# Classification with Belief Decision Trees

Zied Elouedi[1], Khaled Mellouli[1], and Philippe Smets[2]

[1] Institut Supérieur de Gestion de Tunis, 41 Avenue de la liberté, cité Bouchoucha, 2000 Le Bardo, Tunis, Tunisia
{zied.elouedi@isg.rnu.tn}, {khaled.mellouli@ihec.rnu.tn}
[2] IRIDIA, Université Libre de Bruxelles, 50 av., F. Roosvelt, CP194/6, 1050 Brussels, Belgium
{psmets@ulb.ac.be}

**Abstract.** Decision trees are considered as an efficient technique to express classification knowledge and to use it. However, their most standard algorithms do not deal with uncertainty, especially the cognitive one. In this paper, we develop a method to adapt the decision tree technique to the case where the object's classes are not exactly known, and where the uncertainty about the class' value is represented by a belief function. The adaptation concerns both the construction of the tree and its use to classify new objects characterized by uncertain attribute values.

## 1 Introduction

Decision trees are among the well known machine learning techniques. They are widely used in a variety of fields notably in artificial intelligence applications. Their success is explained by their ability to handle complex problems by providing an understandable representation easier to interpret and also their adaptability to the inference task by producing logical rules of classification.

Several methods [1] [5] [7] have been proposed to construct decision trees. These algorithms have as inputs the training set composed by instances where each one is described by the set of attribute values and its assigned class. The output is a decision tree ensuring the classification of new instances.

A major problem faced in the standard decision tree algorithms results from the uncertainty encountered in the data. This uncertainty can appear either in the construction or in the classification phase. Ignoring it can affect the efficiency of the obtained results.

In order to overcome this drawback, probabilistic decision trees have been developed by Quinlan [6]. This kind of trees presents small extensions over the standard one and its use remains limited since it only deals with statistical uncertainty induced by information arisen from random behavior.

The objective of this paper is to develop what we call a belief decision tree, a classification method adapting the decision tree approach to uncertain data, where the uncertainty is represented by belief functions as defined in the Transferable Belief Model (TBM). The choice of the TBM seems appropriate as it provides a convenient framework [2] for dealing with limited and uncertain information, notably those given by experts.

This paper is organized as follows: section2 provides a brief description of standard decision tree algorithms. In section3, the basics of the belief function theory are recalled. Our approach regarding a belief decision tree is described in section4. Both the construction and classification procedures will be detailed. Finally, an example explaining these two procedures is proposed in section5.

## 2 Basics of Decision Tree Algorithms

Several algorithms have been developed for learning decision trees [1] [5] [7]. In the artificial intelligence community, the most used is based on the TDIDT[1] approach. In that approach, the tree is constructed by employing a recursive divide and conquer strategy. Its steps can be defined as follows:

- By using *an attribute selection measure*, an attribute will be chosen in order to partition the training set in an *"optimal"* manner.
- Based on *a partitioning strategy*, the current training set will be divided into training subsets by taking into account the values of the selected attribute.
- When *the stopping criterion* is satisfied, the training subset will be declared as a leaf.

In the literature many attribute selection measures are proposed in [3] [5] [7]. Among the most used, we mention the information gain used within the ID3 algorithm [5]. The information gain of an attribute A relative to a set of objects S measures the effectiveness of A in classifying the training data. It is defined as follows:

$$Gain(S, A) = Info(S) - Info_A(S) \text{ where}$$

$$Info(S) = -\sum_{i=1}^{n} p_i.log_2 p_i \text{ and } Info_A(S) = \sum_{v \in Domain(A)} \frac{|S_v^A|}{|S|}.Info(S_v^A)$$

where $p_i$ is the proportion of objects in S belonging to the class $C_i$ (i = 1..n) and $S_v^A$ is the subset of objects for which the attribute A has the value v.

Although, it has shown good results, this measure has a serious limitation. It favors attributes with large number of values over those with few number of values [7]. To overcome this shortcoming, Quinlan [5] [7] suggests another selection attribute measure called the gain ratio and defined by:

$$Gain \text{ ratio}(S, A) = \frac{Gain(S,A)}{Split\ Info(A)} \text{ where}$$

$$Split\ Info(A) = \sum_{v \in Domain(A)} \frac{|S_v^A|}{|S|}.log_2 \frac{|S_v^A|}{|S|}$$

Split Info(A), measures the information content of the attribute A itself [5]. The gain ratio is the information gain calibrated by Split Info. Note that when the ratio is not defined, this criterion selects attributes among those with an average or better information gain [5].

[1] Top-Down Induction of Decision Tree

Once constructed, the decision tree is used to classify new objects. For a new instance, we start with the root, we evaluate the relative test attribute and we take the branch corresponding to the test's outcome. This process is repeated until a leaf is encountered. The new object belongs to the class labeling the leaf.

## 3 Belief Function Theory

In this section, we briefly review the main concepts underlying the theory of belief functions [8] [10] [11].

### 3.1 Definitions

Let $\Theta$ be a finite set of elementary events called frame of discernment. The basic belief assignment (bba) is a function m: $2^{\Theta} \rightarrow [0,1]$ such that $\sum_{A \subseteq \Theta} m(A) = 1$.

The value m(A) represents the part of belief supporting exactly that the actual event belongs to A and nothing more specific. The subsets A in $\Theta$ such that m(A) $> 0$ are called focal elements.

Associated with m is the belief function [10] defined for A $\subseteq \Theta$ as: $bel(A) = \sum_{\emptyset \neq B \subseteq \Theta} m(B)$. The degree of belief bel(A) given to a subset A of the frame $\Theta$ is defined as the sum of all the masses given to subsets that support A.

The representation of total ignorance is nicely achieved in the belief function theory. It is represented by the so-called vacuous belief function [8], i.e., the belief function which bba satisfies m($\Theta$) $= 1$ and m(A) $= 0$ for all A $\neq \Theta$.

### 3.2 Rules of Combination

Let $m_1$ and $m_2$ be two basic belief assignments induced from two distinct pieces of evidence. These bbas can be combined either conjunctively or disjunctively.

1. *The Conjunctive Rule:* When we know that both sources of information are fully reliable then the bba representing the combined evidence satisfies [12]:

   $$(m_1 \wedge m_2)(A) = \sum_{B,C \subseteq \Theta : B \cap C = A} m_1(B).m_2(C) \text{ for } A \subseteq \Theta$$

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 [12]:

   $$(m_1 \vee m_2)(A) = \sum_{B,C \subseteq \Theta : B \cup C = A} m_1(B).m_2(C) \text{ for } A \subseteq \Theta$$

### 3.3 Vacuous Extension of Belief Functions

Let X and Y be two sets of variables such that Y $\subseteq$ X. Let $m^Y$ be a bba defined on the domain $\Theta_Y$ of Y. The extension of $m^Y$ to $\Theta_X$, denoted $m^{Y \uparrow X}$ means that the information in $m^Y$ is extended to a larger frame X [4]:

$$m^{Y \uparrow X}(A \times \Theta_{X-Y}) = m^Y(A) \text{ for A} \subseteq \Theta_Y$$

$$m^{Y \uparrow X}(B) = 0 \text{ if B is not in the form } A \times \Theta_{X-Y}$$

### 3.4 Pignistic Transformation

The decision making problem is solved in the TBM framework by using the pignistic probability function defined and fully explianed by [10]:

$$BetP(\theta) = \sum_{A \subseteq \Theta, \theta \in A} \frac{m(A)}{|A|.(1-m(\emptyset))}, \text{ for all } \theta \in \Theta$$

It is the only transformation between belief functions and probability functions that satisfies some natural rationality requirements. The major one is described as follows: Suppose two contexts $C_1$ and $C_2$, suppose your beliefs in context $C_i$ is represented by $m_i$ and that the choice of the context obeys to some random process, with $P(C_1) = p$ and $P(C_2) = q$ with $p + q = 1$. Let $\Gamma$ denotes the operator that transforms a bba into a probability function. We want that it satisfies:

$$\Gamma(p\ m_1 + q\ m_2) = p\ \Gamma(m_1)\ +\ q\ \Gamma(m_2).$$

This translates the property that transforming the belief held before knowing the context that will be selected is the same as combining the conditional probability functions one would have obtained if the context had been known. Full details can be found in [10]. The probability function so obtained is then used to compute the expected utilities needed for optimal decision making.

## 4 Belief Decision Tree

In this section, we define the structure of the decision tree within the belief function framework, called belief decision tree then we present the notations that will be used in this paper. Next, we develop the two major procedures of a decision tree: the construction and the classification procedures.

### 4.1 Decision Tree Structure in the Belief Function Context

Any decision tree is constructed from a training set of objects based on successive refinements. Due to the uncertainty, the structure of the training set may be different from the traditional one. In fact, we assume that the uncertainty is lying only on classes of training instances. That is, our training set is composed by objects where the value of each attribute is known with certainty, whereas there is some uncertainty regarding its corresponding class.

We propose to associate for each training instance $\mathrm{I}_j$, $j = 1..p$, a bba, denoted $\mathrm{m}^{\Theta}\{I_j\}$, defined on the set of the possible classes $\Theta$ to which the object $\mathrm{I}_j$ can belong, and representing the beliefs given by an expert (or several experts) on the actual class of the object $\mathrm{I}_j$. This representation is also appropriate to describe the classical case where the object's class is exactly known.

Once the structure of the training set is defined, our belief decision tree is composed by the same elements as in the traditional tree. However, due to the uncertainty in training instances' classes, the structure of the leaves will change. Instead of assigning a unique class to each leaf, it will be labeled by a bba expressing a belief about the actual class of the objects belonging to the leaf.

### 4.2 Notations and Assumptions

In this paper, we use the following notations:

- S: a given set of objects,
- $I_j$: an instance (object, case, example),
- $\mathbf{A} = \{A_1, A_2...A_k\}$: a set of k attributes,
- $D(A_i)$: the domain of the attribute $A_i \in \mathbf{A}$,
- $A(I_j)$: the value of the attribute A for the object $I_j$,
- $S_v^A = \{I_j : A(I_j) = v\}$: the subset of objects which value for attribute A $\in$ $\mathbf{A}$ is v $\in D(A_i)$
- $\Theta = \{C_1, C_2, ..., C_n\}$: the frame of discernment involving the possible classes related to the classification problem.
- $C(I_j)$: the actual class of the object $I_j$,
- $m_g^{\Theta}\{I_j\}[A](C)$ denotes the conditional bba given to C $\subseteq \Theta$ relative to object $I_j$ given by an agent g that accepts that A is true. Useless indices are omitted.

### 4.3 Procedure for Constructing a Belief Decision Tree

As mentioned the algorithm to construct a decision tree, also called the induction task, is based on three major parameters: the attribute selection measure, the partitioning strategy, the stopping criterion. These parameters must take into account the uncertainty encountered in the training set.

**Attribute Selection Measure.** Our attribute selection measure has to take into account the bba of each object in the training set. The idea is to adapt the gain ratio proposed by Quinlan [7] to this uncertain context.

In order to define the gain ratio measure of an attribute A over a set of objects S within the TBM framework, we propose the following steps:

1. For each object $I_j$ in S, we have a bba $m^{\Theta}\{I_j\}$ that represents our belief about the value of $C(I_j)$. Suppose we select randomly and with equi-probability one object in S. What can be said about $m^{\Theta}\{S\}$, the bba concerning the actual class of that object selected in S?
   $m^{\Theta}\{S\}$ is the average of the bbas taken over the objects in the subset S:
   $$m^{\Theta}\{S\}(C) = \frac{\sum_{I_j \in S} m^{\Theta}\{I_j\}(C)}{|S|} \text{ for C} \subseteq \Theta \quad (1)$$
2. Apply the pignistic transformation to $m^{\Theta}\{S\}$ to get the average probability $BetP^{\Theta}\{S\}$ on each singular class of this randomly selected instance.
3. Perform the same computation for each subset $S_v^A$, we get $BetP^{\Theta}\{S_v^A\}$ for v $\in D(A)$, A $\in \mathbf{A}$.
4. Compute Info(S) and $Info_A(S)$ as done initially by Quinlan, but using the pignistic probabilities. We get:
   $$Info(S) = -\sum_{i=1}^{n} BetP^{\Theta}\{S\}(C_i).log_2 BetP^{\Theta}\{S\}(C_i) \quad (2)$$

$$Info_A(S) = \sum_{v \in D(A)} \frac{|S_v^A|}{|S|} Info(S_v^A)$$

$$= - \sum_{v \in D(A)} \frac{|S_v^A|}{|S|} \sum_{i=1}^{n} BetP^{\Theta}\{S_v^A\}(C_i) log_2 BetP^{\Theta}\{S_v^A\}(C_i) \quad (3)$$

Once computed, we get the information gain provided by the attribute A in the set of objects S such that:

$$Gain(S, A) = Info(S) - Info_A(S) \quad (4)$$

5. Using the Split Info, compute the gain ratio relative to each attribute A:

$$Gain \text{ Ratio}(S, A) = \frac{Gain(S, A)}{Split \text{ Info}(A)} \quad (5)$$

In each decision node, the attribute having the highest gain ratio will be selected as the root of the corresponding decision tree.

**Partitioning Strategy.** For the selected attribute, assign a branch corresponding to each attribute value. Thus, we get several training subsets where each one is relative to one branch and regrouping objects having the same attribute value.

**Stopping Criterion.** It allows to stop the development of a path and to declare the treated training subset as a leaf. Three strategies are proposed:

1. There is no more attribute to test.
2. The treated training subset contains only one object.
3. The values of the gain ratio relative to the remaining attributes are equal or less than zero.

Once the stopping criterion is fulfilled, the current node is declared as a leaf characterized by a bba defined on $\Theta$. The leaf's bba is equal to the average bba taken over the objects belonging to the same leaf.

**Constructing Algorithm.** Our algorithm presents an extension of the ID3 algorithm to the uncertain context. It is composed by the following steps:

1. Create the root node of the decision tree including all the objects of the training set T.
2. Verify if this node satisfies or not the stopping criterion. If it is fulfilled, declare it as a leaf node and compute its corresponding bba.
3. Otherwise, look for the attribute having the highest gain ratio. This attribute will be designed as the root of the tree related to the whole training set T.
4. Divide the training set according to the partitioning strategy.
5. Create a root node relative to each training subset.
6. For each node created, repeat the same process from the step 2.

If the bbas over the classes for every instance in the training set are described by a certain bba, i.e., there is no uncertainty about the actual class for all the objects in the training set, then we get the same results as the ID3 algorithm of Quinlan [7] based on the gain ratio.

### 4.4 Procedure of Classifying New Instances

Once constructed, the belief decision tree will be used to ensure the classification of new instances in this uncertain framework. These instances may present some uncertainty regarding the value of one (or several) of its attributes. In fact, the uncertainty related to each attribute $A_i$ can be defined by a bba $m^{A_i}$ on the set $\Theta_{A_i}$ of all the possible values of the attribute. For those, where the value is known with certainty, it would correspond a certain bba having as a focal element only this value. Besides, if an attribute value is unknown, it would be expressed by a vacuous bba.

We have to find the bba expressing beliefs characterizing the different attributes' values of the new instance to classify. To ensure this objective, we have to apply the following steps:

1. Extend the different bbas $m^{A_i}$ to the global frame of attributes $\Theta_A$.
2. Combine the extended bbas $m^{A_i \uparrow A}$ by applying the conjunctive rule:

$$m^{\Theta_A} = \wedge_{i=1..k} m^{A_i \uparrow A} \tag{6}$$

$m^{\Theta_A}$ represents beliefs on the combinations of the attributes of the given instance. We then consider individually the focal elements of this bba . Let x be such a focal element. The next phase is to compute the belief functions $\text{bel}^{\Theta}[x]$.

1. If the treated focal element x is a singleton (only one value for each attribute), then $\text{bel}^{\Theta}[x]$ is equal to the average belief function corresponding to the leaf to which this focal element is attached.
2. If the focal element x is not a singleton (some attributes have more than one value), then we have to explore all the possible paths relative to this combination of values. Two cases are possible:
   – If these paths lead to one leaf, then $\text{bel}^{\Theta}[x]$ is equal to this leaf's bel.
   – If these paths lead to distinct leaves, then $\text{bel}^{\Theta}[x]$ is equal to the result of the combination of each leaf's bel by applying the disjunctive rule.

Finally the belief functions computed with each focal element x are averaged [9] using the $m^{\Theta_A}$:

$$bel^{\Theta}[m^{\Theta_A}](\theta) = \sum_{x \subseteq \Theta_A} m^{\Theta_A}(x).bel^{\Theta}[x](\theta) \text{ for } \theta \in \Theta \tag{7}$$

Note that we have to apply the pignistic transformation in order to take a decision on the class of the instance to classify.

## 5 Example

Let's illustrate our method by a simple example. Assume that a bank wants to develop a loan policy for its clients by taking into account a number of their attributes. Let T be a training set (see Table 1) composed of eight instances (clients) characterized by three symbolic attributes: - *Income* with possible values $\{no, low, average, high\}$,

- *Property* with possible values $\{less, greater\}$ that is to express if the property's value is less or greater than the loan expected by the client,

- *Unpaid-credit* (denoted by Unp-c) with possible values $\{yes, no\}$ in order to know if the client has another credit unpaid or not.

Three classes may be assigned to clients ($\Theta = \{C_1, C_2, C_3\}$): $C_1$ for whom the bank accepts to give the whole loan, $C_2$ for whom the bank accepts to give a part of the loan and $C_3$ for whom the bank refuses to give the loan.

**Table 1.** The training set T

| Income | Property | Unp-c | Class |
|---|---|---|---|
| High | Greater | Yes | $m^\Theta\{I_1\}(C_1) = 0.7; m^\Theta\{I_1\}(\Theta) = 0.3;$ |
| Average | Less | No | $m^\Theta\{I_2\}(C_2) = 0.5; m^\Theta\{I_2\}(C_1 \cup C_2) = 0.4; m^\Theta\{I_2\}(\Theta) = 0.1$ |
| High | Greater | Yes | $m^\Theta\{I_3\}(C_1) = 0.8; m^\Theta\{I_3\}(\Theta) = 0.2;$ |
| Average | Greater | Yes | $m^\Theta\{I_4\}(C_2) = 0.5; m^\Theta\{I_4\}(C_3) = 0.2; m^\Theta\{I_4\}(\Theta) = 0.3$ |
| Low | Less | Yes | $m^\Theta\{I_5\}(C_3) = 0.8; m^\Theta\{I_5\}(C_2 \cup C_3) = 0.1; m^\Theta\{I_5\}(\Theta) = 0.1$ |
| No | Less | Yes | $m^\Theta\{I_6\}(C_3) = 1;\ m^\Theta\{I_6\}(\Theta) = 0$ |
| High | Greater | No | $m^\Theta\{I_7\}(C_1) = 1; m^\Theta\{I_7\}(\Theta) = 0$ |
| Average | Less | Yes | $m^\Theta\{I_8\}(C_3) = 0.6; m^\Theta\{I_8\}(\Theta) = 0.4$ |

Contrary to the 'traditional' training set where it includes only instances which classes are known with certainty, this given training set T is characterized by uncertainty relative to some instances'classes and which is represented by bbas. The training set T offers a more generalized framework than the tradional one. Thanks to our belief decision tree algorithm, we are able to generate the corresponding tree by taking into account this uncertainty.

**Construction Procedure.** Let's now try to construct the induced belief decision tree relative to the training set T. The first step is to find the root of the decision tree. Hence, we have to compute the gain ratio relative to the three attributes by taking into account the uncertainty embedded in instances' classes.

Let's illustrate briefly the computation of the gain ratio relative to the property attribute. Let $m^\Theta\{T\}$ be the average bba relative to T, $m^\Theta\{T^{property}_{greater}\}$ and $m^\Theta\{T^{property}_{less}\}$ be the average bbas relative to the sets of objects in T having as a value of the property attribute respectively greater and less. These bbas are computed by using the equation (1), then their corresponding pignistic probabilities $BetP^\Theta\{T\}$, $\mathrm{BetP}^\Theta\{T^{property}_{greater}\}$ and $\mathrm{BetP}^\Theta\{T^{property}_{less}\}$ have to be calculated.

Once computed, we get Info(T) = 1.535; $\mathrm{Info}_{property}$ = 1.17 and Split Info (property) = 1. So Gain ratio(T, property) = 0.365; By applying the same process, we get Gain ratio(T, income) = 0.405; Gain ratio(T, unpaid-credit) = 0.214

The gain ratio criterion favors the income attribute since it presents the highest value. Thus, it will be chosen as the root of the decision tree and branches are created for each of its possible values (high, average, low, no).

The same steps of the algorithm will be applied recursively. The belief decision tree induced is represented by Fig. 1:

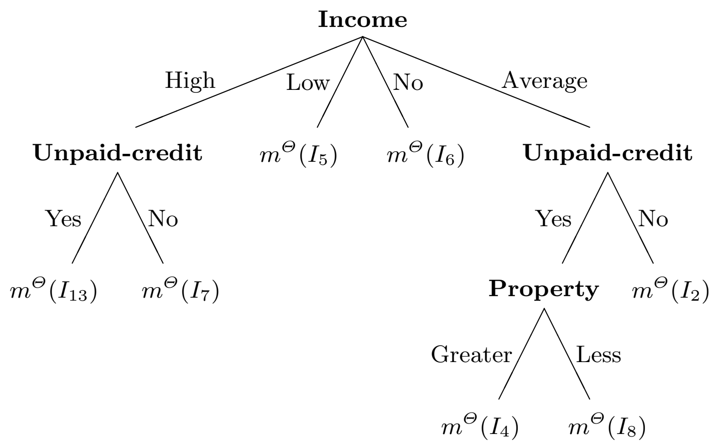


**Fig. 1.** The Final Belief Decision Tree.

Note that the leaf labeled by $\mathrm{m}^{\Theta}\{I_{13}\}$ is the average bba of the set involving the objects $\mathrm{I}_1$ and $\mathrm{I}_3$ defined as: $\mathrm{m}^{\Theta}\{I_{13}\}(C_1) = 0.75$; $\mathrm{m}^{\Theta}\{I_{13}\}(\Theta) = 0.25$;

**Classification Procedure.** Once the belief decision tree relative to the training set T is constructed (see Fig. 1), suppose that we would classify an instance characterized by certain and exact values for its income and unpaid-credit attributes which are respectively the values average and yes. However, there is some uncertainty in the value of the property attribute defined by: $\mathrm{m}^{property}(greater) = 0.4$; $\mathrm{m}^{property}(less) = 0.3$; $\mathrm{m}^{property}(\Theta_{property}) = 0.3$;

Once the attributes' bba are extended to $\Theta_A$ ($\Theta_A = \Theta_{income} \times \Theta_{property} \times \Theta_{unpaid-credit}$), we apply the conjunctive rule. We get a joint bba $m^{\Theta_A}$ on singular or subsets of instances such that: $m^{\Theta_A}(\{(average, greater, yes)\}) = 0.4$; $m^{\Theta_A}(\{(average, less, yes)\}) = 0.3$; $m^{\Theta_A}(\{average\} \times \Theta_{property} \times \{yes\}) = 0.3$;

Next, we have to find beliefs on classes (defined on $\Theta$) given the values of the attributes characterizing the new instance to classify. Three belief functions have to be defined where for each one, we take into account one focal element of $m^{\Theta_A}$. According to the induced belief decision tree (see Fig. 1), we get: $bel^{\Theta}[\{(average, greater, yes)\}] = bel_4$; $bel^{\Theta}[\{(average, less, yes)\}] = bel_8$; $bel^{\Theta}[\{average\} \times \Theta_{property} \times \{yes\}] = bel_4 \vee bel_8$.

Hence, these belief functions will be averaged then computing its corresponding BetP. As a result, we obtain that the new instance to classify has respectively 0.14, 0.38 and 0.48 as probability to belong to the classes $\mathrm{C}_1$, $\mathrm{C}_2$ and $\mathrm{C}_3$. So, it seems most probable to refuse the loan expected by this client.

As we note, our classification method using the induced belief decision tree is able to ensure the classification of new instances characterized by certain attribute values (like in the case of the standard decision tree). It has also the advantage (over the standard tree) to classify instances characterized by uncertain attribute values.

## 6 Conclusion

In this paper, we have developed a classification method providing a formal way to handle uncertainty in decision trees within the belief function framework. In fact, the construction procedure of the belief decision tree is ensured by taking into account the uncertainty about the actual classes of training objects. Then, we have proposed a classification procedure allowing to classify objects characterized by uncertain attributes. This method ensures the classification of instances with certain attributes or even those presenting some missing attribute values.

The major interest of the proposed method is that it can be applied to training sets where the instance classes are uncertain. Belief function theory offers a perfect representation of any form of uncertainty, from total knowledge to total ignorance, in particular more flexible than what probability theory can achieve. The most obvious case where belief decision trees will show their power is encountered where the instance classes are only known to belong to some subsets of the class domain.

## References


1. Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J.: Classification and regression trees. Belmont, CA: Wadsworth, (1984)
2. Elouedi, Z., Mellouli, K., Smets, P.: Decision trees using the belief function theory. To appear in the Proceedings of the International Conference on Information Processing and Management of Uncertainty IPMU'2000 (2000)
3. Lopez De Mantaras, R.: A distance-based attribute selection measure for decision tree induction. Machine learning **6** (1991) 81–92
4. Mellouli, K.: On the propagation of beliefs in network using the Dempster-Shafer theory of evidence. Ph.D dissertation University of Kansas Lawrence KS (1987)
5. Quinlan, J. R.: Induction of decision trees. Machine learning **1** (1986) 81–106
6. Quinlan, J. R.: Decision trees as probabilistic classifiers. Proceedings of the Fourth international Machine Learning (1987) 31–37
7. Quinlan, J. R.: C4.5: Programs for machine learning. Morgan Kaufmann San Mateo Ca (1993)
8. Shafer, G.: A mathematical theory of evidence. Princeton University Press (1976)
9. Smets, P.: Belief functions : the disjunctive rule of combination and the generalized bayesian theorem. International Journal of Approximate Reasoning **9** (1993) 1–35
10. Smets, P., Kennes, R.: The transferable belief model. Artificial Intelligence **66** (1994) 191–234
11. Smets, P.: The transferable belief model for quantified belief representation. D.M. Gabbay and Ph. Smets (eds.) Handbook of Defeasible Reasoning and Uncertainty Management Systems **1** Kluwer Doordrecht (1998) 267–301
12. Smets, P.: The Application of the transferable belief Model to Diagnostic Problems Int. J. Intelligent Systems **13** (1998) 127–158