Feature Selection From Partially Uncertain Data Within the Belief Function Framework

Asma Trabelsi[1], Zied Elouedi[1], and Eric Lefevre[2]

[1] Université de Tunis, Institut Supérieur de Gestion de Tunis, LARODEC , Tunisia
[2] Univ. Artois, EA 3926, Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A), Béthune, F-62400, France

Abstract. With the rapid growth of high dimensional data, feature selection has become a substantial task for several machine learning problems. In fact, it is regarded as an important process for classification performance owing to its ability to remove redundant and inconsistent features. The rough set theory is regarded as a well known tool allowing relevant feature selection. As the task of attribute selection using rough sets is an NP-hard problem, several heuristic algorithms have been introduced. The Johnson's algorithm, handling data characterized by certain and precise attribute values, is one of the most known ones. In this paper, we propose to extend this latter algorithm to an uncertain context, precisely where data contain uncertain condition attribute values represented within the belief function framework. We test the performance of our belief Johnson's algorithm through several experiments on synthetic databases.

Keywords: Classification, feature selection, rough set theory, heuristic algorithms, belief function theory.

1 Introduction

Classification is a substantial problem in the area of machine learning and data mining and it has become increasingly challenging owing to the exponential data growth in both simple size and dimensionality [21]. Dimensionality reduction techniques such as feature selection are widely used to deal with high-dimensionality [15, 20]. In fact, they allow to exclude as much as possible irrelevant and redundant attributes from the ordinal set of attributes for the purpose of reducing the computational cost and the dimensionality space of huge data sets as well to improve the classification accuracy. Mainly, there exist two feature selection approaches: wrapper and filter. These former methods incorporate classification algorithms to search and select a subset of attributes, while these latter methods select a subset of attributes independently of any classification algorithm. As the classification of each attribute or subsets of attributes is costly in term of computation time, we resort, in this investigation, to filter approaches. Rough set theory is one of the most popular filter methods allowing to find out the minimal set of relevant attributes also called reduct [14]. The main advantages of the reduct is its ability to predict the decision concepts as well as the whole set of attributes. Basically, finding the set of all reducts or finding the optimal reduct is regarded as an NP-hard problem which has led to the introduction of several heuristic approaches such as the QuickReduct algorithm [3], the Johnson's algorithm [8], etc. In this paper, we propose to adopt this latter heuristic algorithm thanks to its capacity to discover only one reduct with the minimal number of attributes generally close to the optimal from a given data. It is substantial to note that Johnson's algorithm handles only the case of perfect data. However, real world databases may be susceptible to imprecision, incompleteness and uncertainty. Such cases require to adopt the concept of rough sets heuristic algorithms to an uncertain environment. In several domains, uncertainty may exist either in decision attributes or in condition attributes or in both decision and condition attributes. For instance, in medicine, patients' symptoms (condition attributes) or patients' diseases (decision attribute) can be uncertain. Therefore, in this paper, we propose to adopt the Johnson's heuristic algorithm to an uncertain environment. Several theories have been discussed in the literature to handle uncertainty such as the bayesian theory [1], the fuzzy theory [6], the belief function theory [5], etc. As this latter formalism has the advantage to deal with partial or even total ignorance, we propose a belief Johnson algorithm to find reducts from a partially uncertain decision table. More precisely, we tackle the problem where uncertainty exists only in the condition attributes. This paper is organized as follows. Section 2 is dedicated to hightailing the basic concepts of the rough set theory. We detail Johnson's algorithm in Section 3. Section 4 provides an overview of the fundamental concepts of the belief function theory. Our novel approach for feature selection based on the belief Johnson's algorithms has been presented in Section 5. Section 6 describes the experimental results yielded from several uncertain databases under the classifier fusion framework, in order to evaluate the performance of our novel approach. In Section 7, we draw conclusion and we highlight some future works.

2 Rough Set Theory

Rough Sets (RS), introduced by Pawlak [14], is a valid mathematical tool for dealing with imperfect knowledge (vague, imprecise and uncertain) in variety of applications related to machine learning area which mainly includes the problems of knowledge discovery, clustering [13], classification [7, 10], feature selection [2, 11], etc. This latter paradigm consists of extracting the smallest subsets of relevant features, also called reducts, from the original set of features of a given data. In a practical point of view, information and knowledge are represented by a decision information system which is defined as a pair $A = (U, R)$, where $U = \{O_1, \dots, O_n\}$ is a non-empty, finite set of objects called the universe and $R = C \cup D$ is a finite set of attribute, $C = \{c_1, \dots, c_K\}$ is a non-empty, finite set of K condition attributes, v_{c_k} is a non-empty set of values of $c_k \in C$, $D = \{\mathrm{d}\}$ is the decision attribute set and v_d is the decision attribute value [23]. Given a subset of attributes $B \subseteq C$, an indiscernibility relation, denoted $IND(B)$, is defined as follows $\forall\ k = \{1, \dots, K\}$:

$$IND(B) = \{(O_i, O_j) \in U \times U | \forall c_k \in B, v_{c_k}(O_i) = v_{c_k}(O_j)\} \qquad (1)$$

Let $B \subseteq C$ and $X \subseteq U$. We can approximate X by using only the information contained by constructing the B-lower and B-upper approximations of X, denoted respectively by $\underline{B}(X)$ and $\overline{B}(X)$ and defined by:

$$\underline{B}(X) = \{O_j | [O_j]_B \subseteq X\} \tag{2}$$

and

$$\overline{B}(X) = \{O_j | [O_j]_B \cap X = \emptyset\} \tag{3}$$

where

$$[O_j]_B = \{O_i | \forall c_k \in B, v_{c_k}(O_i) = v_{c_k}(O_j)\} \tag{4}$$

Let B and D be equivalence relations over U, then the positive region can be set as:

$$Pos_B(D) = \bigcup_{x \in U/D} \underline{B}(X) \tag{5}$$

The positive region embraces all objects of U that can be classified to classes of U/D by the use of the information in attributes B. Keeping only attributes that preserve the positive region is regarded as a practical way for feature reduction. It is noteworthy that there exist several subsets of condition attributes and those which are minimal are called reducts. A subset $B \subseteq C$ is a reduct of C with respect to the decision attribute d, if B is minimal and:

$$Pos_B(d) = Pos_C(d) \tag{6}$$

In other terms, the attributes that do not belong to any reduct are unnecessary for the classification of the universe elements. Authors in [17], have introduced the notation of discernibility matrix and function as other ways for finding reducts for a decision table DT. The discernibility matrix of DT, denoted by M, is a $|U| \times |U|$ matrix, in which the element $M(O_i, O_j)$ for an object pair (O_i, O_j) is defined by:

$$M(O_i, O_j) = \{v_c \in C | v_c(O_i) \neq v_c(O_j) \text{ and } v_d(O_i) \neq v_d(O_j)\} \ \forall\ i, j = \{1, \ldots, n\}$$

The matrix element $M(O_i, O_j)$ represents the set of all condition attributes discerning objects O_i and O_j that do not have the same value of the decision attribute d. The notion of discernibility function can be defined from the discernibility matrix as follows:

$$f(M) = \wedge\{\vee(M(O_i, O_j)) | \forall O_i, O_j \in U, M(O_i, O_j) = \emptyset\} \tag{7}$$

Reducts may be yielded by transforming the discernibility function from conjunctive normal form into disjunctive normal form. The major shortcoming of this solution is its costly operation which makes it impractical for medium sized or large sized data sets. Therefore, several heuristic algorithms have been discussed to overcome this drawback. The Johnson's heuristic algorithm [8] is one of the most known ones.

3 Johnson's heuristic algorithm

Johnson's algorithm proposed in [8] is an heuristic algorithm that uses a greedy search technique which consists of picking out attributes having the most frequency appearing in the discernibility matrix. Algorithm 1 below underlines the main steps of the Johnson algorithm.

Algorithm 1 Johnson's Algorithm(U,C ∪ d)

```
input:U: a finite set of instances, C: a set of conditional attributes, d: a set of
decision attributes
Output: R:reduct, R ⊆ C
R ← ∅
M ← DiscernibilityMatrix (U,C ∪ d)
repeat
   c ← SelectAttributeWithMaxWeight(M)
   R ← R ∪ {c}
   for i=1 to |U| do
      for j=1 to |U| do
         if c ∈ M(O_i, O_j) then
            M(O_i, O_j) = ∅
         end if
      end for
   end for
until (M(O_i, O_j) =∅ ∀ i,j)
```

Johnson's algorithm begins by setting the reduct candidate, denoted by R, to an emptyset. Subsequently, it computes the number of occurrences of each attribute in the discernibility matrix. The attribute that has the highest count of appearances will be added to R and all cells containing this attribute will be removed from the discernibility matrix. This process should be repeated until all non empty cells are removed. Then, the algorithm returns R as a final reduct. Though Johnson's algorithm guarantees to uncover a single reduct, it is unuseful in the case where data sets are characterized by uncertain attributes. Thus, we propose to extend this algorithm to an uncertain context, more particulary to the context of the belief function theory.

4 Belief function theory

The belief function theory, also known as Dempster-Shafer Theory (DST) or theory of evidence [16], is considered as a useful theory for representing and managing uncertain knowledge. In what follows, we briefly introduce the main concepts of the belief function theory as interpreted in the Transferable belief Model (TBM) [19].

Let Θ be a finite non-empty set of N elementary events related to a given problem, these events are assumed to be exhaustive and mutually exclusive.

Such Θ is called the frame of discernment. The power set of Θ, denoted by 2^{Θ}, is composed of all the subsets of Θ.

The impact of evidence assigned to each subsets of the frame of discernment Θ is named basic belief assignment (bba). It is defined as:

$$m : 2^{\Theta} \rightarrow [0, 1]$$
$$\sum_{A \subseteq \Theta} m(A) = 1 \tag{8}$$

The amount $m(A)$, known as basic belief mass (bbm), expresses the degree of belief committed exactly to the event A.

To make decision within the belief function framework, we must transform the bba into a probability measure called pignistic probability denoted $BetP$ and defined as follows [18]:

$$BetP(A) = \sum_{B \subseteq \Theta} \frac{|A \cap B|}{|B|} \frac{m(B)}{1 - m(\emptyset)} \ \forall \ A \in \Theta \tag{9}$$

5 Belief Johnson's algorithm for partially uncertain data

This Section is devoted to describing our heuristic approach for feature selection form partially uncertain decision table. Our proposed solution, namely belief Johnson's algorithm, aims to extract the subset of relevant attributes which enables the same classification ability as the entire set of attributes. In what follows, we provide firstly a brief description of a partially uncertain decision table under the belief function framework and then we detail our heuristic approach.

5.1 Partially uncertain decision table

Our partially uncertain decision table will be defined as a pair UDT=(U, $uC \cup d$) where U is a finite set of n objects U=$\{O_1$,. . .,$O_n\}$ described by a set of K uncertain condition attributes denoted by uC=$\{c_1, \ldots, c_K\}$ and a certain decision attribute denoted by $\{d\}$. In this work, we suggest to represent the uncertainty of each condition attribute within the belief function framework. Thus, a basic belief assignment $m_i^{\Theta_k}$, defined on the frame of discernment Θ_k which represents all possible values of a condition attribute $c_k \in uC$, will be assigned to each condition attribute value v_{c_k} of an instance O_i. These bbas can be induced by one or several agents and they may express the case of total certainty ($m_i^{\Theta_k}(\{v_{c_k}\}) = 1$ and $m_i^{\Theta_k}(\Theta_k) = 0$) or even the case of total ignorance ($m_i^{\Theta_k}(\{v_{c_k}\}) = 0$ and $m_i^{\Theta_k}(\Theta_k) = 1$).

Example: Let Table 1 be our uncertain decision table composed with eight instances characterized by three uncertain categorical condition attributes uC =

$\{Hair, Eye, Height\}$ and a certain decision attribute d with possible values $\{d_1,\ d_2\}$. To simplify the notations, we will use 1, 2 and 3 instead of *Hair*, *Eye* and *Height*. The basic belief assignments, which are randomly affected to the condition attribute values, will be defined on the frame of discernments $\Theta_1 = \{Blond, Dark\}$, $\Theta_2 = \{Brown, Blue\}$ and $\Theta_3 = \{Short, Middle, Tall\}$.

Table 1. Uncertain decision table

	Hair	*Eye*	*Height*	d
O_1	$m_1^{\Theta_1}(\{Dark\})$=0.5 $m_1^{\Theta_1}(\Theta_1)$=0.5	$m_1^{\Theta_2}(\{Brown\})$=1 $m_1^{\Theta_2}(\Theta_2)$=0	$m_1^{\Theta_3}(\{Middle\})$=0.95 $m_1^{\Theta_3}(\Theta_3)$=0.05	d_1
O_2	$m_2^{\Theta_1}(\{Blond\})$=0.1 $m_2^{\Theta_1}(\Theta_1) = 0.9$	$m_2^{\Theta_2}(\{Blue\})$=0.82 $m_2^{\Theta_2}(\Theta_2)$=0.18	$m_2^{\Theta_3}(\{Middle\})$=1 $m_2^{\Theta_3}(\Theta_3)$=0	d_1
O_3	$m_3^{\Theta_1}(\{Blond\})$=0.6 $m_3^{\Theta_1}(\Theta_1) = 0.4$	$m_3^{\Theta_2}(\{Brown\})$=0.2 $m_3^{\Theta_2}(\Theta_2)$=0.8	$m_3^{\Theta_3}(\{Tall\})$=0.55 $m_3^{\Theta_3}(\Theta_3)$=0.45	d_2
O_4	$m_4^{\Theta_1}(\{Dark\})$=0.7 $m_4^{\Theta_1}(\Theta_1) = 0.3$	$m_4^{\Theta_2}(\{Brwon\})$=0 $m_4^{\Theta_2}(\Theta_2)$=1	$m_4^{\Theta_3}(\{Short\})$=1 $m_4^{\Theta_3}(\Theta_3)$=0	d_1
O_5	$m_5^{\Theta_1}(\{Blond\})$=1 $m_5^{\Theta_1}(\Theta_1) = 0$	$m_5^{\Theta_2}(\{Blue\})$=0.18 $m_5^{\Theta_2}(\Theta_2)$=0.82	$m_5^{\Theta_3}(\{Middle\})$=0.15 $m_5^{\Theta_3}(\Theta_3)$=0.85	d_2
O_6	$m_6^{\Theta_1}(\{Blond\})$=0.3 $m_6^{\Theta_1}(\Theta_1) = 0.7$	$m_6^{\Theta_2}(\{Brwon\})$=0.13 $m_6^{\Theta_2}(\Theta_2)$=0.87	$m_6^{\Theta_3}(\{Tall\})$=0.8 $m_6^{\Theta_3}(\Theta_3)$=0.2	d_2
O_7	$m_7^{\Theta_1}(\{Dark\})$=1 $m_7^{\Theta_1}(\Theta_1) = 0$	$m_7^{\Theta_2}(\{Brown\})$=0.8 $m_7^{\Theta_2}(\Theta_2)$=0.2	$m_7^{\Theta_3}(\{Tall\})$=0.25 $m_7^{\Theta_3}(\Theta_3)$=0.75	d_1
O_8	$m_8^{\Theta_1}(\{Dark\})$=0.5 $m_8^{\Theta_1}(\Theta_1) = 0.5$	$m_8^{\Theta_2}(\{Blue\})$=0.22 $m_8^{\Theta_2}(\Theta_2)$=0.78	$m_8^{\Theta_3}(\{Short\})$=0.1 $m_8^{\Theta_3}(\Theta_3)$=0.9	d_1

5.2 Reducts for partially uncertain decision table

Let us remind that the reduct, using Johnson's algorithm, is constructed by sequentially adding the most discernable attributes for a given decision attribute. Therefore, the computation of the discernibility matrix M will be a preliminary step in Johnson's algorithm. However, computing M from partially uncertain decision table UDT=(U, $uC \cup d$) remains really a challenging task which has not attracted great attention yet. To cope with this problem, we propose to adopt Johnson's heuristic algorithm to an uncertain environment, precisely to the belief function framework. Our belief Jonson's algorithm tackles mainly the problem where the uncertainty exists only in the condition attributes and represented within the framework of belief functions. In such cases, dissimilarity metrics must be used to discern all pairs of objects with different decision values. Consequently, entries of the discernibility matrix should be set as follows $\forall\ i, j \in \{1, \ldots, n\}$ and $k \in \{1, \ldots, K\}$:

$$M'(O_i, O_j) = \{c_k \in C | dist(m_i^{\Theta_k}, m_j^{\Theta_k}) > S \text{ and } v_d(O_i) \neq v_d(O_j)\} \qquad (10)$$

where S denotes a tolerance threshold and $dist$ corresponds to a distance measure between two bbas. Different distance metrics have been investigated in the literature such as the Tessems distance [22], the Euclidean distance [4], the Jousselme distance [9], etc. This latter is one of the most commonly used distances.

Given two bbas m_1 and m_2, the Jousselme distance measure is computed as follows:

$$dist(m_1, m_2) = \sqrt{\frac{1}{2}(m_1 - m_2)^T D(m_1 - m_2)} \quad (11)$$

with D is the Jaccard index matrix, the elements of which are calculated as follows:

$$D(A, B) = \begin{cases} 1 & \text{if } A{=}B{=}\ \emptyset \\ \dfrac{|A \cap B|}{|A \cup B|} & \forall\ A{,}B \in 2^{\Theta} \end{cases} \quad (12)$$

Once the discernibility matrix is computed, the reduct will be incrementally composed by adding the condition attribute that occurs with the most frequency and then removing any cells contain this attribute. This procedure must be repeated until all non-empty cells will be eliminated.

Example: In order to extract the reduct relative to our partially uncertain decision table (see Table 1), we start by computing the discernibility matrix M' where the threshold S is setting to 0.1 (see Table 2). To simplify, we use the notations Ha, E and He respectively for $Hair$, Eye and $Height$. For instance, $M'(O_1, O_2){=}\emptyset$ due to the fact that $v_d(O_1) = v_d(O_2)$. Another example, $M'(O_1, O_5){=}\{Ha, E, He\}$ because $dist(m_1^{\Theta_1}, m_5^{\Theta_1}) = 0.5 > 0.1$, $dist(m_1^{\Theta_2}, m_5^{\Theta_2}) = 0.7185 > 0.1$, $dist(m_1^{\Theta_3}, m_5^{\Theta_3}) = 0.6532 > 0.1$ and $v_d(O_1) \neq v_d(O_5)$.

Table 2. Discernibility matrix M'

	O_1	O_2	O_3	O_4	O_5	O_6	O_7	O_8
O_1	-							
O_2	-	-						
O_3	E, He	Ha, E, He						
O_4	-	-	Ha, E, He					
O_5	Ha, E, He	Ha, E, He	-	Ha, E, He	-			
O_6	Ha, E, He	Ha, E	-	Ha, He	-	-		
O_7	-	-	Ha, E, He	-	Ha, E, He	Ha, E, He		
O_8	-	-	E, He	-	Ha, E, He	Ha, E, He	-	-

Let us now compute the reduct using our belief Johnson's algorithm. Firstly, we count the number of occurrences relative to each condition attribute and the feature with the highest frequency will be added to the reduct. In our discernibility matrix (Table 2), the attributes Eye and $Height$ appear 14 times, while the attribute $Hair$ appears 13 times. As attributes Eye and $Height$ have equal weights, we randomly add one among them to the reduct R. If the attribute Eye is chosen then we remove all cells containing Eye from M' and the next best feature will be selected. By removing Eye, we still have $Hair$ and $Height$ with weights equal to 1. As $Hair$ and $Height$ have equal weights, we add either

$Hair$ or $Height$ to R and then we remove the chosen attribute for M' and the: if we remove the attribute $Hair$, R will be set to $R = \{Eye, Hair\}$ and M' will be empty. By against, if we remove the attribute $Height$, R will be equal to $R = \{Eye, Height\}$ and M' will be empty.

6 Experimentations

In order to evaluate the performance of our heuristic feature selection approach, we propose to carry out several experimental tests on real world databases obtained from the U.C.I. repository [12]. Table 3 gives a brief description of the databases where #Instances and #Attributes denote respectively the total number of instances and the total number of condition attributes.

Table 3. Description of databases

Databases	**#Instances**	**#Attributes**
Tic-Tac-Toe	958	9
SPECT Heart	267	22
Lymphography	148	18
Wine	178	13
Zoo	101	17

As all these databases do not contain uncertain condition attributes represented within the belief function framework, we propose to generate synthetic databases by taking into account the original database D and a degree of uncertainty P to transform actual condition attribute value v_{c_k} of each object O_i, where $c_k \in uC$, into a basic belief assignment as follows:

$$\begin{aligned} m_i^{\Theta_k}(\{v_{c_k}\}) &= 1 - P \\ m_i^{\Theta_k}(\Theta_k) &= P \end{aligned} \tag{13}$$

The degree of uncertainty P takes value in the interval [0,1]: Certain Case (P=0), Low Uncertainty ($0 \leq P < 0.4$), Middle Uncertainty ($0.4 \leq P < 0.7$) and High Uncertainty ($0.7 \leq P \leq 1$).

To check the validity of our proposed heuristic approach, we try to perform an empirical comparison in terms of dimensionality space and classification accuracy criterion (PCC) between results yielded by our initial databases and those obtained by our belief Johnson's algorithm in both certain and uncertain cases. In order to compare PCCs, we resort to three well known classification algorithms, namely the Decision Tree classifier (DT), the Naive Bayes classifier (NB) and the k-Nearest Neighbor classifier (k-NN) with k equals to 1. As these classification algorithms cannot handle data characterized by uncertain condition attributes represented within the framework of belief functions, we perform the pignistic transformation, using Equation 9, to make decision about condition attribute values which should be chosen. Once computing the pignistic probability of all condition attribute beliefs, we run the three mentioned classifiers using the leave one out cross validation approach which divides a data set with N instances into N-1 instances for training and the remaining instance for testing. This procedure will be repeated N times where each existing instance is used once as a test set. Experimental results are given from Table 4 to Table 6 where $\#F$ denotes the number of selected attributes. Note that, for the sake of simplification, we have replaced the attribute names in the reduct by numbers according to their order in the databases.

Table 4. Classification accuracy (%) without dimensionality reduction

Databases	NB	DT	1-NN
Tic-Tac-Toe	82.04	69.41	99.16
SPECT Heart	84.64	79.40	82.39
Lymphography	79.05	83.78	82.43
Wine	92.69	98.31	98.87
Zoo	92.07	95.04	96.03

Table 5. Belief Johnson's algorithm: certain case

Databases	Reduct	#F	PCC (%)		
			NB	DT	1-NN
Tic-Tac-Toe	R={1, 2, 3, 4, 5, 6, 8, 9}	8	80.58	71.71	81.41
SPECT Heart	R={1, 3, 4, 5, 6, 7, 8, 9, 10, 13, 14, 16, 19, 20, 21, 22}	16	79.40	79.77	80.52
Lymphography	R={2, 13, 14, 15, 16, 18}	6	77.02	80.40	81.75
Wine	R={1, 5, 7, 11, 13}	5	97.19	98.87	87.87
Zoo	R={4, 7, 9, 11, 14}	5	94.05	90.09	96.03

We remind that our ultimate objective is to reduce dimensionality space as well as the computational time and keep or increase the classification accuracy. Let us note that in certain case our belief Johnson's algorithm gives exactly the same results as the original Johnson's algorithm. From the results given in Tables 4, 5 and 6, we have deduced that in both certain and uncertain cases, our belief Johnson's algorithm allows a significant dimensionality reduction. For instance, applying our belief Johnson's algorithm to the Lymphography database containing 18 condition attributes we obtain 6 selected features in certain case, while applying this proposed algorithm to synthetic Lymphography database we obtain 5 selected features for both low and middle uncertainty cases and 6 selected feature for high uncertainty case.

Table 6. Belief Johnson's algorithm: uncertain case

Databases	Low					Middle					High				
	Reduct	#F	PCC (%)			Reduct	#F	PCC (%)			Reduct	#F	PCC (%)		
			NB	DT	1-NN			NB	DT	1-NN			NB	DT	1-NN
Tic-Tac-Toe	R={1, 2, 3, 4, 5, 6, 7, 9}	8	82.56	73.48	92.90	R={1, 2, 3, 5, 7, 8, 9}	7	78.81	69.83	89.53	R={1, 2, 3, 4, 5, 6, 7, 8, 9}	9	82.04	69.41	99.16
SPECT Heart	R={1, 3, 4, 5, 6, 7, 8, 9, 10, 13, 14, 15, 20, 22 }	14	82.77	79.77	80.52	R={1, 3, 7, 8, 9, 10, 11, 13, 14, 16, 17, 20, 21, 22 }	14	84.26	80.14	81.64	R={1, 5, 6, 7, 8, 9, 10, 13, 16, 17, 20, 22 }	12	83.52	81.64	82.02
Lymphography	R={2, 13, 14, 15, 18}	5	75.67	76.35	79.72	R={2, 13, 14, 15, 18}	5	75.67	76.35	79.72	R={2, 13, 14, 15, 16, 18}	6	79.05	78.37	70.05
Wine	R={1, 7, 10, 13}	4	93.25	98.31	97.19	R={5, 7, 11, 13}	4	96.62	95.50	94.38	R={1, 2, 3, 4, 7, 10, 11, 13}	8	97.75	99.43	98.97
Zoo	R={2, 4, 7, 9, 14}	5	93.06	87.12	95.04	R={4, 5, 7, 9, 14}	5	94.05	85.14	95.04	R={2, 5, 6, 7, 8, 9, 13, 14}	8	95.04	91.08	91.08

In terms of the PCC criterion, we emphasize that for our certain case feature reduction allows the improvement of the PCC criterion compared to those yielded by initial databases, though not always. However, the PCCs yielded following to the feature reduction process are often close to those obtained with the initial databases. For example, for the initial Spect-Heart database, we have 84.64%, 79.40% and 82.39% as PCCs relative to respectively DT, NB and 1-NN classifiers, while applying our belief Johnson's algorithm in certain case, we get 79.40%, 79.77% and 80.52% as PCCs relative to respectively DT, NB and 1-NN classifiers. Consequently, we can admit that feature reduction allows not only to reduce dimensionality space and computational time, but also to provide significant classification accuracies and thus, it is worth applying belief Johnson's algorithm to partially uncertain databases. Concretely, we have tackled three levels of uncertainty: Low, Middle and High. From Table 6, we can deduce that the Decision Tree, the Naive Bayes and the 1-Nearest Neighbors classifiers have yielded interesting PCC values for the different synthetic databases obtained by using the three levels of uncertainty. For instance, for Wine database, we have gotten 93.25%, 98.31% and 97.19% as PCCs obtained respectively by the DT, the NB and the 1-NN classifiers in low uncertainty case, for the middle uncertainty case, we have obtained 96.62%, 95.50% and 94.38% as PCCs relative respectively to the DT, the NB and the 1-NN classifiers. Also, we have reported 97.75%, 99.43% and 98.97% as PCCs obtained respectively by the DT, the NB and the 1-NN classifiers in high uncertainty case.

7 Conclusion

In this paper, we have proposed a new heuristic approach for relevant feature selection from partially uncertain decision table, precisely where uncertainty exists only in the condition attributes and represented within the belief function framework. Our experimental tests have shown the efficiency of our proposed method in terms of dimensionality reduction and classification accuracy. However, it is important to note that our heuristic method does not give optimal reduct. So, as a future work, we look forward to improving our proposed method by allowing the optimal reduct. We further intend to introduce uncertainty in both condition and decision attributes.

References

1. J. M. Bernardo and A. F. Smith. *Bayesian theory*, volume 405. John Wiley & Sons, 2009.
2. R. B. Bhatt and M. Gopal. On fuzzy-rough sets approach to feature selection. *Pattern recognition letters*, 26(7):965–975, 2005.
3. A. Chouchoulas and Q. Shen. Rough set-aided keyword reduction for text categorization. *Applied Artificial Intelligence*, 15(9):843–873, 2001.
4. F. Cuzzolin. A geometric approach to the theory of evidence. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 38(4):522–534, 2008.
5. T. Denœux. Reasoning with imprecise belief structures. *International Journal of Approximate Reasoning*, 20(1):79–111, 1999.

6. T. Haag and M. Hanss. Comprehensive modeling of uncertain systems using fuzzy set theory. *Nondeterministic Mechanics*, pages 193–226, 2012.
7. R. Jensen and C. Cornelis. Fuzzy-rough nearest neighbour classification and prediction. *Theoretical Computer Science*, 412(42):5871–5884, 2011.
8. D. S. Johnson. Approximation algorithms for combinatorial problems. In *The 5th annual ACM symposium on Theory of computing*, pages 38–49. ACM, 1973.
9. A. Jousselme, D. Grenier, and E. Bossé. A new distance between two bodies of evidence. *Information fusion*, 2(2):91–101, 2001.
10. L. Khoo, S. Tor, and L. Zhai. A rough-set-based approach for classification and rule induction. *International Journal of Advanced Manufacturing Technology*, 15(6):438–444, 1999.
11. P. Kumar, P. Vadakkepat, and L. A. Poh. Fuzzy-rough discriminative feature selection and classification algorithm, with application to microarray and image datasets. *Applied Soft Computing*, 11(4):3429–3440, 2011.
12. P. Murphy and D. Aha. UCI repository databases. *http://www.ics.uci.edu/mlearn*, 1996.
13. D. Parmar, T. Wu, and J. Blackhurst. MMR: an algorithm for clustering categorical data using rough set theory. *Data & Knowledge Engineering*, 63(3):879–893, 2007.
14. Z. Pawlak. Rough sets. *International Journal of Computer & Information Sciences*, 11(5):341–356, 1982.
15. N. J. Pizzi and W. Pedrycz. Aggregating multiple classification results using fuzzy integration and stochastic feature selection. *International journal of approximate reasoning*, 51(8):883–894, 2010.
16. G. Shafer. *A mathematical theory of evidence*, volume 1. Princeton university press Princeton, 1976.
17. A. Skowron and C. Rauszer. The discernibility matrices and functions in information systems. In *Intelligent Decision Support*, pages 331–362. Springer, 1992.
18. P. Smets. The transferable belief model for quantified belief representation. *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, 1:267–301, 1998.
19. P. Smets and R. Kennes. The transferable belief model. *Artificial intelligence*, 66(2):191–234, 1994.
20. C. O. S. Sorzano, J. Vargas, and A. P. Montano. A survey of dimensionality reduction techniques. *arXiv preprint arXiv:1403.2877*.
21. C. Sun, N. Rampalli, F. Yang, and A. Doan. Chimera: Large-scale classification using machine learning, rules, and crowdsourcing. *Proceedings of the VLDB Endowment*, 7(13):1529–1540, 2014.
22. B. Tessem. Approximations for efficient computation in the theory of evidence. *Artificial Intelligence*, 61(2):315–329, 1993.
23. Y. Yao and Y. Zhao. Discernibility matrix simplification for constructing attribute reducts. *Information sciences*, 179(7):867–882, 2009.