Assessing Sensor Reliability for Multisensor Data Fusion within the Transferable Belief Model

Zied Elouedi[1], Khaled Mellouli[1], Philippe Smets[2]
[1] Institut Supérieur de Gestion de Tunis,
41 Av. de la liberté, Cité Bouchoucha, 2000 Le Bardo, Tunis, Tunisia
[2] IRIDIA, Université Libre de Bruxelles,
50 av., F. Roosevelt, CP194/6, 1050 Brussels, Belgium
zied.elouedi@isg.rnu.tn, khaled.mellouli@ihec.rnu.tn,
psmets@ulb.ac.be

Abstract

This paper presents a method for assessing the reliability of a sensor in a classification problem based on the transferable belief model. First, we develop a method for the evaluation of the reliability of a sensor when considered alone. The method is based on finding the discounting factor minimizing the distance between the pignistic probabilities computed from the discounted beliefs and the actual values of data. Next, we develop a method for assessing the reliability of several sensors that are supposed to work jointly and their readings are aggregated. The discounting factors are computed on the basis of minimizing the distance between the pignistic probabilities computed from the combined discounted belief functions and the actual values of data.

1 Introduction

A multisensor data fusion system is an important component in many fields dealing with pattern recognition, identification, diagnosis, etc. It is used with the hope that the aggregation of several sensors achieves better results. In this paper, we consider sensors delivering classification (categorical) readings. In the simplest cases, each sensor reading is a set of classes containing the actual class. In more complex cases, some or all sensors communicate several such sets, each weighted by a measure of confidence. Some classical measures are based on the possibility theory and the fuzzy set theory, on the probability theory and on the belief function theory. This paper considers only the belief function theory as understood in the transferable belief model (TBM) [8, 10, 12].

Our discussion of sensor fusion is also potentially applicable to related problems, such as collating expert opinions. Expert opinions are equivalent to sensor readings, and pooling is equivalent to fusion [3]. Experts differ in their level of expertise, some of them are more reliable than others due to their better knowledge, training, experience, intelligence, etc. To express their opinions, experts may use different background, methodology and knowledge. Hence, the necessity to consider the expert reliability and consequently their judgments must be appropriately 'discounted'. Similarly, sensors do not have the same degree of reliability. This may be due not only to the same reasons as already mentioned for experts, but also to other factors more specific to sensors. For instance, measurements can differ from one sensor to another in terms of completeness, precision, and certainty. Additionally, the working environment can also affect sensor reliability since some of them could be better adapted to the conditions encountered in the considered environment than others. Thus, sensor reliabilities must be assessed before using their readings.

The TBM provides a highly flexible model to manage the uncertainty encountered in the multisensor data fusion problems [1, 2, 5, 6]. The sensor reading about the actual value of a variable is represented by a belief function. The reliability of the sensor is represented by a discounting factor i.e., a coefficient that 'weights' the belief function produced by the sensor. Reliability and discounting are linked, the smaller the reliability, the larger the discounting. Methods for evaluating the experts discounting factors are presented in [4, 13]. In [13], the authors have introduced the idea of estimating discounting factors by minimizing an error function, an idea which is the starting point of our paper. In [13], each observation of the training set is seen, using our terminology, as a 'sensor' which receives a discounting factor. This latter depends on some distance between itself and the one to classify. The solution is different from the one presented here.

Our paper develops methods to assess discounting factors to be applied to sensor readings. In the first case, a sensor is considered alone. The method to assess the discounting factor is based on the comparison of the discounted sensor readings (represented by belief functions on the domain of the actual classes) with what we know about the actual classes. In the second case, several sensors are considered jointly. The assessment method is based on the comparison of the combined discounted readings (also represented by belief functions) with the actual classes. Section 2 recalls briefly the basics of the belief function theory as interpreted in the TBM. Section 3 and 4 represent the major part of this paper where the methods for the evaluation of the reliability of sensors will be detailed. An illustrative example will be presented to explain the different methods.

2 Belief Function Theory

In this section, we briefly recall some basics of the belief function theory as explained in the transferable belief model (TBM). More details can be found in [7, 10, 12].

2.1 Basic concepts

Let Θ be a non empty finite set including all the elementary events related to a given problem. These events are assumed to be exhaustive and mutually exclusive. The set Θ is called the **frame of discernment**. A basic belief assignment (bba) is a function m from 2^Θ to $[0,1]$ satisfying:

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

The basic belief mass $m(A)$, represents the part of belief exactly committed to the subset A of Θ given a piece of evidence, or equivalently to the fact that all we know is that A holds. When $m(\emptyset) = 0$, m is called a normalized bba. When $m(\Theta) = 1$, m is called a vacuous bba.

2.2 Combination

Combining bba's induced from distinct pieces of evidence is achieved by the conjunctive rule of combination. The bba obtained by the conjunctive combination rule applied to the bba's m_1 and m_2 is denoted by $m_1 \textcircled{\cap} m_2$ and defined for all $A \subseteq \Theta$ as:

$$(m_1 \textcircled{\cap} m_2)(A) = \sum_{B,C \subseteq \Theta : B \cap C = A} m_1(B) m_2(C) \tag{2}$$

2.3 Discounting

Reliability is expressed here by the user opinion about the 'value' of the sensor reading. The idea is to weight most heavily the opinions of the best sensors and conversely for the less reliable ones. The result is a discounting of the bba m^Θ produced by the sensor into the new bba $m^{\Theta,\alpha}$ where:

$$\begin{cases} m^{\Theta,\alpha}(A) = (1-\alpha)\ m^\Theta(A),\ \forall A \subseteq \Theta, A \neq \Theta \\ m^{\Theta,\alpha}(\Theta) = \alpha + (1-\alpha)\ m^\Theta(\Theta) \end{cases} \tag{3}$$

The coefficient $(1-\alpha)$ can be regarded as the degree of 'trust', the strength of reliability assigned to the sensor reading. The rule is justified in [9].

2.4 Pignistic Transformation

In the TBM, when a decision has to be made we build a probability function $BetP$ on Θ, called the pignistic probability function, by applying the pignistic transformation. $BetP$ is defined as [12] :

$$BetP(A) = \sum_{B \subseteq \Theta} \frac{|A \cap B|}{|B|} \frac{m(B)}{1 - m(\emptyset)},\ \forall A \subseteq \Theta \tag{4}$$

This solution is a classical probability measure from which expected utilities can be computed in order to take optimal decisions. Full details and justifications can be found in [11, 12].

3 Assessing the Discounting Factor of one Sensor

Usually the discounting factor applicable to a sensor is unknown and the users would like to find objective ways to assess it. We present a method to assess the optimal discounting factor applicable to a sensor reading. The method can easily be adapted to other domains, the underlying schema being quite general. The idea is to compare the readings of the sensor with what the user knows about the actual values. We consider several states of knowledge about the actual class of the data in the learning set.

3.1 Evaluation of the Discounting Factor

Let $\mathcal{T} = \{o_1, \ldots, o_n\}$ be a set of n objects and $\Theta = \{\theta_1, \theta_2, \ldots, \theta_p\}$ be a set of classes. For each object $o_j \in \mathcal{T}$, we know its class, denoted c_j, with $c_j \in \Theta$. The reading of the sensor about the class of the object o_j is represented by a bba, denoted $m^\Theta\{o_j\}$ defined on Θ.

Our first method for assessing the discounting factor considers one sensor alone. Finding the sensor discounting factor is achieved by comparing the bba's $m^\Theta\{o_j\}$ produced by the sensor about the class of each of the n objects in $\mathcal{T}$ and their real classes.

Assume the discounting factor α relative to this sensor is known, then its bba should be discounted taking into account α. We get $m^{\Theta,\alpha}\{o_j\}$ by using equation (3). In order to make a decision about the class to which the object belongs to, we apply the pignistic transformation on the bba $m^{\Theta,\alpha}\{o_j\}$. So, we get the pignistic probability, denoted $BetP^{\Theta,\alpha}\{o_j\}$ representing the probability of the object o_j to belong to each individual class. This probability function is then compared with what the user knows about the actual value c_j of the object o_j. Let the indicator function $\delta_{j,i}$ be defined as: $\delta_{j,i} = 1$ if $c_j = \theta_i$ and 0 otherwise. The idea is to compute the distance between the pignistic probability computed from the discounted sensor reading and the indicator function $\delta_{j,i}$. We have chosen the Euclidean distance, the choice of the distance is of course arguable. Other distances could of course be considered. The one we use is by far the most classical one. This distance is defined as follows:

$$Dist(o_j, \alpha) = \sum_{i=1}^{p} (BetP^{\Theta,\alpha}\{o_j\}(\theta_i) - \delta_{j,i})^2 \tag{5}$$

The following property holds for the above distance:

Theorem 1

$$0 \leq Dist(o_j, \alpha) \leq 2 \tag{6}$$

Proof. Without loss of generality, we can take $\delta_{j,i} = 1$ if $i = 1$, and 0 otherwise. We write P_i for $BetP^{\Theta,\alpha}\{o_j\}(\theta_i)$. So $Dist(o_j, \alpha) = (1 - P_1)^2 + \sum_{i=2}^{p} P_i^2 = \sum_{i=1}^{n} P_i^2 + 1 - 2P_1$. Being a sum of squares, it is non negative. We also have always $\sum_{i=1}^{n} P_i^2 \leq 1$ so $Dist(o_j, \alpha) \leq 2$ holds. The boundaries are reached with $P_1 = 1$, in which case $Dist(o_j, \alpha) = 0$, and with $P_1 = 0$ and $P_2 = 1$, in which case $Dist(o_j, \alpha) = 2$. □

The next step consists of computing the distance between the pignistic probability of each object o_j in $\mathcal{T}$ and its corresponding $\delta_{j,i}$. The sum of these distances, denoted $TotalDist$, reflects sensor reliability. $TotalDist$ is defined as follows:

$$TotalDist = \sum_{j=1}^{n} Dist(o_j, \alpha) = \sum_{j=1}^{n} \sum_{i=1}^{p} (BetP^{\Theta,\alpha}\{o_j\}(\theta_i) - \delta_{j,i})^2 \tag{7}$$

We then estimate the discounting factor as the coefficient $\alpha \in [0,1]$ that minimizes $TotalDist$. In other words, α makes the values of $BetP^{\Theta,\alpha}\{o_j\}$ as close as possible to the truth represented by $\delta_{j,i}$.

3.2 The Case of Normalized Belief Functions

In the case of normalized bba's, it is possible to produce the solution for the value of α that minimizes $TotalDist$. Let $BetP^{\Theta}\{o_j\}$ be the pignistic probability function computed from $m^{\Theta}\{o_j\}$, before discounting. The value of the factor α is presented in the following theorem.

Theorem 2 *Let a set of normalized bba's $m^{\Theta}\{o_j\}$ defined on the set of classes $\Theta = \{\theta_1, \theta_2, \ldots, \theta_p\}$ for objects $o_j, j = 1, \ldots, n$. Let the indicator function $\delta_{j,i} = 1$ if $c_j = \theta_i$ (and the object o_j belongs to the class c_j), and 0 otherwise. Let $BetP^{\Theta,\alpha}\{o_j\}$ be the pignistic probability function computed from the discounted bba $m^{\Theta,\alpha}\{o_j\}$. The discounting factor α that minimizes:*
$TotalDist = \sum_{j=1}^{n} \sum_{i=1}^{p} (BetP^{\Theta,\alpha}\{o_j\}(\theta_i) - \delta_{j,i})^2$ is given by:

$$\alpha = \min(1, \max(0, \frac{\sum_{j=1}^{n} \sum_{i=1}^{p} (\delta_{j,i} - BetP^{\Theta}\{o_j\}(\theta_i))BetP^{\Theta}\{o_j\}(\theta_i)}{n/p - \sum_{j=1}^{n} \sum_{i=1}^{p} BetP^{\Theta}\{o_j\}(\theta_i)^2})) \tag{8}$$

Proof. Given the bba $m^{\Theta}\{o_j\}$, its α-discounted bba is:

$$\begin{aligned} m^{\Theta,\alpha}\{o_j\}(\theta) &= (1-\alpha)m^{\Theta}\{o_j\}(\theta) && \text{if } \theta \subset \Theta \\ &= (1-\alpha)m^{\Theta}\{o_j\}(\Theta) + \alpha && \text{if } \theta = \Theta \end{aligned}$$

Hence, $BetP^{\Theta,\alpha}\{o_j\}$ computed from $m^{\Theta,\alpha}\{o_j\}$ can be defined as a function of $BetP^{\Theta}\{o_j\}$ which is computed directly from the initial bba $m^{\Theta}\{o_j\}$. We get $BetP^{\Theta}\{o_j\}(\theta_i) = \sum_{\theta:\theta_i \in \theta} m^{\Theta}\{o_j\}(\theta)/|\theta|$

$BetP^{\Theta,\alpha}\{o_j\}(\theta_i) = \sum_{\theta:\theta_i \in \theta} m^{\Theta,\alpha}\{o_j\}(\theta)/|\theta|$

$= \sum_{\theta:\theta_i \in \theta} (1-\alpha)m^{\Theta}\{o_j\}(\theta)/|\theta| + \alpha/p = (1-\alpha)BetP^{\Theta}\{o_j\}(\theta) + \alpha/p$

Let $P_{ij} = BetP^{\Theta}\{o_j\}(\theta_i)$. The value of $TotalDist$ to be minimized becomes $TotalDist = \sum_{j=1}^{n} \sum_{i=1}^{p} (BetP^{\Theta,\alpha}\{o_j\}(\theta_i) - \delta_{j,i})^2 = \sum_{j=1}^{n} \sum_{i=1}^{p} ((1-\alpha)P_{ij} + \alpha/p - \delta_{j,i})^2$

Its extremum is reached when its derivative is zero:

$$\begin{aligned} 0 = \frac{d\, TotalDist}{d\alpha} &= 2\sum_{j,i} ((1-\alpha)P_{ij} + \alpha/p - \delta_{j,i})(-P_{ij} + 1/p) \\ &\propto \sum_{j,i} -(1-\alpha)P_{ij}^2 - \alpha n/p + \sum_{j,i} \delta_{j,i}P_{ij} + (1-\alpha)n/p + \alpha n/p - n/p \\ &= \sum_{j,i} -(1-\alpha)P_{ij}^2 - \alpha n/p + \sum_{j,i} \delta_{j,i}P_{ij} \end{aligned}$$

Thus $\alpha = \frac{\sum_{j,i}(\delta_{j,i} - P_{ij})P_{ij}}{n/p - \sum_{j,i} P_{ij}^2}$. In order that $\alpha \in [0,1]$, we get the limit constraints. To prove that our solution is a minimum, the second derivative is shown to be positive.

We have:

$$\frac{d^2\ TotalDist}{d\alpha^2} \propto \sum_{j,i} P_{ij}^2 - n/p$$

For any probability distribution function $q_i : i = 1, \dots, p$, we have $q_i^2 \leq q_i, \forall i$, thus $\sum_i q_i^2 \leq \sum_i q_i = 1$. Furthermore the minima of $\sum_i q_i^2$ is reached when $q_i = 1/p, \forall i$ (obtained by minimizing $\sum_i q_i^2 - \lambda(\sum_i q_i - 1)$ where λ is a Lagrange multiplier). Coming back to the initial relations, we get $n \geq \sum_{i,j} P_{ij}^2 \geq n/p$, thus $\frac{d^2\ TotalDist}{d\alpha^2} \geq 0$, and the extremum for α is a minimum. □

Hence, in the case the sensor produces normalized bba's, we may easily find its discounting factor.

3.3 Example

Our example deals with a simple classification problem of aerial targets. Assume two sensors S_1, S_2 are applied to present their readings concerning the class of the detected target. Three classes are possible: $\Theta = \{Airplane, Helicopter, Rocket\}$. The sensor readings on the classes are expressed by the bba's detailed in table 1 where we consider 4 objects which classes are known by us but not by S_1 and S_2. At the first row of the table, we have the actual class of each object, then the bba's produced by the two sensors about the classes of the four objects.

Table 1: The bba's produced by two sensors.

Truth	Airplane	Helicopter	Airplane	Rocket
S_1	o_1	o_2	o_3	o_4
$\emptyset$	0	0	0	0
Airplane	0	0	0	0
Helicopter	0	0.5	0.4	0
Rocket	0.5	0.2	0	0
Airplane ∪ Helicopter	0	0	0	0
Airplane ∪ Rocket	0	0	0.6	0.6
Helicopter ∪ Rocket	0.3	0	0	0.4
Θ	0.2	0.3	0	0
S_2	o_1	o_2	o_3	o_4
$\emptyset$	0	0	0	0
Airplane	0	0.3	0.2	0
Helicopter	0	0	0	0
Rocket	0	0	0	0
Airplane ∪ Helicopter	0.7	0.4	0	0
Airplane ∪ Rocket	0	0	0	0
Helicopter ∪ Rocket	0	0	0.6	1
Θ	0.3	0.3	0.2	0

Let α_1 be the discounting factor applicable to sensor S_1. We discount bba's produced by S_1 relative to the four objects. Then, we compute their corresponding $BetP$'s (see table 2):

Next, the whole distance relative to the sensor S_1 will be equal to $TotalDist = \sum_{j=1}^{4} \sum_{i=1}^{3} (BetP^{\Theta,\alpha_1}\{o_j\}(\theta_i) - \delta_{j,i})^2$. Hence, $TotalDist = 0.41\alpha_1^2 - 0.54\alpha_1 + 2.83$.

Table 2: $BetP$'s computed from the four discounted bba's of S_1

S_1	o_1	o_2	o_3	o_4
Airplane	$0.06 + 0.26\alpha_1$	$0.1 + 0.24\alpha_1$	$0.3 + 0.03\alpha_1$	$0.3 + 0.03\alpha_1$
Helicopter	$0.22 + 0.12\alpha_1$	$0.6 - 0.27\alpha_1$	$0.4 - 0.06\alpha_1$	$0.2 + 0.13\alpha_1$
Rocket	$0.72 - 0.38\alpha_1$	$0.3 + 0.03\alpha_1$	$0.3 + 0.03\alpha_1$	$0.5 - 0.16\alpha_1$

Minimizing $TotalDist$ under the constraint $0 \leq \alpha_1 \leq 1$ results in $\alpha_1 = 0.66$. Applying the same procedure for readings produced by the sensor S_2, we get the discounting factor $\alpha_2 = 0.52$. Thus, sensor S_2 is (a little) better than the sensor S_1, in other words, it is (a little) more reliable than S_1.

3.4 Categorical Imprecise Sensors

Sensors are qualified as 'categorical' when there is only one focal element, so there is a $\tau_j \subseteq \Theta$ with $m^\Theta\{o_j\}(\tau_j) = 1$ for all objects o_j. Categorical sensors are 'precise' if $|\tau_j| = 1$ for all objects o_j.

We say that a sensor is 'blind' if the class τ_j is selected irrespective of the object o_j, i.e. the sensor allocates the object to a class at random, using the same distribution of every object, irrespective of true object class. It is 'unbiased' if the distribution gives equal probability for each class. A sensor is 'worth using' if its results are better than those a blind sensor would achieve.

We consider the case of a categorical but imprecise sensor, i.e. a sensor such that for all object o_j, there is a $\tau_j \subseteq \Theta$ with $m^\Theta\{o_j\}(\tau_j) = 1$ and $|\tau_j| \geq 1$, with a strict inequality for some objects. Thanks to the simplified structure of the data, it is possible to derive the value of the discounting factor to assign to the sensor.

Theorem 3 *Let $\Theta = \{\theta_1, \ldots, \theta_p\}$ be a set of p mutually exclusive and exhaustive classes, and $\{o_1, \ldots, o_n\}$ be a set of n objects. For each o_j, let c_j be the actual class of object o_j, and $\tau_j \subseteq \Theta$ be the subset of Θ in which all we know about c_j is $c_j \in \tau_j$. Let $n_{i,\theta} = |\{j : c_j = \theta_i, \tau_j = \theta\}|$ and $n_{.,\theta} = \sum_i n_{i,\theta}$. The value α that minimizes $TotalDist$ in relation (7) is given by:*

$$\alpha = \min(1, \frac{S - PCC}{S - 1/p}) \tag{9}$$

$$\textit{where } PCC = \sum_i \sum_{\theta \subseteq \overline{\theta}_i} \frac{n_{i,\theta_i \cup \theta}}{|\theta| + 1} \tag{10}$$

$$\textit{and } S = \frac{1}{n} \sum_{\emptyset \neq \theta \subseteq \Theta} \frac{n_{.,\theta}}{|\theta|} \tag{11}$$

Proof. Suppose α is the discounting factor. For every j, $m^\Theta\{o_j\}(\tau_j) = 1$ and its *alpha* discounted bba is $m^{\Theta,\alpha}\{o_j\}(\tau_j) = 1 - \alpha, m^{\Theta,\alpha}\{o_j\}(\Theta) = \alpha$. The pignistic probabilities computed from $m^{\Theta,\alpha}\{o_j\}$ are:

$$\begin{aligned} BetP^{\Theta,\alpha}\{o_j\}(\theta_i) &= \alpha/p + (1-\alpha)/r & \forall \theta_i \in \tau_j \\ &= \alpha/p & \forall \theta_i \notin \tau_j \end{aligned}$$

where $r = |\tau_j|$ and $p = |\Theta|$.

The distance $Dist(o_j, \alpha)$ becomes:

- If $c_j \in \tau_j$ (possibly exact classification): $-1 - \alpha^2/r + 1 - \alpha^2/p$.
- If $c_j \notin \tau_j$ (wrong classification): $(1-\alpha)^2/r + 1 - \alpha^2/p$.

The overall distance becomes:

$$\begin{aligned}
TotalDist &= \sum_i \Big(\sum_{\theta \subseteq \overline{\theta}_i} n_{i,\theta_i \cup \theta} \frac{-(1-\alpha^2)}{|\theta|+1} + \sum_{\theta \subseteq \overline{\theta}_i} n_{i,\theta} \frac{(1-\alpha)^2}{|\theta|}\Big) + n\Big(1 - \frac{\alpha^2}{p}\Big) \\
&= -(1-\alpha^2) \sum_i \sum_{\theta \subseteq \overline{\theta}_i} \frac{n_{i,\theta_i \cup \theta}}{|\theta|+1} + (1-\alpha)^2 \sum_i \sum_{\theta \subseteq \overline{\theta}_i} \frac{n_{i,\theta}}{|\theta|} + n\Big(1 - \frac{\alpha^2}{p}\Big) \\
&= -n(1-\alpha^2)PCC + n(1-\alpha)^2(S - PCC) + n(1 - \alpha^2/p)
\end{aligned}$$

where $PCC = \sum_i \sum_{\theta \subseteq \overline{\theta}_i} \frac{n_{i,\theta_i \cup \theta}}{|\theta|+1}$ and $S = \frac{1}{n} \sum_{\emptyset \neq \theta \subseteq \Theta} \frac{n_{.,\theta}}{|\theta|}$.

The value of α in [0,1] that minimizes $TotalDist$ is obtained by finding the value that makes the derivative equals to zero , and satisfies the domain constraints.

$$\frac{d\ TotalDist}{d\alpha} = 2\alpha n PCC - 2(1-\alpha)n(S - PCC) - 2n\alpha/p = 0.$$

Hence, $\alpha = \min(1, \frac{S-PCC}{S-1/p})$ which is a minima as the second derivative $S - 1/p$ is positive. Indeed with $m(\theta) = n_{.,\theta}/n$, $S = \sum_{\emptyset \neq \theta \subseteq \Theta} m(\theta)/|\theta| \geq \sum_{\emptyset \neq \theta \subseteq \Theta} m(\theta)/|\Theta| = 1/p$. □

The nature of the terms in relation (9) deserves consideration. Define $BetP^{\Theta}(\theta_i) = \sum_{\theta_i \in \theta \subseteq \Theta} \mu_i(\theta)/|\theta|$ where $\mu_i(\theta) = n_{i,\theta}/n_{i,.}$. By construction μ_i is a normalized bba's on Θ. The term $BetP^{\Theta}(\theta_i)$ is the probability with which you would bet correctly when the case is a θ_i case. It is the analogous of the PCC (Percent of Correct Classification) for the θ_i cases. This value would be the classical PCC if:

- all cases with $c_j = \tau_j$ are correctly classified,
- $1/m$ cases with $c_j \in \tau_j$ for $|\tau_j| = m$ are correctly classified.

A global PCC computed for the whole learning set can be defined in this context of imprecise data as $PCC = \sum_i \frac{n_{i,.}}{n} BetP^{\Theta}\{o_j\}(\theta_i)$, hence its name.

Using the same argument, S can be seen as the maximal number of objects that we can expect to classify correctly when data are imprecise.

Finally, the term $1/p$ can be seen as the expected number of objects correctly classified by a blind unbiased sensor.

The only case when the domain constraint is used is when $PCC < 1/p$, thus when the sensor is worse than a blind unbiased sensor that allocates objects on pure chance. When $\alpha < 1$, the sensor is worth using. When $\alpha = 1$, all bba's produced by the classifier are so discounted that they become vacuous. In that case why bother with such a sensor?

When the classifier is categorical and precise, i.e., when $|\tau_j| = 1$ for all objects o_j, then $S = 1$ and $\alpha = \min(1, \frac{1-PCC}{1-1/p})$.

3.5 Uncertainty about the Truth

Suppose the actual classes of the objects used to assess the discounting factors are not exactly known, but we only have a bba $m_0^\Theta\{o_j\}$ that expresses what we know about the true class of object o_j. The adaptation of the distance, in such a case, is immediate. Let $p_{j,k}$ be the value of the pignistic probability induced from the $m_0^\Theta\{o_j\}$ with $p_{j,k} = BetP_0^\Theta$ (c_j class of o_j is θ_k). If we knew that the class of o_j was θ_k, we would compute $Dist(o_j, \alpha) = \sum_{i=1}^p (BetP^{\Theta,\alpha}\{o_j\}(\theta_i) - \delta_{j,i})^2$. According to equation (5), where $\delta_{j,i} = 1$ if $c_j = \theta_i$, and 0 otherwise. The probability that the class of o_j was θ_k is $p_{j,k}$, so we weight this distance by $p_{j,k}$ and compute its expectation taken over k.

$$Dist(o_j, \alpha) = \sum_{k=1}^{p} p_{j,k} \sum_{i=1}^{p} (BetP^{\Theta,\alpha}\{o_j\}(\theta_i) - \delta_{j,i})^2 \tag{12}$$

We then proceed as previously (see section 3.1).

3.6 Comparing α and κ

A classical criterion for evaluating the classifier quality is the κ coefficient. This coefficient is normally defined for categorical and precise classifiers, as follows:

$$\begin{aligned}\kappa &= \frac{PCC - \text{ proportion correctly classified by chance}}{\text{max possible } PCC - \text{ proportion correctly classified by chance}}\\ &= \frac{\sum_i n_{i,\theta_i} - \sum_i n_{.,\theta_i} n_{i,.}/n}{n - \sum_i n_{.,\theta_i} n_{i,.}/n}\end{aligned}$$

Of course max possible PCC is 1 in the classical context.

Within our framework, the α coefficient can be used to generalize the meaning of κ in context where learning set data are imprecise and non categorical. Consider the following sensors:

- Suppose the sensor is categorical, precise and worth using. Further suppose the sensor allocates the same number of objects in each class. Then $1 - \alpha = \kappa$. This illustrates the direct link between the two coefficients.
- Suppose the sensor is categorical, imprecise and worth using. The proportion correctly classified by chance becomes S, what fits indeed with the explanation given for relation (11). So, α in relation (9) is a direct generalization of κ.
- Suppose the sensor is not categorical and worth using, then α is a further generalization of the previous case adapted to the non categorical cases.

In order to better understand what α means, we can propose another link that might help. For categorical and precise sensors, the PCC can be seen as the expected utility obtained from the use of the sensor when utilities are 1 for a correct decision ($c_j = \tau_j$), and 0 for a wrong decision. For the categorical imprecise sensor, use utilities $1/|\tau_j|$ when $c_j \in \tau_j$, thus when the sensor is not wrong, and 0 when it is wrong. The coefficient fits nicely with the natural idea that if for instance all we know is $\tau_j = \{\theta_1, \theta_2, \theta_3\}$, then such a decision is worth 1/3 as on the average one can expect that 1/3 of those objects classified as τ_j are correctly classified and 2/3 are wrongly classified. Relation (8) could be seen as generalizing this idea of expected utilities to non categorical sensors.

3.7 Unclassified Data

Suppose an object o_j has not been classified by the sensor, and still we add it to the data base to assess the discounting factor. It seems that this case should not interfere with the assessment of α. And so is it indeed. In that case, $BetP^{\Theta,\alpha}\{o_j\}(\theta_i) = 1/p$ whatever α. Equation (12) becomes:

$Dist(o_j, \alpha) = \sum_{k=1}^{p} p_{j,k} \sum_{i=1}^{p} (1/p - \delta_{j,i})^2$ which does not depend on α and therefore the evaluation of α will not be affected by these unclassified cases. This property results from the fact that the discounting of a vacuous belief function is the vacuous belief function itself.

3.8 Conclusions for one sensor

We have considered the case where data in the learning set are categorical precise, categorical imprecise, or uncertain. The computation of the discounting factor α is obtained by a minimization procedure. In the most classical case where bba's are normalized, explicit solutions are presented. The smaller the α, the best the sensor. The discounting factor can be applied to the beliefs generated by the sensor when facing new data. They can also be used to order sensors by their 'quality'.

4 Assessing the Discounting Factors of Several Sensors Used Jointly

4.1 Assessing the Discounting Factors

The second method developed in this paper permits the evaluation of the discounting factors when there are several sensors and their readings are aggregated. Pooling sensor readings together is done in order to derive, hopefully, a better predictor. This pooling is achieved by conjunctively combining the bba's produced by each sensor. Before combining them, they must be discounted appropriately in order to take into account their individual reliability. Let $m_{S_\nu}^{\Theta}\{o_j\}$ be the bba collected from sensor S_ν about the actual value of the class of object o_j, denoted c_j. In order to get the optimal set of discounting factors, the following steps are applied:

- Assume a set of discounting factors α_ν, one for each sencor.
- Discount $m_{S_\nu}^{\Theta}\{o_j\}$ by its discounting factor α_ν given to sensor S_ν. We get $m_{S_\nu}^{\Theta,\alpha_\nu}\{o_j\}$. This process is applied for each sensor and for each object.
- For each object o_j $(j = 1, ..., n)$, apply the conjunctive rule in order to compute the overall bba $m^{\Theta,\overline{\alpha}}\{o_j\}$ about the class to which o_j belongs.

$$m^{\Theta,\overline{\alpha}}\{o_j\} = m_{S_1}^{\Theta,\alpha_1}\{o_j\} \cap \text{.........} \cap m_{S_e}^{\Theta,\alpha_e}\{o_j\} \tag{13}$$

  where $\overline{\alpha}$ denotes the vector $(\alpha_1, \ldots, \alpha_e)$.
- Compute the corresponding $BetP^{\Theta,\overline{\alpha}}\{o_j\}$ representing the pignistic probability on the class of object o_j.

- For each object o_j, compute the distance $Dist(o_j, \overline{\alpha})$ between $BetP^{\Theta,\overline{\alpha}}\{o_j\}$ and $\delta_{j,i}$, the indicator function for the real class of o_j. $Dist(o_j, \overline{\alpha}) = \sum_{i=1}^{p}(BetP^{\Theta,\overline{\alpha}}\{o_j\}(\theta_i) - \delta_{j,i})^2$ where $\delta_{j,i} = 1$ if $c_j = \theta_i$ and 0 otherwise.
- Compute $TotalDist$ as follows:

$$TotalDist = \sum_{j=1}^{n} Dist(o_j, \overline{\alpha}) \tag{14}$$

  $TotalDist$ is expressed on the terms of the sensor discounting factors α_1, $\alpha_2, \ldots, \alpha_e$.
- To find the optimal discounting factors, we minimize $TotalDist$ on the α_ν's under the constraints $0 \leq \alpha_\nu \leq 1$, $\forall \nu \in \{1, .., e\}$

4.2 Example 2

Let us consider the same data in the example 1 (see table 1) but assume that the two sensor readings will be taken into account together. So, let's apply our second method on the two sensor readings in order to get their merged reading. Once the bba's of S_1 and S_2 are discounted, we get respectively $m_{S_1}^{\Theta,\alpha_1}\{o_j\}$ and $m_{S_2}^{\Theta,\alpha_2}\{o_j\}$ where $j = 1, 2, 3, 4$, which are linear functions of the discounting factors. For each object o_j, we compute the joint bba $m^{\Theta,\overline{\alpha}}\{o_j\}$ as : $m^{\Theta,\overline{\alpha}}\{o_j\} = m_{S_1}^{\Theta,\alpha_1}\{o_j\} \bigcirc\!\!\!\!\cap\, m_{S_2}^{\Theta,\alpha_2}\{o_j\}$ where the terms containing the $\alpha_i' s$ are at worst of the form $\prod_{\nu=1,\ldots,I} \alpha_\nu$ where I is the number of sensors ($I = 2$ in the present case). The corresponding discounted $BetP$'s relative to these bba's are also linear functions of the same product terms. The value of $Dist(o_j, \overline{\alpha})$ relative to the objects, as well as $TotalDist$, are quadratic functions of the previous product terms. So its minimization on the α_i is simple and can be achieved by any minimization program.

When we work with more than two sensors, the minimization program should produce the values of α_ν's. Minimization might become problematic when the number of sensors is large as it is non-linear in the alphas and of exponential size. This computational problem is not analyzed here as we focus on the principle, not on its implementation.

In the present case, $\alpha_1 = 0.28$ and $\alpha_2 = 0.12$. It should be emphasized that the discounting factors computed in this second method should not be compared to those computed with the first method described in section 3. Here we need the α_ν's such that the multisensor itself is 'optimal', whereas in the first method, we compute the α_ν's in order to evaluate the quality of the individual sensors taken individually. Mathematically, in the first case, we minimize for each sensor individually, whereas in the second case, we minimize the distance with the result of the combination of the discounted sensor belief functions.

4.3 Sensors Observing Different Data Sets

Suppose an object o_j has not been classified by the sensor S_ν. This is equivalent to using a vacuous bba for $m_{S_\nu}^{\Theta}\{o_j\}$. As before, these data do not interfere with the assessment of the α's. The term $m^{\Theta,\overline{\alpha}}\{o_j\}$ encountered in equation (13) is not changed as $m_{S_\nu}^{\Theta,\alpha_\nu}\{o_j\}$ is vacuous. The same holds for $BetP^{\Theta,\overline{\alpha}}\{o_j\}$, $Dist(o_j,\overline{\alpha})$ and $TotalDist$. Hence, the α's will be the same. The fact that the assessment of α's does not depend on the addition of any vacuous bba implies that we can apply the previous method to the case where the data sets observed by each sensor differ from sensors to sensors. One just considers all possible cases, add (fictiously) vacuous bba for all the missing bba's, and proceed as before. The only problematic case would be if S_ν had not observed any data, and produce only vacuous bba's. In that case $TotalDist$ becomes independent of α_ν, and any value would be as good as any other, as it should be indeed. How could we assess the quality of a sensor that does not report anything?

5 Conclusion

In this paper, we have presented two methods for assessing reliability factors of non ideal sensors. The first method treats the case where each sensor is considered alone and consists of finding the discounting factor that will make its readings as close as possible to reality. The second method treats the case where we have several sensors that must be used jointly in order to assess their discounting factors. The method can be adapted to handle partially known data where the user is not sure about the actual values of classes. We can also apply the method to a case where the sets of objects used for assessing the reliability of the various sensors varies from sensor to sensor. The methods we present can easily be extended to other problems of prediction in contexts of supervised learning. All it requires is a 'distance' between the sensor readings and the reality. The technique consists of finding the discounting factors that will minimize this distance.

Acknowledgments

The authors would like to thank Thierry Denoeux for his interest in our work and for his help. Many thanks are also due to anonymous referees for their detailed comments, criticisms, and suggestions that have helped greatly in improving this paper.

References

[1] I. Bloch. Information combination operator for data fusion: a comparative review with classification. *IEEE Trans. SMC*, 26:52–67, 1996.

[2] D. M. Buede and P. Girardi. A target identification comparison of bayesian and Dempster-Shafer multisensor fusion. *IEEE Trans. SMC: A*, 27:569–577, 1997.

[3] R. M. Cooke. *Experts in Uncertainty.* Oxford Univ. Press, New York, 1991.

[4] Z. Elouedi, K. Mellouli, and P. Smets. The evaluation of sensors' reliability and their tuning for multisensor data fusion within the transferable belief

model. In *Proceedings of the 6th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 350–361. ECSQARU 2001, Springer LNAI2143, 2001.

[5] N. Milisavljevic. *Analysis and Fusion Using Belief Functions Theory of Multisensor Data for Close-Range Humanitarian Mine Detection.* PhD thesis, Ecole Normale Supérieure de Télécomminications, Paris, France, 2001.

[6] N. Milisavljevic, I. Bloch, and M. Acheroy. Modeling, combining and discounting mine detection sensors within Dempster-Shafer framework. In *Detection Technologies for Mines and Minelike Targets*, volume 4038, pages 1461–1472. SPIE Press, Orlando, USA, 2000.

[7] G. Shafer. *A Mathematical Theory of Evidence.* Princeton Univ. Press. Princeton, NJ, 1976.

[8] P. Smets. The combination of evidence in the transferable belief model. *IEEE Pattern analysis and Machine Intelligence*, 12:447–458, 1990.

[9] P. Smets. Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. *International Journal of Approximate Reasoning*, 9:1–35, 1993.

[10] P. Smets. The transferable belief model for quantified belief representation. In D. M. Gabbay and P. Smets, editors, *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, volume 1, pages 267–301. Kluwer, Doordrecht, The Netherlands, 1998.

[11] P. Smets. Decision making in a context where uncertainty is represented by belief functions. In R. P. Srivastava and T. J. Mock, editors, *Belief Functions in Business Decisions.* Physica-Verlag, Heidelberg, 2002.

[12] P. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–234, 1994.

[13] L. M. Zouhal and T. Denœux. An evidence theoretic k-nn rule with parameter optimisation. *IEEE Transactions on Systems, Man and Cybernetics - Part C*, 28:263–271, 1998.