) = 1$.
If Muscle-pain = no and Temperature = high Then $m_{2,6}(\{\text{no}\}) = 1$.
If Muscle-pain = yes and Temperature = normal Then $m_{3,4,5}(\{\text{yes}\}) = 0.6$ $m_{3,4,5}(\Theta) = 0.4$.
If Muscle-pain = yes and Temperature = high Then $m_8(\{\text{yes}\}) = 1$.

Where $m_{2,6}$ is the combined bba of m_2 and m_6 using the operator mean of combination. The same thing for the combined bba $m_{3,4,5}$.

We take the second solution in Table 6, the belief decision rules generated are as follows:

If Headache = yes and Temperature = very high Then $m_1(\{\text{yes}\}) = 1$.
If Headache = yes and Temperature = high Then $m_{2,6}(\{\text{no}\}) = 1$.
If Headache = yes and Temperature = normal Then $m_3(\{\text{yes}\}) = 0.5$ $m_3(\Theta) = 0.5$.
If Headache = no and Temperature = normal Then $m_{4,5}(\{\text{no}\}) = 0.7$ $m_{4,5}(\Theta) = 0.3$.
If Headache = no and Temperature = very high Then $m_7(\{\text{yes}\}) = 1$.
If Headache = no and Temperature = high Then $m_8(\{\text{yes}\}) = 1$.

We can conclude that the first solution gives better number of combined belief decision rules than the second solution. These sets of belief decision rules could be used to classify unseen objects.

4.3 *Adaptation of one heuristic method for attribute selection*

In Zhong *et al.* (2001), a heuristic approach is presented based on rough set theory. This algorithm uses the attributes from the core (those attributes that cannot be removed without introducing inconsistencies) as an initial attribute subset. Next, it selects attributes one by one from unselected ones using some strategies and adds them to the attribute subset until a reduct approximation is obtained. A threshold value is required as a stopping criterion to determine when a reduct candidate is 'near enough' to being a reduct. On each iteration, those objects that are consistent with the current reduct candidate are removed (an optimisation can speed up the algorithm on each iteration as it removes objects that are already in the positive region). In this subsection, we will adapt this heuristic method to extract one reduct from the uncertain decision system using the redefined basic concepts of rough sets described in the previous subsection.

The following notations are used to introduce the algorithm:
Notations:

- U: the set of objects (instances),
- C: the set of all conditional features,
- ud: the uncertain decision feature,
- R: the set of selected condition attributes,
- P: the set of unselected condition attributes,
- ε: reduct threshold.

Algorithm:

Initial state
(1) R $\leftarrow$ calculateCore()
$P = C - Core(C)$
(2) **while** $(\gamma_R(\{ud\}) < \varepsilon)$
(3) U $\leftarrow$ $U - UPos_R(\{ud\})$//optimisation
(4) $\forall\, a \in P$
(5) $v_a = card(UPos_{R\cup\{a\}}(\{ud\}))$
(6) $m_a = max\text{-}size(UPos_{R\cup\{a\}}(\{ud\})/R \cup \{a\} \cup \{ud\}))$
(7) $x_a = max\text{-}size(U/R \cup \{a\})$
(8) Choose a with largest $v_a{*}m_a{*}x_a$
(9) $R \leftarrow R \cup \{a\}$
(10) P = P − {a}
(11) **return** R

The strategy for attribute selection used in this algorithm can be described as follows: select a given attribute a, if by adding it to the subset R of attributes, the card($U\,\mathrm{Pos}_{R\cup\{a\}}(\{\mathrm{ud}\})$) increases faster and the max-size($U\,\mathrm{Pos}_{R\cup\{a\}}(\{\mathrm{ud}\})/R \cup \{a\} \cup \{\mathrm{ud}\}$)) and max-size($U/R \cup \{a\}$) are larger than by adding any other attribute. The discussed conditions can be competitive. So, we choose in our quality criterion the result of multiplication of the three values.

Example. Let us continue with the same example in Table 1 to compute the relative reduct using the heuristic.

We begin by computing the relative core (the set of indispensable condition attributes) as follows:

- Remove the attribute Headache from the condition attributes:

$$U\,\mathrm{Pos}_{\{\text{Muscle-pain,Temperature}\}}(\{\mathrm{ud}\}) = \{o_1, o_2, o_6, o_7, o_8\} = U\,\mathrm{Pos}_C(\{\mathrm{ud}\}).$$

  So, the attribute *Headache* is not indispensable.
- Remove the attribute Temperature from the condition attributes:

$$U\,\mathrm{Pos}_{\{\text{Headache,Muscle-pain}\}}(\{\mathrm{ud}\})\{o_8\} \neq U\,\mathrm{Pos}_C(\{\mathrm{ud}\}).$$

  So, the attribute Temperature is indispensable.

- Remove the attribute Muscle-pain from the condition attributes:

$$U\,\mathrm{Pos}_{\{\mathrm{Headache,Temperature}\}}(\{\mathrm{ud}\}) = \{o_1, o_2, o_6, o_7, o_8\} = U\,\mathrm{Pos}_C(\{\mathrm{ud}\}).$$

So, the attribute Muscle-pain is not indispensable.

Only the attribute Temperature is ud-indispensable. So, it is the relative core.

We have, in the initial state:

$$R = \mathrm{Core}(C) = \{\mathrm{Temperature}\} \quad \text{and} \quad P = C - \mathrm{Core}(C) = \{\text{Muscle-pain}, \mathrm{Headache}\}.$$

The {Temperature}-positive region of {ud}: $U\,\mathrm{Pos}_{\{\mathrm{Temperature}\}}(\{\mathrm{ud}\}) = \{o_1, o_7\}$.

Setting reduct threshold: $\varepsilon = \gamma_C(\{\mathrm{ud}\}) = 5/8$, the termination condition will be $\gamma_R(\{\mathrm{ud}\}) \geq 5/8$. Since $\gamma_R(\{\mathrm{ud}\}) = 2/8 < 5/8$, R is not a reduct, and we must continue adding other condition attributes to R until a reduct is obtained.

From U, the consistent objects $\{o_1, o_7\}$ should be removed. The initial state is shown in Table 7 with $U = \{o_2, o_3, o_4, o_5, o_6, o_8\}$.

Next, we have two candidates Muscle-pain and Headache. Tables 8 and 9 give the results of adding Muscle-pain and Headache to R, respectively.

From Tables 8 and 9, we obtain the following positive regions:

Table 7. Initial state.

U	*Temperature*	*Flu*
o_2	High	$m_2(\{\mathrm{no}\}) = 1$
o_3	Normal	$m_3(\{\mathrm{yes}\}) = 0.5$, $m_3(\Theta) = 0.5$
o_4	Normal	$m_4(\{\mathrm{no}\}) = 0.6$, $m_4(\Theta) = 0.4$
o_5	Normal	$m_5(\{\mathrm{no}\}) = 1$
o_6	High	$m_6(\{\mathrm{no}\}) = 1$
o_8	High	$m_8(\{\mathrm{yes}\}) = 1$

Table 8. Selecting Muscle-pain.

U	*Muscle-pain*	*Temperature*	*Flu*
o_2	No	High	$m_2(\{\mathrm{no}\}) = 1$
o_3	Yes	Normal	$m_3(\{\mathrm{yes}\}) = 0.5$, $m_3(\Theta) = 0.5$
o_4	Yes	Normal	$m_4(\{\mathrm{no}\}) = 0.6$, $m_4(\Theta) = 0.4$
o_5	Yes	Normal	$m_5(\{\mathrm{no}\}) = 1$
o_6	No	High	$m_6(\{\mathrm{no}\}) = 1$
o_8	Yes	High	$m_8(\{\mathrm{yes}\}) = 1$

Table 9. Selecting Headache.

U	*Headache*	*Temperature*	*Flu*
o_2	Yes	High	$m_2(\{\mathrm{no}\}) = 1$
o_3	Yes	Normal	$m_3(\{\mathrm{yes}\}) = 0.5$, $m_3(\Theta) = 0.5$
o_4	No	Normal	$m_4(\{\mathrm{no}\}) = 0.6$, $m_4(\Theta) = 0.4$
o_5	No	Normal	$m_5(\{\mathrm{no}\}) = 1$
o_6	Yes	High	$m_6(\{\mathrm{no}\}) = 1$
o_8	No	High	$m_8(\{\mathrm{yes}\}) = 1$

$U\,\mathrm{Pos}_{\{\text{Muscle-pain,Temperature}\}}(\{\mathrm{ud}\}) = \{o_2, o_6, o_8\}$,
$U\,\mathrm{Pos}_{\{\text{Headache,Temperature}\}}(\{\mathrm{ud}\}) = \{o_2, o_6, o_8\}$,
$v_{\text{Muscle-pain}} = |U\,\mathrm{Pos}_{\{\text{Muscle-pain,Temperature}\}}(\{\mathrm{ud}\})| = 3$,
$v_{\text{Headache}} = |U\,\mathrm{Pos}_{\{\text{Headache,Temperature}\}}(\{\mathrm{ud}\})| = 3$.

The two candidates have the same v_a. So, we should check the value of m_a

$$U\,\mathrm{Pos}_{\{\text{Muscle-pain,Temperature}\}}(\{\mathrm{ud}\})/\{\text{Muscle-pain, Temperature, Flu}\} = \{\{o_2, o_6\}, \{o_8\}\},$$

$$U\,\mathrm{Pos}_{\{\text{Headache,Temperature}\}}(\{\mathrm{ud}\})/\{\text{Headache, Temperature, Flu}\} = \{\{o_2, o_6\}, \{o_8\},$$

$$m_{\text{Muscle-pain}} = 2,$$

$$m_{\text{Headache}} = 2,$$

The two candidates have the same m_a. So, we should check the value of x_a

$$U/\{\text{Muscle-pain, Temperature}\} = \{\{o_2, o_6\}, \{o_3, o_4, o_5\}, \{o_8\}\},$$

$$U/\{\text{Headache, Temperature}\} = \{\{o_2, o_6\}, \{o_3\}, \{o_4, o_5\}, \{o_8\}\},$$

$$x_{\text{Muscle-pain}} = 3,$$

$$x_{\text{Headache}} = 2.$$

One can see that by selecting the attribute Muscle-pain or Headache, we can reduce the number of contradictory instances. Since the maximal set is in $U/\{$Muscle-pain, Temperature$\}$, then, according to our selection strategies, Muscle-pain should be selected first. After adding Muscle-pain to R, $\gamma_R(\{\mathrm{ud}\}) = 3/6 \geq 5/8$. The process is finished. Thus, the selected attribute subset is {Muscle-pain, Temperature}.

5. Experimental results

In our experiments, we have performed several tests on real databases obtained from the UCI repository.[1] A brief description of these databases is presented in Table 10. These databases are modified in order to include uncertainty in decision attribute. We take 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 $m(\Theta) = P$. The different results carried out from these tests will be presented and analysed in order to evaluate our proposed heuristic attribute selection method for certain and uncertain cases. The PCC representing the percentage of correct classification of the objects belonging to testing set is a relevant criterion to judge the performance of classifying unseen objects using the learned and simplified belief decision

Table 10. Description of databases.

Database	*No. of instances*	*No. of attributes*	*No. of decision values*
Wisconsin breast cancer database	690	8	2
Balance scale database	625	4	3
Congressional voting records database	497	16	2
Zoo database	101	17	7
Nursery database	12,960	8	3

Table 11. Experimental results using exhaustive search.

Database	PCC certain case (%)	PCC low unc (%)	PCC middle unc (%)	PCC high unc (%)	Times complexity (s)
Wisconsin breast cancer database	77.78	74.12	72.36	70.23	110
Balance scale database	65.7	62.3	60.9	54.34	91
Congressional voting records database	86.52	83.74	80.21	79.53	85
Zoo database	79.37	78.21	74.61	72.13	110
Nursery database	64.9	61.56	60.08	53.14	293

Table 12. Experimental results using heuristic search.

Database	PCC certain case (%)	PCC low unc (%)	PCC middle unc (%)	PCC high unc (%)	Times complexity (s)
Wisconsin breast cancer database	77.61	72.2	71.53	69.93	41
Balance scale database	65.7	62.3	60.9	54.34	35
Congressional voting records database	86.11	83.45	79.23	74.66	27
Zoo database	77.78	74.12	72.36	70.23	110
Nursery database	63.9	60.34	59.7	52.3	127

rules. Besides to PCC, we take the time complexity as a second criterion to compare the exhaustive and heuristic algorithms of finding reduct.

Each database is divided into 10 parts. Nine parts are used as the training set, the last is used as the testing set. The procedure is repeated 10 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.

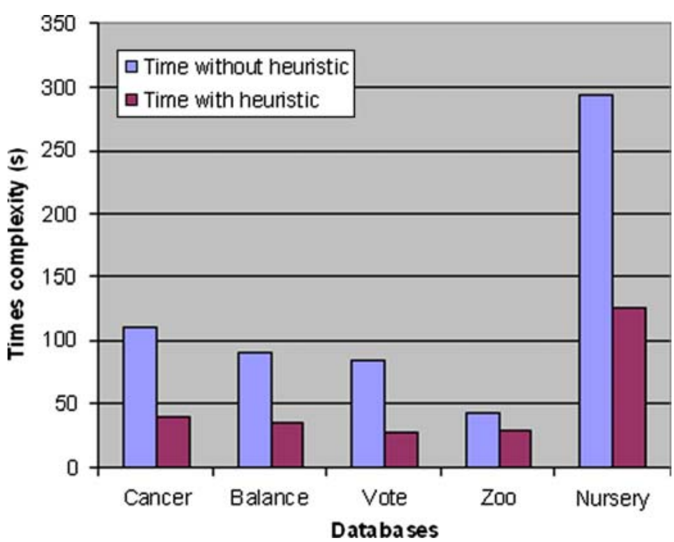

Figure 1. Time complexity for the different databases.

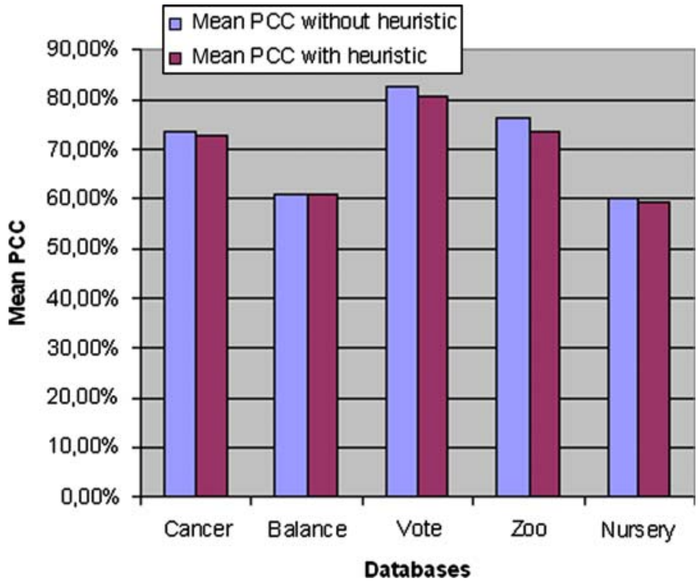

Figure 2. The mean PCC for the different databases.

Tables 11 and 12 summarise different results relative to these databases by applying the two algorithms of finding the reduct using complete and heuristic search. We start by the time complexity criterion; it is significantly decreased by applying our attribute selection heuristic method using rough sets. For instance, the time complexity goes from 110 to 41s for Wisconsin breast cancer database (Figure 1).

For the second criterion of evaluation, the two algorithms have a better PCC in the certain case and also in all degrees of uncertainty for all databases. However, when the uncertainty increases, the accuracy is lightly decreased. When we compare the PCC between the two algorithms, we find that finding the reduct without the heuristic gives slightly better PCCs than with the heuristic search. The reason is that the heuristic does not guarantee the optimal reduct. However, the difference in PCC is very small. For example, the PCC goes from 86.52 to 86.11% for the certain case of Congressional voting records database, whereas the PCC for the Balance scale database is the same for the two algorithms due to the small number of attributes (Figure 2).

6. Conclusion and future work

In this paper, we have adapted the basic concepts of rough sets in an uncertain context, and then we have proposed a heuristic method for feature selection in order to extract the more relevant condition attributes from partially uncertain data without costly calculation. We handle uncertainty in decision attributes (classes) under the belief function theory as understood by the TBM. With this simplification, the belief decision rules generated will be more significant for the classification process. Experiments on different databases show interesting results especially on the time complexity which is significantly depleted using our heuristic. As a future work, we will handle the uncertainty in condition attribute values.

Acknowledgements

The author would like to thank Richard Jensen for his help and for the interesting discussions with him during the writing of this paper. Many thanks are also due to anonymous referees for their detailed comments, criticisms and suggestions that have helped greatly in improving the quality of this paper.

Note

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

Notes on contributors

Salsabil Trabelsi is a PhD student at the Tunis Higher Institute of Management and a member of Larodec Laboratory. Her PhD is related to classification under uncertainty based on rough sets. Her research interests are classification problems, decision making, belief function, rough sets, decision trees etc.

Zied Elouedi is an Associate Professor in Computer Science at the Tunis Higher Institute of Management and a member of the Larodec Laboratory. He received his PhD in 2002. His research papers have been published in refereed journals such as the International Journal of Approximate Reasoning, Fuzzy Sets and Systems, IEEE Transactions on Systems Man and Cybernetics – Part B, etc. His research topics are related to uncertainty theories, machine learning, datamining, bioinformatics, intrusion detection, etc.

References

Ben Hariz, S., Elouedi, Z., and Mellouli, K., 2006. Clustering approach using belief function theory. *In*: *12th international conference*. AIMSA 2006, Varna, Bulgaria, 12–15 September, LNAI 4183 Berlin/Heidelberg: Springer, 162–171.

Chouchoulas, A. and Shen, Q., 2001. Rough set-aided keyword reduction for text categorisation. *Journal of applied artificial intelligence*, 15 (9), 843–873.

Denoeux, T., 1995. A *k*-nearest neighbor classification rule based on Dempster–Shafer theory. *IEEE transactions on systems, man and cybernetics*, 25 (5), 804–813.

Denoeux, T., 2000. A neural network classifier based on Dempster–Shafer theory. *IEEE transactions on systems, man and cybernetics, part A*, 30 (2), 131–150.

Elouedi, Z., Mellouli, K., and Smets, P., 2001. Belief decision trees: theoretical foundations. *International journal of approximate journal*, 28 (2–3), 91–124.

Grzymala-Busse, J.W., 2003. 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, FL, 5663.

Jensen, R. and Shen, Q., 2003. Finding rough set reducts with ant colony optimization. In: *Proceedings of the 2003 UK workshop on computational intelligence*, 15–22.

Kryszkiewiez, M., 1995. Rough set approach to incomplete information systems. *In*: *Proceedings of the second annual joint conference on information science*, Wrightsville, NC, 194–197.

Modrzejewski, M., 1993. Feature selection using rough sets theory. *In*: *Proceedings of the 11th international conference on Machine Learning*, Madison, WI, pp. 213–226.

Orlowska, E., 1994. Reasoning with incomplete information: rough set based information systems. *In*: *Incompleteness and uncertainty in information systems*. New York, NY: Springer-Verlag, 16–33.

Pawlak, Z., 1982. Rough sets. *International journal of computer and information sciences*, 11 (5), 341–356.

Pawlak, Z., 1991. *Rough sets: theoretical aspects of reasoning about data*. Dordrecht: Kluwer Academic Publishing.

Shafer, G., 1976. *A mathematical theory of evidence*. Princeton, NJ: Princeton University Press.

Skowron, A. and Rauszer, C., 1992. The discernibility matrices and functions in information systems. *In*: R. Slowinski, ed. *Intelligent decision support*. Boston, MA: Kluwer Academic Publishers, 331–362.

Slowinski, R. and Stefanowski, J., 1989. Rough classification in incomplete information systems. *Mathematical and computer modelling*, 12 (10–11), 1347–1357.

Smets, P., 1994. The transferable belief model. *Artificial intelligence*, 66 (2), 191–236.

Smets, P., 1998a. The transferable belief model for quantified belief representation. *In*: D.M. Gabbay and P. Smets, eds. *Handbook of defeasible reasoning and uncertainty management systems*. Vol 1. Doordrecht: Kluwer, 207–301.

Smets, P., 1998b. Application of belief transferable belief model to diagnostic problems. *International journal of intelligent systems*, 13 (2–3), 127–157.

Trabelsi, S. and Elouedi, Z., 2008. Learning decision rules from uncertain data using rough sets. *In*: *The 8th international FLINS conference on computational intelligence in decision and control*. Madrid, Spain, 21–24 September, Singapore: World Scientific, 114–119.

Trabelsi, S. and Elouedi, Z., 2009. Attribute selection from partially uncertain data using rough sets. *The third international conference on modeling simulation, and applied optimization*, UAE, January 20–22.

Wroblewski, J., 1995. Finding minimal reducts using genetic algorithms. *In*: *Proceedings of the 2nd annual joint conference on information sciences*. Wrightsville Beach, NC 186–189.

Zhong, N., Dong, J.Z., and Ohsuga, S., 2001. Using rough sets with heuristics for feature selection. *Journal of intelligent information systems*, 16 (3), 199–214.