Data Fusion in the Transferable Belief Model.

Philippe Smets
IRIDIA
Université Libre de Bruxelles
50 av. Roosevelt, CP 194-6, 1050 Bruxelles, Belgium
psmets@ulb.ac.be
http://iridia.ulb.ac.be/~psmets

Abstract - *When Shafer introduced his theory of evidence based on the use of belief functions, he proposed a rule to combine belief functions induced by distinct pieces of evidence. Since then, theoretical justifications of this so-called Dempster's rule of combination have been produced and the meaning of distinctness has been assessed. We will present practical applications where the fusion of uncertain data is well achieved by Dempster's rule of combination. It is essential that the meaning of the belief functions used to represent uncertainty be well fixed, as the adequacy of the rule depends strongly on a correct understanding of the context in which they are applied. Missing to distinguish between the upper and lower probabilities theory and the transferable belief model can lead to serious confusion, as Dempster's rule of combination is central in the transferable belief model whereas it hardly fits with the upper and lower probabilities theory.*

Keywords: belief function, transferable belief model, Dempster-Shafer theory, Dempster's combination rules.

1 Introduction.

People working in the fusion of uncertain data have been interested in the so-called Dempster-Shafer theory essentially because of two of its tools:

- a nice and flexible way to represent uncertainty, be it total ignorance or any form of partial or total knowledge, that is more general than what the probabilistic approach provides, and

- a rule to combine uncertain data, called the Dempster's rule of combination, that seems to provide an excellent tool for data-aggregation.

Nevertheless, confusion in the understanding of the exact nature of the model lead many authors to criticize it as inadequate, if not erroneous. These criticisms have permitted to clarify the exact nature of the models that might be regrouped under the Dempster-Shafer theory label.

We will try here to clarify what are these models represented by Dempster-Shafer theory, hoping that a correct understanding of their natures will lead to correct applications.

2 Belief functions based models.

Several models have been proposed that are based on the mathematical object called a belief function. We examine them, as confusing them is a major source of misdirected criticisms.

2.1 Various models

2.1.1 The belief functions as mathematical objects.

The common mathematical object encountered in every model that receive the 'belief function' or 'Dempster-Shafer theory' label is the belief function itself. As such it is nothing but a Choquet capacity monotone of infinite order. It means that a belief function, denoted bel, is a function from some power set on the $[0,1]$ interval. Let Ω be a finite space, and let 2^Ω be its power set. Then $bel : 2^\Omega \rightarrow [0,1]$ satisfies

$$bel(\emptyset) = 0$$
$$\forall n \geq 1, \forall A_1, A_2, ...A_n \subseteq \Omega$$
$$bel(A_1 \cup A_2 \cup ...A_n) \geq \sum_i bel(A_i)...$$
$$... - \sum_{i>j} bel(A_i \cap A_j).. - (-1)^n bel(A_1 \cap A_2 ...A_n) \quad (1)$$

As such these inequalities are hardly meaningful, but the special case with $n = 2$ and $A_1 \cap A_2 = \emptyset$ is worth considering:

$$bel(A_1 \cup A_2) \geq bel(A_1) + bel(A_2) \text{ if } A_1 \cap A_2 = \emptyset.$$

This last relation just illustrates that the belief given to the union of two disjoint subsets of Ω is larger or equal to the sum of the beliefs given to each subset individually.

When all the inequalities of relations (1) are replaced by equalities, the resulting function bel would then be a classical probability function.

A belief function can also be mathematically defined by introducing another set function, called the basic belief assignment (bba for short) and denoted m, with $m : 2^\Omega \rightarrow [0,1]$ and:

$$\sum_{A \subseteq \Omega} m(A) = 1. \quad (2)$$

Note. We do not require $m(\emptyset) = 0$ or equivalently that $bel(\Omega) = 1$, as initially required by Shafer. Our approach is more general, and the gain in some cases is even important, as illustrated in section 4.3.

Given a bba m, we can define bel so that

$$bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B), \ \forall A \subseteq \Omega. \quad (3)$$

The functions bel and m are in one-to-one correspondence, and the function bel as defined is indeed a belief function, i.e., satisfies relations (1).

Other functions have also been introduced, as often they are more convenient than bel or m. They are all in one-to-one correspondence with m, so they never add nor loose any information. They are: the plausibility function pl, the commonality function q and the the implicability function b, where for all $A \subseteq \Omega$:

$$pl(A) = bel(\Omega) - bel(\overline{A}) = \sum_{B: A\cap B \neq \emptyset} m(B). \quad (4)$$

$$q(A) = \sum_{B: A \subseteq B} m(B) \quad (5)$$

$$b(A) = bel(A) + m(\emptyset) = \sum_{B \subseteq A} m(B) \quad (6)$$

Up to now, we have only defined a mathematical function, called the belief function (and all its related functions). The next mathematical concept is Dempster's rule of combination.

2.1.2 Dempster's rule of combination.

Shafer introduces also a rule to combine two belief functions, called Dempster's rule of combination. It is an associative and commutative operation that maps a pair of belief functions defined both on the same space Ω into a new belief function on Ω. Let bel_1 and bel_2 be two belief functions on Ω, with m_1 and m_2 their related bba's. Then $bel_1 \cap bel_2$ is defined through its related bba $m_{1\cap 2}$ where:

$$m_{1\cap 2}(A) = \sum_{B\cap C = A} m_1(B) m_2(C), \ \forall A \subseteq \Omega \quad (7)$$

The same result can be conveniently expressed with the commonality function.

$$q_{1\cap 2}(A) = q_1(A) q_2(A) \ \forall A \subseteq \Omega \quad (8)$$

Usually the $\oplus$ symbol is used to denote Dempster's rule of combination, but for later symmetry we prefer the $\cap$ symbol.

A special, but essential, case of Dempster's rule of combination is the so called Dempster's rule of conditioning. Let m_A be so that $m_A(A) = 1$ and all other m_A values are null. The result of the combination of m with m_A produces a new belief function, denoted bel_A with:

$$bel_A(B) = bel(B \cup \overline{A}) - bel(\overline{A}). \quad (9)$$

The concepts of belief functions and Dempster's rules makes what is classically denoted by Dempster-Shafer theory. What are the meaning of these objects is another problem that authors took too fast for granted.

2.1.3 The representation of beliefs.

Shafer had the idea that

- bel represents a quantified belief about the exact value of the actual world, i.e., the strength of an agent's opinion about the value of the actual world,
- Dempster's rule of conditioning is the appropriate way to condition a belief function on a new piece of evidence that states that the actual world does not belong to some subset of worlds in Ω,
- Dempster's rule of combination is the appropriate way to combine the belief functions derived from two 'distinct' sources of evidence.

Of course the concept of 'belief' is not easy to define, and confusion appears because of the different understandings that can be given to this term. We can describe three major families of models where belief functions can appear, and where Dempster's rules may be appropriate. They are:

- the upper and lower probability models
- Dempster's model and the hint model of Kohlas and Monney
- the transferable belief model of Smets.

We examine successively these three models. But we introduce first a new notation that has proved to be very convenient in practical applications.

2.1.4 Notation

The full notation for bel and its related functions is:

$$bel^{\Omega,\Re}_{Y,t}[EC_{Y,t}](\omega_0 \in A) = x.$$

It denotes that the degree of belief held by the agent Y (shortcut for You) at time t that the actual world ω_0 belongs to the set A of worlds is equal to x, where A is a subset of the frame of discernment Ω and $A \in \Re$ where $\Re$ is a Boolean algebra of subsets of Ω. The belief is based on the evidential corpus $EC_{Y,t}$ held by Y at t, where $EC_{Y,t}$ represents all what agent Y knows at t.

Fortunately, in practice many indices can be omitted for simplicity sake. Usually $\Re$ is 2^Ω, the power set of Ω. When $\Re$ is not explicitly stated, it means that bel is defined on 2^Ω. '$\omega_0 \in A$' is denoted as 'A'. Y, t and/or Ω are omitted when the values of the missing elements are clearly defined from the context. So $bel^\Omega[E](A)$ or even $bel(A)$ are the most often used notation. Furthermore, E is usually just a conditioning event, and a subset of Ω. So the classical conditional probabilities $P(x|\theta)$ for $x \subseteq X$ and $P(.|\theta)$ are denoted here as $P^X[\theta](x)$ and $P^X[\theta]$ (indicating the domain avoids many errors).

In the above notation, bel can be replaced by any of m, pl, q, b, etc... The indices should made it clear what the links are.

2.2 The upper and lower probability models

In this context, one assumes

- either the existence of a probability function which represents the agent's beliefs but which values are not precisely known, and all that can be stated about the probability function is that it belongs to a family Π of probability functions,
- or that the state of belief of an agent is defined by such a family of probability functions.

Usually one will also assume that the family is convex, in which case its lower envelop completely defined the family. That lower envelop function of Π, denoted P_* is defined as

$$P_*(A) = \min_{P \in \Pi} P(A), \ \forall A \subseteq \Omega$$

This function is usually a Choquet capacity monotone of order two, and only in some special cases is it a belief function. The applicability of Dempster's rules is usually unjustified. Indeed, suppose You learn from one source that the unknown probability function belongs to a family Π_1, and from a second source that it belongs to a family Π_2, it is obvious that if You trust both sources of information, You would conclude that the unknown probability function belongs to $\Pi_1 \cap \Pi_2$. The lower envelop of this new family is usually not a belief function even if the lower envelop of Π_1 and of Π_2 were belief functions. In this example, Dempster's rules should not be applied.

Many criticisms against Dempster-Shafer theory were based on the use of Dempster's rules in such upper and lower probability contexts (Zadeh, 1986; Kyburg, 1987; Voorbracht, 1991).

As an example we just illustrate a case of conditioning were both Dempster's rule of conditioning and another conditioning rule are justified, but where these rules reflects different conditioning events. It just shows that a blind application of some 'universal' rule is non-sensical.

Example. Imprecise Database.(Smets, 1998a) Suppose a database with seven cases and one field containing the age (table 1). It happens that the ages are imprecisely known. The age can be interval valued like for case 1 whose age is in the interval [13-17], or disjunctive like for case 2 whose age is either 15 or 40. Because of the imprecision, some proportions cannot be exactly known. Depending on the values given to the actual ages of the seven cases, we can construct a family of functions giving the proportion of cases falling in every subset of the age domain $\Omega = [0, 100]$ and compatible with the data.

Suppose the interval [20, 30] of Ω. In order to compute the minimal and maximal proportions of cases that belong to [20, 30], we determine those cases that must belong to the interval whatever the values of their actual age, and those who might belong to it. In table 1, the columns Nec and Pos (columns 3 and 4) indicate those two groups of cases by a *. Cases 4 and 5 belong to [20, 30] whatever the actual ages of these two cases. Cases 4 to 7 might belong to it as it is possible that the actual ages of these cases are in the interval [20, 30].

Let these two proportions be called the upper and lower proportions, denoted Prop*([20-30]) and Prop_*([20-30]), respectively. We have:
$\text{Prop}_*([20\text{-}30]) = 2/7$, $\text{Prop}^*([20\text{-}30]) = 4/7$.

Suppose an individual will be selected from the database and the selection procedure is such that every individual has the same chance of being selected. Now we can speak of probabilities: they result from the selection procedure, and the probabilities happen to be equal to the proportions because of the equiprobability of being selected. What would be the probability that should be given to the fact that the age of the selected individual, denoted w_0, would fall in the interval [20-30]?

Due to the imprecision of the proportions that results from the imprecision in the data, we can only build the family of probability functions compatible with the available data. By construction, this family is in one-to-one correspondence to the family of proportion functions compatible with the data. Let Π denote the family of probability functions Prob defined over Ω and compatible with the data. For any subset of Ω, we can only determine the upper and lower probabilities, denoted Prob^* and Prob_*, respectively, by taking the extremes values given to that subset when Prob is constrained within Π. We have:
$\text{Prob}_*([20\text{-}30]) = \min_{\text{Prob} \in \Pi} \text{Prob}([20\text{-}30]) = 2/7$
$\text{Prob}^*([20\text{-}30]) = \max_{\text{Prob} \in \Pi} \text{Prob}([20\text{-}30]) = 4/7$

These probabilities are numerically equal to the corresponding proportions because of the equi-probability of the sampling procedure. If the chance of being selected had depended on the individual, this direct relation between the probabilities and the proportions would disappear.

Now suppose we receive the information that the selected individual happens to be younger than 25 years old. This is a *factual* (Dubois & Prade, 1998; Smets, 1998a) knowledge as it tells something about w_0 itself. Among other it tells that case 4 could not have been the one selected. Depending on the values given to the actual ages in agreement with the available data, we can compute the upper and lower values that Prob([20-30]||[0-25]) could achieve. This is done by taking every probability function Prob that belongs to Π, computing Prob([20-30]||[0-25]) through the application of the Bayesian conditioning rule, and finding the extreme values these conditional probabilities could reach. It can be shown that the conditional probability that the age of the randomly selected individual is between 20 and 30 becomes:

$$\text{Prob}_*([20-30]||[0-25]) = ...$$

$$... = \frac{\text{Prob}_*([20-25])}{\text{Prob}_*([20-25]) + \text{Prob}^*([0-19])} = 1/6$$

$$\text{Prob}^*([20-30]||[0-25]) =$$

Table 1: Value of the age of seven individuals.

				Case is < 25				All cases < 25			
		20-30		20-25	0-19	20-25	0-19			20-30	
Case	Age	Nec	Pos	Nec	Pos	Pos	Nec	Age	Incl	Nec	Pos
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	13-17				*		*	13-17	*		
2	15 or 40				*			15	*		
3	11 or 50				*			11	*		
4	27	*	*					-			
5	21-23	*	*	*	*	*	*	21-23	*	*	*
6	23 or 45		*			*	*	23	*	*	*
7	15-23		*		*	*	*	15-23	*		*
Total		2	4	1	5	3	4		6	2	3

$$... = \frac{\text{Prob}^*([20-25])}{\text{Prob}^*([20-25]) + \text{Prob}_*([0-19])} = 3/7.$$

Columns 5 to 8 of table 1 provide the details of the needed data to compute these upper and lower probabilities. This rule is called the 'natural extension' by Walley (1991).

Let us forget about the previous conditioning event, and suppose now You learn a piece of information that states that none of the individuals in the database were older than 25. This is a *generic* knowledge as it concerns all the individuals in the database and not only the selected one. In that case, the database is transformed into a new data base, as given in columns 9 to 12 of table 1. For every case, the 'age' is obtained by intersecting the previous subsets with the interval [0-25]. Case 2 was known to be either 15 or 40. As nobody was older than 25, we know now that case 2 is 15 years old, etc... For case 4, one might wonder why he was initially in the data base. Indeed the intersection is empty, he was known to be 27, and now we learn that everybody was younger that 25. There are two ways to handle that case. We can consider it as an error and eliminate it from the database, in which case only 6 cases are left over, all probabilities are normalized, and:

$\text{Prob}_*([20\text{-}30]|[0\text{-}25]) = 2/6$,
$\text{Prob}^*([20\text{-}30]|[0\text{-}25]) = 3/6$.

We can also decide to keep it as an indication of some incoherence in the database and not to normalize the data, keeping 1/7 as an amount of conflict, in which case:

$\text{Prob}_*([20\text{-}30]|[0\text{-}25]) = 2/7$,
$\text{Prob}^*([20\text{-}30]|[0\text{-}25]) = 3/7$.

This last conditioning corresponds to what would have been obtained by the application of Dempster's rule of conditioning on the Prob_* function.

Note that if the values of the database had been known precisely, not only the upper and lower probabilities would have been equal, but also the conditional probabilities that would result from both the factual and the generic knowledge would have been both equal to the conditional probabilities obtained by the application of the Bayesian rule of conditioning. This degenerescence explains why the distinction between generic and factual knowledge is not important in probability theory.

2.3 The hint model.

Initially, Dempster (1967, 1968, 1972) studied belief functions while trying to solve the problem of fiducial inference. He introduced a model characterized by three components:

- two finite spaces X and Y,
- a probability measure P defined on X,
- a one-to-many mapping $\Gamma : X \to 2^Y$.

The question to be solved is the computation of the probability on Y. Should we know all the conditional probabilities on Y given each $x \in X$, the problem is trivial. But it happens we do not know these conditional probability functions. All we know is they are zero when $y \notin \Gamma(x)$. If one consider all the possible values for these conditional probability functions, one can then deduce the family of probability functions on Y, which lower envelop happens to be a belief function . As such, Dempster's model is then a special case of upper and lower probability model. But it can also get another understanding as shown by the hint model developed by Monney and Kohlas (1995).

These authors assume Dempsters original structure (X, P, Γ, Y) where X and Y are two sets, P is a probability measure on X and Γ is a one-to-many mapping from X to Y. They assume a question, whose answer is unknown. The set y is the set of possible answers to the question. One and only one element of Y is the correct answer to the question. 'The goal is to make assertions about the answer in the light of the available information. We assume that this information allows for several different interpretations, depending on some unknown circumstances. These interpretations are regrouped into the set X and there is exactly one correct interpretation. Not all interpretations are equally likely and the known probability measure P on X reflects our information in that respect. Furthermore, if the interpretation $x \in X$ is the correct one, then the answer is known to be in the subset $\Gamma(x) \subseteq Y$. Such a structure $H = (X, P, \Gamma, Y)$ is called a hint... An interpretation $x \in X$ supports the hypothesis $H \subseteq Y$

if $\Gamma(x) \subseteq H$ because in that case the answer is necessarily in H. The degree of support of H is defined as the probability of all supporting interpretation of H' (Kohlas & Monney, 1995, page vi).

The hint theory corresponds to Dempsters original approach. Kohlas and Monney call their measure a degree of support, instead of belief, to avoid personal, subjective connotation, but degrees of support and degrees of belief are mathematically equivalent and conceptually very close. In the hint theory, the primitive concept is the hint from which degrees of supports are deduced, whereas the transferable belief modeland Shafer's initial approach (Shafer, 1976) both consider the degrees of belief as a primitive concept.

The approach of Cholvy (see these proceedings) can be seen as a special case of the hint model.

2.4 Probabilities on modal propositions.

Ruspini (1986, 1987) and Pearl (1988) have suggested that $bel(A)$ should be understood as the probability that A is known or proved, respectively. This extension of the domain of the probability functions from the propositional logic domain to the modal propositional logic domain leads indeed to belief functions. Let $bel(p) = Proba(p\ is\ known)$ or $bel(p) = Proba(p\ is\ proved)$, then bel, seen as a function on the propositions, is a belief function. The problem with this approach is that Dempster's rules need a justification these authors did not provide (see (Smets, 1991, 1993b)). So what they propose is quite limited and does not cover what is usually understood as Dempster-Shafer theory as it lacks the concept of conditioning and combination that really makes what is called Dempster-Shafer theory.

3 The transferable belief model.

3.1 Scope.

The TBM (for transferable belief model) provides a model for the representation of quantified beliefs (Smets & Kennes, 1994; Smets, 1998b). One assumes that there are several possible worlds, one of them corresponding to the actual world, but the agent, denoted You hereafter (but it may be a robot, a sensor...), does not know which among the possible worlds is the actual one. All You can state is the strength of Your opinion / belief that the actual world belong to this or that subset of Ω. The value $bel(A)$ represents the agent's belief that the actual world belongs to $A \subseteq \Omega$.

Beware that no concept of probability measure underlies the description of the TBM. We are presenting a theory for the representation of quantified beliefs that can exist without regard to the concept of probability functions. We only assume that degrees of beliefs are represented by a number, like in $bel(A) = .67$, that satisfies some 'natural' constraints.

A study of the rationality properties that should be satisfied by a function which purpose is to quantify someone's beliefs leads to the use of belief functions (Smets, 1997, 1993c). These axiomatic studies lead also to the derivation of Dempster's rule of conditioning. From this construction, we have derived (and often justified) many other concepts like :

- the conjunctive combination rule (that is Dempster's rule of combination) to compute $bel^{\Omega}[E_1 \wedge E_2]$ from $bel^{\Omega}[E_1]$ and $bel^{\Omega}[E_2]$ (see relation 7),
- the disjunctive rule of combination to compute $bel^{\Omega}[E_1 \vee E_2]$ from $bel^{\Omega}[E_1]$ and $bel^{\Omega}[E_2]$, (see section 3.5)
- the specialization concept: mass given to a set is redistributed among its subsets, (see section 3.4)
- the least commitment principle: 'never give more support than justified' what means that we should select the belief function which values of $pl(A)$ are as large as possible for every $A \subseteq \Omega$,
- the cautious combination rule: a conjunctive combination rule that is associative, commutative and idempotent, and accepts possible correlations between the sources (see section 3.6)
- the generalized Bayesian theorem to compute $bel^{\Theta}[x]$ for $\{bel^{X}[\theta_i] : \theta_i \in \Theta\}$, (see section 3.10)
- the measure of information content,
- the concept of doxastic independence,
- the pignistic transformationto build the probability function needed for taking 'optimal' decisions using the expected utility theory, (see section 3.9)...

The TBM is a largely extended model inspired by what is described in Shafer's book (note that some of Shafer's later papers enhance other interpretations).

3.2 Credal and pignisitc levels.

The TBM is based on the assumption that beliefs manifest themselves at two mental levels: the 'credal' level where beliefs are entertained and the 'pignistic' level where beliefs are used to make decisions (from 'credo' I believe and 'pignus' a bet, both in Latin).

Usually these two levels are not distinguished and probability functions are used to quantify beliefs at both levels. The justification for the use of probability functions is usually linked to "rational" behavior to be held by an ideal agent involved in some decision contexts (Ramsey, 1964; Savage, 1954; DeGroot, 1970). This result is accepted here, except that these probability functions quantify the uncertainty only when a decision is really involved.

At the credal level, we defend that beliefs are represented by belief functions. When a decision must be made, the beliefs held at the credal level induce a probability function at the pignistic level. This probability function is needed to compute the expected utilities and we call it the pignistic probability function, denoted by $BetP$. The transformation between the belief function and the pignistic probability function is called the pignistic transformation (see section 3.9).

3.3 Belief and plausibility.

In the TBM, the bba receives a natural interpretation. For $A \subseteq \Omega$, $m(A)$ is that part of Your belief that supports that all You know is that the actual world ω_0 belongs to A, and that, due to lack of information, does not support any strict subset of A.

If some further pieces of evidence become available to You and You accept them as valid, and if their only impact bearing on Ω is that they imply that the actual world ω_0 does not belong to $\overline{B}$, then the mass $m(A)$ initially allocated to A is transferred to $A \cap B$. Indeed, some of Your belief (quantified by $m(A)$) was allocated to A, and now You accept that $\omega_0 \notin \overline{B}$, so that mass $m(A)$ is transferred to $A \cap B$ (hence the name of the model). The resulting new basic belief assignment is the one obtained by the application of Dempster's rule of conditioning (relation (9)).

The degree of belief $bel(A)$ quantifies the total amount of justified specific support given to A. The degree of plausibility $pl(A)$ for $A \subseteq \Omega$ quantifies the maximum amount of potential specific support that could be given to A.

3.4 The conjunctive combination of two belief functions.

In the TBM, if m_1 and m_2 are the two bba's produced by two distinct sources of information, and You consider that both sources are fully reliable (so You believe what they state), then the two bba's are conjunctively combined by the equivalent of Dempster's rule of combination (relation (7)).

This rule is justified through the concept of *specialization*. It is assumes that whenever a new piece of information is accepted, the basic belief mass $m(A)$ previously given to $A \subseteq \Omega$ is distributed among the subsets of A. Indeed that mass $m(A)$ represents a part of belief that supports A and might support anything more specific than A, should we get some new information, and this is just what happens. We get some new information, so m(A) is going to be redistributed to the subsets of A (of course it may also stay allocate to A). Such an operation is what is called a 'specialization' (Yager, 1983).

One way to justify Dempster's rule of combination is as follows (Klawonn & Smets, 1992). Other justifications have been provided, but the next one seems the simplest one.

If You assume that learning that $A \subseteq \Omega$ is true, then You must define a new bba on Ω built as a specialization of the initial bba. You want that after specialization, the plausibility given to $\overline{A}$ be null, (indeed, now You know that $\overline{A}$ is impossible, so its plausibility must be zero). Among all the specializations that satisfy this requirement, there is a least committed one, i.e., one that gives as few support to every propositions. It correspond to the principle 'never give more support than justified'. In that case the least committed solution that satisfies $pl(\overline{A}) = 0$ is the solution described by Dempster's rule of conditioning (relation 9).

If then You want to build a conjunctive combination rule that is associative and commutative, and commute with the conditioning operation, the solution is again unique, and corresponds to Dempster's rule of combination (relation 7). This last relation can also, and very conveniently, be written as:

$$f_{1\cap 2}(A) = \sum_{B \subseteq \Omega} f_1[B](A) m_2(B), \ \forall A \subseteq \Omega$$

where f can be any of m, bel, pl, q, b.

These results explain the essential role played by Dempster's rules in data fusion. The associativity an commutativity are usually required for data fusion, and are just those constrains that lead to Dempster's rules.

3.5 The disjunctive combination of two belief functions.

But the conjunctive combination rule is not the only rule for combining belief functions. Sometimes, You receive two bba's, and You are not sure both sources are reliable. If all You know is that at least one of the sources is reliable, You end up with the disjunctive rule of combination, i.e., a combination rule where the product of the masses $m_1(A_1)m_2(A_2)$ is given, not to $A_1 \cap A_2$ as in the conjunctive case, but to $A_1 \cup A_2$. The result, denoted $m_{1\cup 2}$ where the $\cup$ symbol indicates the disjunctive nature, is :

$$m_{1\cup 2}(A) = \sum_{B \cup C} m_1(B) m_2(C), \ \forall A \subseteq \Omega \tag{10}$$

$$b_{1\cup 2}(A) = b_1(A) b_2(A), \ \forall A \subseteq \Omega \tag{11}$$

The last relation shows the interest of b, the implicability function.

To see the origin of thus rule, consider one sources states the actual world is in A_1 and second source states it is in A_2, and You only know that one of them is reliable but You don't know which one, all You can conclude is that the actual world is in $A_1 \cup A_2$.

In particular, this disjunctive rule can be used to compute $bel^X[\theta_1 \cup \theta_2]$ from $bel^X[\theta_1]$ and $bel^X[\theta_2]$.

3.6 The cautious conjunctive combination.

Still other cases can be considered, the most important one being probably the cautious conjunctive combination, a combination rule that is not only commutative and associative but also idempotent. It is a rule able to cope with correlated sources of information.

Suppose You have two sources of evidence who express their beliefs about the value of the actual world. Let m_1 and m_2 be two bba's on Ω. We do not assume the two pieces of evidence underlying these two bba's are 'distinct'. Each bba may result from some information common for both sources and some individual information. Thus we do not assume that the sources are distinct. They may be 'correlated', they can even repeat the same information twice as they are in fact just one source (but You did not know it).

The aim is to build a new bba, that will be denoted $m_{1\wedge 2}$, from m_1 and m_2. To achieve this we consider, for $i = 1$ and 2, all the bba's that could be built from m_i by its conjunctive combination with any bba on Ω. So let for $i = 1, 2$,

$$B_i = \{m : m \text{ on } \Omega, m = m_i \cap m^* \\ \text{for } m^* \text{ belief function on } \Omega\}.$$

B_i are the belief functions reachable from m_i by combining it with any belief function on Ω.

All we know is that the result of the 'combination' of m_1 and m_2 must be in $B_1 \cap B_2$. In particular, $m_1 \cap m_2$ belongs to that family. Not knowing which element of $B_1 \cap B_2$ is appropriate, the cautious attitude consists in selecting the least committed elements of that family. This is the idea of the 'cautious conjunctive combination rule'. Solutions have been found in several important cases, but finding the least committed solution is not yet fully solved for the general case.

3.7 Discounting.

Suppose a source tells You that the belief functionon Ω is m, but You are not sure the source is really reliable (maybe it refers to another question than the one You are interested in). Suppose You believe at level α that the source is reliable, and $1 - \alpha$ it is not, then Your belief m^* on Ω becomes :

$$m^*(A) = \alpha\ m(A),\ \forall A \neq \Omega \tag{12}$$

$$m^*(\Omega) = 1 - \alpha + \alpha\ m(\Omega). \tag{13}$$

This is just an application of the disjunctive rule of combination and was called the discounting by Shafer. Its interest comes from the possibility to 'discount' sources of information when You feel they are not fully reliable.

Beware that combination and discounting do not commute, so the order with which they are applied is important.

3.8 Distinctness.

The concept of 'disitinct' pieces of evidence is usually left undefined. We propose that it means in fact that once You know the bba m_1 produced by one source of information, what You know about the domain of the bba that could be produced by another source is unchanged in comparison to what it was before You learn the value of m_1 (see also (Smets, 1992, 1998b)).

3.9 Decision making.

When a decision must be made, we use the expected utility theory, what implies the need to construct a probability function on Ω. This is achieved by the so-called pignistic transformation which transforms a belief function into a probability function, called the pignistic probability and denoted by $BetP$. The nature of this transformation and its full justification is described in Smets and Kennes (1994).

The value of the pignistic probability if given by :

$$BetP(A) = \sum_{X \subseteq \Omega} \frac{|A \cap X|}{|X|} \frac{m(X)}{1 - m(\emptyset)} \tag{14}$$

where $|A|$ denotes the number of worlds in the set A.

It is easy to show that the function $BetP$ so obtained is indeed a probability function. Decisions are then achieved by computing the expected utilities of the acts using $BetP$ as the probability function needed to compute the expectations. Why this probability function is adequate for decision making and its use to provide an operational definition to the values of a belief function is detailed in Smets and Kennes (1994).

3.10 The Generalized Bayesian Theorem.

In probability theory, Bayes theorem permits the computation of a probability over some space Θ given the value of some variable $x \in X$ from the knowledge of the probabilities over X given each $\theta_i \in \Theta$, and some *a priori* probability function over Θ. The same ideas have been extended in the TBM context where we will build a belief function over Θ given an observation $x \subseteq X$ and the knowledge of the belief function over X given $\theta_i \in \Theta$ and a vacuous *a priori*, i.e., an *a priori* describing a state of total ignorance (therefore solving the delicate problem of choosing the appropriate *a priori*).

Suppose the finite spaces X and Θ. Suppose that for each $\theta_i \in \Theta$, there is a basic belief assignment on X, denoted $m^X[\theta_i]$. Given this set of basic belief assignments, what is the belief induced on Θ if You come to know that $x \subseteq X$ holds?

In statistics, this is the classical inference problem where Θ is a parameter space and X an observation space. In target detection, θ_i is an hypothesis and x the observation. In diagnosis, θ_i is the causal event, the disease, the failure, and x is the observable, the symptom, the collected data...

The Generalized Bayesian Theorem (GBT) performs the same task as the Bayesian theorem but within the TBM context. The major point is that the needed prior can be a vacuous belief function, what is the perfect representation of total ignorance. No informative prior belief is needed, avoiding thus one of the major criticisms against the classical approach, in particular when used for diagnostic applications. Of course, should some *a priori* belief over Θ be available, it would be combined by the conjunctive combination rule with the result obtained by the GBT. Whenever this *a priori* belief is considered, before or after applying the GBT, the result is the same.

Given the set of basic belief assignments $m^X[\theta_i]$ known for every $\theta_i \in \Theta$ and their related functions then for $x \subseteq X$ and for every $A \subseteq \Theta$:

$$b^{\Theta}[x](A) = \prod_{\theta_i \in \overline{A}} b^X[\theta_i](\overline{x}) \quad (15)$$

$$pl^{\Theta}[x](A) = 1 - \prod_{\theta_i \in A} (1 - pl^X[\theta_i](x)) \quad (16)$$

$$q^{\Theta}[x](A) = \prod_{\theta_i \in A} pl^X[\theta_i](x) \quad (17)$$

where $[x]$ is the piece of evidence that states 'x holds'.

Should You have some non vacuous beliefs on Θ, represented by $m^{\Theta}[E_0]$, than this belief is simply combined with $m^{\Theta}[x]$ by the application of the conjunctive rule of combination.

This rule has been derived axiomatically by Smets (1978, 1986, 1993a) and by Appriou (1991). When the belief function on X given θ_i is a probability function, as it will often be the case for practical applications, we just replace $bel^X[\theta_i](x)$ and $pl^X[\theta_i](x)$ by $P(x|\theta_i)$.

Some particular cases are worth mentioning.

Case 1. We consider the case of two 'independent' observations x defined on X and y defined on Y, and the inference on Θ obtained from their joint observation.

Suppose the two variables X and Y satisfy the Conditional Cognitive Independence property defined as:

$$pl^{X \times Y}[\theta_i](x, y) = pl^X[\theta_i](x) pl^Y[\theta_i](y),$$
$$\forall x \subseteq X, \forall y \subseteq Y, \forall \theta_i \in \Theta.$$

When $pl^X[\theta_i]$ and $pl^Y[\theta_i]$ are probability functions, this property is just the classical conditional independence property.

The GBT could be applied in two different ways.

Let $b^{\Theta}[x]$ and $b^{\Theta}[y]$ be computed by the GBT (with a vacuous a priori belief on Θ) from the set of basic belief assignment $m^X[\theta_i]$ and $m^Y[\theta_i]$ known for every $\theta_i \in \Theta$. We then combine by the conjunctive rule of combination these two functions in order to build the belief $b^{\Theta}[x, y]$ on Θ induced by the pair of observations.

We could as well consider the basic belief assignment $m^{X \times Y}[\theta_i]$ built on the space $X \times Y$ thanks to the Conditional Cognitive Independence property, and compute $b^{\Theta}[x, y]$ from it using the GBT.

Both results are the same, a property that is essential and at the core of the axiomatic derivations of the rule.

Case 2. If for each $\theta_i \in \Theta$, $b^X[\theta_i]$ is a probability function $P(.|\theta_i)$ on X, then the GBT for $|\theta_i| = 1$ becomes:

$$pl^{\Theta}[x](\theta_i) = P(x|\theta_i), \qquad \forall x \subseteq X.$$

That is, on the singletons θ_i of Θ, $pl^{\Theta}[x]$ reduces to the likelihood of θ_i given x. The analogy stops there as the solution for the likelihood of subsets of Θ are different.

If, furthermore, the *a priori* belief on Θ is also a probability function $P_0(\theta)$, then the normalized GBT becomes:

$$bel^{\Theta}[x](A) = \frac{\sum_{\theta_i \in A} P(x|\theta_i) P_0(\theta_i)}{\sum_{\theta_i \in \Theta} P(x|\theta_i) P_0(\theta_i)} = P(A|x)$$

i.e. the (normalized) GBT reduces itself into the classical Bayesian theorem (as it should), which explains the origin of its name.

4 Applying the TBM to the fusion of uncertain data.

4.1 The TBM classifier.

4.1.1 Partially known classes.

Discriminant analysis is probably the most classical tool used for classifying cases into one of several categories given the values of some measurement variables. Normally, we use a set of data, called the learning set (LS). For each case in LS, we know the values taken for each measurement variable and the classification variable that tells the class to which the case belongs. The classes are finite and unordered. Let Ω denote the set of possible classes: $\Omega = c_1, c_2, ..., c_n$.

A learning set with N cases and p measurement variables is the set $\{(c_i, x_{1i}, x_{2i}, ...x_{pi}) : i = 1, 2...N\}$ where X_i is the 'name' of the i'th case, c_i is the class to which X_i belongs, and x_{ji} is the value of the measurement variable j for X_i. The data of a new case, denoted $X_?$, is collected, but the class to which $X_?$ belongs, denoted $c_?$, is unknown. We want to predict the value of $c_?$ given the observed values of the measurement variables of $X_?$. Solutions to this problem are well established. One of them, called discriminant analysis, is fully described in most textbooks of statistics.

Let us now suppose that instead of the ideal learning set LS as described here above, we have a learning set PKLS where the classes of the cases are only partially known. For instance suppose we only know that case X_1 belongs either to c_1 or c_2 class, that case X_2 does not belong to class c_1, case X_3 belongs either to c_2 or c_5 or c_7 class... Can we adapt the discriminant analysis method to such 'messy' data context? In fact we face a problem of 'partially supervised learning'. For some cases, classes are known as in the supervised learning approach, for some cases, class in completely unknown as in the unsupervised approach. But here we also have all the cases where we know partially their class. Probabilistic solutions could be based on:

- a Bayesian approach where we assess for each case a probability function that describes the class to which it belongs. We then allocate every case to a class (and compute the probability to get that learning set), compute the needed parameters as in a supervised learning approach and average the results weighted by the probability of the learning sets,
- a maximum likelihood approach (like with the EM algorithm) where we estimate the unknown parameters, including the probability with which the case belong to a given class,
- an adaptation of cluster analysis where partial constraints are introduced that represent the

knowledge about the class to which each case belong.

Whatever method is used, the computational complexity is a serious problem and an adequate tuning of some parameters is not a small matter. The transferable belief model provides another approach that can handle elegantly and efficiently such a messy problem. The method was invented by Denoeux (1995).

We present the method, called the TBM classifier. The comparison of this method with the classical ones can be found in (Denoeux, 1995; De Smet, 1998; Zouhal & Denœux, 1998; Smets, 1999).

4.1.2 Discriminant Analysis with Partially Known Classes.

Let pkc_i denote the subset of Ω that represents what we know about the class to which case X_i belongs. The learning set PKLS is now the set $\{(pkc_i, x_{1i}, x_{2i}, ...x_{pi}) : i = 1, 2...N\}$.

Intuitively the method can be described by an anthropomorphic model. Each case X_i in PKLS is considered as an individual. Let c_{i0} denote the true class to which X_i belongs. All X_i knows about c_{i0} is that $c_{i0} \in pkc_i$. ((Denœux & Zouhal, 1999) generalizes to the case where this knowledge is represented by a belief function or a possibility function on Ω). Then X_i looks at the unknown case and expresses 'his' belief bel_i about $c_?$. If $X_?$ is 'close' to X_i, X_i would defend that $c_? = c_{i0}$. As all what X_i knows about c_{i0} is that $c_{i0} \in pkc_i$, then all what X_i can express about case $X_?$ is that $c_? \in pkc_i$. If $X_?$ is not 'close' to X_i, X_i cannot say anything about c_{i0}.

This description is formalized as follows. X_i can only state: case $X_?$ belongs to the same set of classes as myself, what is represented by a belief function with $m_{i0}(pkc_i) = 1$. Let $d(X_i, X_?)$ be the 'distance' between X_i and $X_?$. If $d(X_i, X_?)$ is small, then what X_i states is reliable, if $d(X_i, X_?)$ is large, it is not reliable, the largest $d(X_i, X_?)$, the less reliable. The impact of this reliability is represented by a discounting on m_{i0} into m_i. So $m_i(pkc_i) = f(d(X_i, X_?))$ and $m_i(\Omega) = 1 - f(d(X_i, X_?))$ where $f(d) \in [0, 1]$ and is decreasing with d. Thus every case X_i generates such a simple support function bel_i on Ω that concerns the value of $c_?$.

Consider now what information $X_?$ collects. Case $X_?$ receives all these simple support functions bel_i, and combines them by the conjunctive rule of combination into a new belief function bel on Ω that represents the belief held by case $X_?$ about $c_?$ and induced by the collected belief functions bel_i:

$$bel_? = \cap_{i=1N} bel_i.$$

If a decision must be made on the value of $c_?$, we build the pignistic probability $BetP_?$ on Ω from $bel_?$ by the application of the pignistic transformation and use the classical expected utility theory in order to take the optimal decision.

De Smet (1998) applied this approach to many sets of data. Results were very satisfactory (see also (Smets, 1999)). We feel this method could be part of the toolbox for classification when data are ill known. It has been extended to regression problems, using a similar approach, and has even been adapted to fuzzy sets by Denoeux.

4.2 Sensors on partially overlapping frames.

Suppose a sensor S_1 that has been trained to recognize A objects and B objects and another sensor S_2 that has been trained to recognize B objects and C objects (like A = airplanes, B = helicopters and C = rockets). Sensor S_1 never saw a C object, and we know nothing on how S_1 would react if it looks at a C object. Beliefs provided by S_1 are always on the frame of discernment $\{A, B\}$. The same holds for S_2 with A and C permuted. A new object X is presented to the two sensors. Both sensors S_1 and S_2 express their beliefs m_1 and m_2, the first on the frame $\{A, B\}$, the second on the frame $\{B, C\}$. How to combine these two beliefs on a common frame $\Omega = \{A, B, C\}$? Some solutions have been proposed in (Janez, 1996; Janez & Appriou, 1996a, 1996b).

One solution is based on the next constraint. If both m_1 and m_2 are conditioned on B, and combined by Dempster's rule of combination (unnormalized), the resulting belief function should be the same as the one obtained after 'combining' the original m_1 and m_2 on $\{A, B, C\}$, and conditioning the result on B. The problem is of course how to 'combine' m_1 and m_2. The original Dempster's rule of combination is inadequate as it requires that both belief functions are defined on compatible frames of discernment, what is not the case here.

A general solution is as follows. Let Ω_1 and Ω_2 be the frame of discernment of m_1 and m_2, respectively. Let $\Omega = \Omega_1 \cap \Omega_2$. For all $A \subseteq \Omega_1 \cup \Omega_2$, let $A_1 = A \cap \Omega_1$, $A_2 = A \cap \Omega_2$, $A_0 = A \cap \Omega$, then let m be the result of the combination with:

$$m(A) = \frac{m_1(A_1)}{m_1[\Omega](A_0)} \frac{m_2(A_2)}{m_2[\Omega](A_0)} (m_1[\Omega] \cap m_2[\Omega])(A_1 \cap A_2)$$

where $m_1[\Omega]$ and $m_2[\Omega]$ are the basic belief assignments obtained by conditioning m_1 and m_2 on Ω.

In table 2, we illustrate the computation. We have $m_1[B] \cap m_2[B](B) = (.1 + .3) * (.7 + .1) = .32$. This mass is distributed on $\{B\}, \{A, B\}, \{B, C\}$ and $\{A, B, C\}$ according to the next ratios: (.1/.4).(.7/.8), (.3/.4).(.7/.8), (.1/.4).(.1/.8), and (.3/.4).(.1/.8). The mass $m_1[B] \cap m_2[B](\emptyset) = .68$ is given to $\{A, C\}$.

In this example the first sensor supports that X is an A, whereas the second claims that X is a B. If X had been a B, how comes the first sensor did not say so? So the second sensor is probably facing an A and just states B because it does not know what an A is. So we feel that the most plausible solution is $X = A$, what is confirmed by $BetP$ as it is the largest for A: $BetP(A) = .45$.

Table 2: Basic belief assignment m_1 and m_2 on the two partially overlapping frames, with their combination m and its related plausibility and pignistic probability functions.

Ω	m_1	m_2	m	pl	$BetP$
A	.6		.00	.92	.455
B	.1	.7	.07	.32	.190
C		.2	.00	.72	.355
A, B	.3		.21	1	
A, C			.68	.93	
B, C		.1	.01	1	
A, B, C			.03	1	

Table 3: The simple support functions generated by the four sensors on the frame of discernment $\Omega = \{C_1, C_2\}$.

Ω	m_1	m_2	m_3	m_4
C_1	.7	.8		
C_2			.6	.9
Ω	.3	.2	.4	.1

4.3 The data association problem.

Suppose a piece of equipment has failed. We collect data from four sensors S_1, S_2, S_3 and S_4. Each sensor produces a belief function on the set of possible component that might have failed. Table 3 presents a highly simplified example where each sensor produces a simple support function pointing toward one component. S_1 and S_2 both point toward component C_1, whereas S_3 and S_4 point toward component C_2. If the four sources S_1 to S_4 were highly reliable, You would conclude that both C_1 and C_2 are broken. Indeed if only one has failed, the source are contradictory, whereas if two components have failed, results are coherent if S_1 and S_2 report on one broken component and S_3 and S_4 report on a second broken component.

How do we translate this problem into belief functions language? The solution is obtained by considering the mass $m(\emptyset)$ given to $\emptyset$ that may be positive in the transferable belief model. When applying Dempster's rule of combination to two basic belief assignments m_1 and m_2, the result is given by: $m_{1\cap 2}(A) = \sum_{X\cap Y=A} m_1(X)m_2(Y)\ \forall A \subseteq \Omega$. We do not normalize the resulting basic belief assignment $m_{1\cap 2}$. The mass $m(\emptyset)$ is among the computed masses and it does not have to be 0 like in Shafer's original presentation. The mass $m(\emptyset)$ quantifies the amount of contradiction between the various sources of belief functions.

Schubert (1995) has proposed a strategy to decide the number of events under consideration by the various sensors producing the several collected belief functions. He analyses $m(\emptyset)$ and finds the association between sensors and events that somehow brings the total conflict to an acceptable level .

Suppose the data of table 3. If there is only one broken component the four sensors are speaking about the same event. The contradiction computed after combining the four basic belief assignments is 0.90, what reflect an enormous conflict between the four sources. If there is two broken components, then some sources might speak about one, the other about the second. So we split the four sensors into two groups, compute what is the contradiction within each group, and sum these contradictions. For instance, suppose sensors S_1, S_2 and S_3 speak about one component, then the contradiction is 0.56, whereas there is no contradiction for sensor 4. Total contradiction is thus 0.56. Now if we consider that sensor S_1 and S_3 speak about one component, whereas S_2 and S_4 speak about the other, the total contradiction is 1.14. Contradiction completely disappears when S_1 and S_2 are grouped as reporting on one component, and S_3 and S_4 on the second. This result fits with common sense analysis of the data. In real life applications, the basic belief assignments are usually quite elaborated, and finding an adequate grouping is not obvious. The technique of 'peeling' the mass given to the empty set (to the contradiction) is nevertheless still applicable. The level of 'tolerable contradiction' is itself determined by the analysis of the conflict present in the given belief functions (and obtained by the use of the canonical decomposition of the belief functions (Smets, 1995)).

Table 4: Masses $m(\emptyset)$ computed from the belief functions included in each group when considering two objects.

Groups		Conflict		
G_1	G_2	G_1	G_2	total
1234	-	.90	-	.90
123	4	.56	.00	.56
124	3	.85	.00	.85
134	2	.67	.00	.67
234	1	.77	.00	.77
12	34	.00	.00	.00
13	24	.42	.72	1.14
14	23	.63	.48	1.11

The mass $m(\emptyset)$ acts in fact as a measure of discrepancy between several belief functions. The proposed algorithm leads to grouping sources which belief functions are 'close' to each other. In probability theory using cross-entropy or chi-square coefficients can achieve this purpose. Comparisons between these approaches are not available (as far as we know). The advantage of the belief function approach resides in the well-founded nature of the approach. The mass $m(\emptyset)$ is part of the transferable belief model, whereas the cross-entropy, the chi-square and the likes need always extra assumptions in order to justify their use.

4.4 Tuning the discounting.

Discounting is often used in the TBM, therefore it might be nice to have a way to estimate the value of the discounting factor. This is possible as illustrate in the next scenario.

Suppose we must build a classifier. We collect a set of data that we submit for classification to my experts.

Let C be the set of possible classes. Let C_{ij} be the conclusion of expert j on the class to which object i belongs, with $C_{ij} \subseteq C$ (note that the expert opinion could even be a belief function over C).

Suppose we feel that some discounting should be introduced. Let d_j be the discounting factor applicable to expert j, and to be determined. If we knew d_j, we would compute for each object a bba m_{ij} over C that represent the discounted beliefs we have based on what expert j states and that depends directly on d_j. All experts' results would be combined by the conjunctive rule of combination. The result is $m_i = \cap_j m_{ij}$, from which we compute $BetP_i$, the pignistic probability over C with which we would decide the class to which object i belongs.

Suppose now that we know the actual classes c_i for these objects. We would like that $BetP_i$ points as strongly as possible to the actual class of object i. So we would like that $BetP_i(c_i) = 1$ is the class of the object is c_i, and 0 otherwise. Let D_i be the distance between the pignistic probability computed for object i and the indicator function of the class c_i:

$$D_i = \sum_k (BetP_i(c_k) - \delta_{ik})^2$$

where $\delta_{ik} = 1$ if object i belongs to class k, and 0 otherwise.

We then add these distances over the objects and determine the set of discounting factors d_j that minimize that sum. These will be the discounting factor we will use in the future for the evaluated experts.

4.5 Comments on efficiency.

It is often claimed that the use of belief function is poised by the computational complexity. Indeed, in theory, we work on the power set, and not on the set as it is the case for probability theory and possibility theory. Nevertheless, this is not necessarily the case. There is no need to build always all the possible values of *bel*, etc... There are many cases where the knowledge is very simple, and where there are very few non null masses, in which case all that must be stored and computed is proportional to the number of non null masses, and this often will be smaller than the number of elements, making the belief function computation lighter than its competitors.

Studies have been published that claim that probability approach is faster than belief function one. Unfortunately, these authors usually ignore the pignistic transformation and the General Bayesian Theorem, and, even worse, use erroneous relations. In fact it has been shown that, for a target detection problem, the TBM approach was even more efficient that the classical probability approach[1]. So before rejecting the TBM for its computational load, potential users should be very skeptical with the published conclusions.

[1]Delmotte, personal communications.

5 Conclusions.

In many applications, the information needed to apply the probability approach is unfortunately not available. One could of course try to fit the missing information by some 'educated guesses'. Quality of the results is of course directly related to the quality of the 'fixing'.

On the contrary, the belief function models is well adapted to work with the information as really available. This power comes from the ability of belief functions to represent any form of uncertainty: full knowledge, partial ignorance, total ignorance (and even probability knowledge). Probability functions do not have such expressiveness power. Equi-probability is not full ignorance, it is already a quite precise form of knowledge.

Before applying the TBM, it is of course necessary to have a good understanding of the theoretical foundations and the exact nature of the various tools. This is of course not specific to the TBM. It is the same for any approach. Of course, users are usually more familiar with probability theory, and often jump on that tool without realizing that other tolls are maybe better adapted to their problems. For someone whose only tool is a hammer, the world looks like a nail. Sophisticated users must be open to all the available tools, using the right one for the right problem. When facing uncertain data, the user must not limited himself to probability models, but also consider the new models like those based on fuzzy set theory, on possibility theory or on belief functions and select the one that fits the problem.

References

Appriou, A. (1991). Probabilités et incertitude en fusion de données multi-senseurs. *Revue Scientifique et Technique de la Défense*, *11*, 27-40.

De Smet, Y. (1998). *Application de la théorie des fonctions de croyances aux problèmes de classification.* Unpublished master's thesis, Université Libre de Bruxelles.

DeGroot, M. H. (1970). *Optimal statistical decisions.* McGraw-Hill, New York.

Dempster, A. P. (1967). Upper and lower probabilities induced by a multplevalued mapping. *Ann. Math. Statistics*, *38*, 325–339.

Dempster, A. P. (1968). A generalization of Bayesian inference. *J. Roy. Statist. Soc. B.*, *30*, 205–247.

Dempster, A. P. (1972). A class of random convex polytopes. *Annals of Mathematical Statistics*, *43*, 250–272.

Denoeux, T. (1995). A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man and Cybernetics:*, *25*, 804–813.

Denœux, T., & Zouhal, L. M. (1999). Handling possibilistic labels in pattern classification using evidential reasoning. *Submitted.*

Dubois, D., & Prade, H. (1998). Introduction: Revising, updating and combining knowledge. In D. M. Gabbay & P. Smets (Eds.), *Handbook of defeasible reasoning and uncertainty management systems* (Vol. 3, pp. 1–16). Kluwer, Doordrecht, The Netherlands.

Gabbay, D. M., & Smets, P. (Eds.). (1998). *Handbook of defeasible reasoning and uncertainty management systems.* Kluwer, Doordrecht, The Netherlands.

Janez, F. (1996). *Fusion de sources d'information définies sur des référentiels non exhaustifs différents.* Unpublished doctoral dissertation, Université d'Angers, France.

Janez, F., & Appriou, A. (1996a). Fusion of sources defined on different non-exhaustive frames. In IPMU-96 (Ed.), *Information processing and management of uncertainty.*

Janez, F., & Appriou, A. (1996b). Théorie de l'évidence et cadres de discernements non exhaustifs. *Revue Traitement du Signal.*, *13*.

Klawonn, F., & Smets, P. (1992). The dynammic of belief in the transferable belief model and specialization- generalization matrices. In D. Dubois, M. P. Wellman, B. D'Ambrosio, & P. Smets (Eds.), *Uncertainty in artificial intelligence 92* (pp. 130–137). Morgan Kaufman, San Mateo, Ca.

Kohlas, J., & Monney, P. A. (1995). *A mathematical theory of hints: An approach to dempster-shafer theory of evidence.* Lecture Notes in Economics and Mathematical Systems No. 425. Springer-Verlag.

Kyburg, H. E., Jr. (1987). Bayesian and non-Bayesian evidential updating. *Artificial Intelligence*, *31*, 271–294.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* Morgan Kaufmann Pub. San Mateo, Ca, USA.

Ramsey, F. P. (1964). Truth and probability. In H. E. Kyburg, Jr. & H. E. Smokler (Eds.), *Studies in subjective probability* (pp. 61–92). Wiley, New York.

Ruspini, E. H. (1986). *The logical foundations of evidential reasoning* (Tech. Rep.). Technical note 408, SRI International, Menlo Park, Ca, US.

Ruspini, E. H. (1987). Epistemic logics, probability and the calculus of evidence. In IJCAI-87 (Ed.), *Intern. joint conf. on artificial intelligence* (pp. 924–931). Morgan Kaufman, San Mateo, Ca.

Savage, L. J. (1954). *Foundations of statistics.* Wiley, New York.

Schubert, J. (1995). Cluster-based specification techniques in Dempster-Shafer theory. In C. Froidevaux & J. Kohlas (Eds.), *Symbolic and quantitative approaches to uncertainty.* Springer-Verlag, Berlin.

Shafer, G. (1976). *A mathematical theory of evidence.* Princeton Univ. Press. Princeton, NJ.

Smets, P. (1978). *Un modèle mathématico-statistique simulant le processus du diagnostic médical.* Unpublished doctoral dissertation, Université Libre de Bruxelles, (available through University Microfilm International, 30-32 Mortimer street, London W1N 7RA, thesis 80-70,003).

Smets, P. (1986). Bayes' theorem generalized for belief functions. In *Proc. ecai-86* (pp. 169–171).

Smets, P. (1991). Probability of provability and belief functions. *Logique et Analyse*, *133*, 177–195.

Smets, P. (1992). The concept of distinct evidence. In IPMU-92 (Ed.), *Information processing and management of uncertainty* (pp. 89–94).

Smets, P. (1993a). Belief functions: the disjunctive rule of combination and the generalized bayesian theorem. *International Journal of Approximate Reasoning*, *9*, 1–35.

Smets, P. (1993b). Probability of deductibility and belief functions. In M. Clarke, R. Kruse, & S. Moral (Eds.), *Symbolic and quantitative approaches to uncertainty* (pp. 332–341). Springer-Verlag, Berlin.

Smets, P. (1993c). An axiomatic justifiaction for the use of belief function to quantify beliefs. In IJCAI-93 (Ed.), *Intern. joint conf. on artificial intelligence* (pp. 598–603). Morgan Kaufman, San Mateo, Ca.

Smets, P. (1997). The normative representation of quantified beliefs by belief functions. *Artificial Intelligence*, *92*, 229–242.

Smets, P. (1998a). Numerical representation of uncertainty. In D. M. Gabbay & P. Smets (Eds.), *Handbook of defeasible reasoning and uncertainty management systems* (Vol. 3, pp. 265–311). Kluwer, Doordrecht, The Netherlands.

Smets, P. (1998b). The transferable belief model for quantified belief representation. In D. M. Gabbay & P. Smets (Eds.), *Handbook of defeasible reasoning and uncertainty management systems, vol. 1* (pp. 267–301). Doordrecht, The Netherlands: Kluwer.

Smets, P. (1999). Practical uses of belief functions. In *Proceedings of the 14th conference on uncertainty in artificial intelligence, stockholm, sweden*

Smets, P., & Kennes, R. (1994). The transferable belief model. *Artificial Intelligence*, *66*, 191–234.

Voorbracht, F. (1991). On the justification of Dempster's rule of combination. *Artificial Intelligence*, *48*, 171–197.

Walley, P. (1991). *Statistical reasoning with imprecise probabilities.* Chapman and Hall, London.

Yager, R. R. (1983). Entropy and specificity in a mathematical theory of evidence. *International Journal of General Systems*, *9*(4), 249-260.

Zadeh, L. A. (1986). A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination. *AI Magazine*, *7*, 85–90.

Zouhal, L. M., & Denœux, T. (1998). An evidence theoretic k-nn rule with parameter optimisation. *IEEE Transactions on Systems, Man and Cybernetics - Part C*, *28*, 263-271.