Refined sensor tuning in the belief function framework using contextual discounting

David Mercier
Université de Technologie
de Compiègne
UMR CNRS 6599 Heudiasyc
Compiègne, France
Email: dmercier@hds.utc.fr

Thierry Denœux
Université de Technologie
de Compiègne
UMR CNRS 6599 Heudiasyc
Compiègne, France
Email: tdenoeux@hds.utc.fr

Marie-Hélène Masson
Université de Picardie Jules Verne
UMR CNRS 6599 Heudiasyc
Compiègne, France
Email: mmasson@hds.utc.fr

Abstract

This paper presents an objective way of assessing the reliability of a sensor or an expert expressing its opinion by way of a belief function. Using the contextual discounting, labeled data and an error function, we generalize an approach proposed by Elouedi, Mellouli and Smets (2004). Applied to a single sensor, the result of this new sensor tuning is a richer learning process allowing to learn the reliability of a sensor for different situations to be recognized, and the possibility to compare sensors on each situation. Applied to improve the combination of a group of sensors, it allows to create some refined combination by automatically detecting strenghs and weaknesses of each sensor.

Keywords: Assessing sensor reliability, improving a combination of sensors, belief function, contextual discounting.

1 Introduction

Multisensor fusion, in the context of pattern recognition, aims at combining information obtained from different sensors in order to achieve better recognition results. To reach this goal, it is usually of great importance to take into account the reliability of the different sources in the evidence aggregation process [2, 4, 5, 7, 12]. In practice, the information about the reliability of each source can either come from expert knowledge or be inferred from data. In this paper the latter approach, as refered to *expert tuning*, is considered and is addressed using the Transferable Belief Model (TBM) [14], an interpretation of the Dempster-Shafer theory of belief functions [8, 11], a powerful and flexible framework for representing and reasoning with various forms of imperfect information and knowledge.

The knowledge about the reliability of a source of information is classically achieved by the discounting operation introduced by Shafer [8], which transforms each belief function provided by a source into a weaker, less informative one. In this *classical discounting operation*, knowledge about sensor or expert reliability is described by a single number: the discount rate α taking values between 0 and 1, and the value $\beta = (1 - \alpha)$ quantifies the degree of belief in the fact that the source is reliable [10]. In certain cases, however, more refined knowledge is available. In particular, the reliability of the source of information can be expected to depend on the actual value of the variable of interest. In medical diagnosis, for instance, a physician may be, due to his/her past experience or training, particularly competent to diagnose some types of diseases, while being less competent for other types. In target recognition, the performances of a data acquisition system may depend not only on weather conditions, but also on background and target properties [1, 4], making the reliability of the decision system dependent on the target at hand. Thus, in [6], a generalization of this discounting operation has been proposed, allowing to use more detailed information regarding the reliability of the source in different contexts, i.e., conditionally on different hypotheses regarding the variable of interest. This *contextual discounting operation* is parameterized by a vector of discount rates representing degrees of belief in sensor reliability, conditionally on the variable of interest taking certain values or sets of values. These discount rates thus allow to express metaknowledge about sensor reliability, at a level of detail chosen by the user. The classical discounting operation is recovered as a special case and corresponds to coarse grained knowledge of sensor reliability.

In [3], a method for learning classical discount rates has been proposed. This method is based on the determination of the discount rates which minimize a distance between sensors decisions and the truth.

In this paper, a method generalizing this approach to the contextual discounting is presented allowing to learn more refined knowledge with a user-defined granularity. A new distance between sensors opinion and reality, more suitable for computation is also presented. Like [3], two cases are distinguished: learning the discount rate of a single sensor, such its decision is improved, and learning discount rates of a group of sensors, such that the whole combination of these sensors is improved.

The rest of this paper is organized as follows. Background material on the TBM is first recalled in Section 2. The general notion of Θ-contextual discounting, its main properties usefull for the expert tuning and in particular its link with the classical discounting are summarized in Section 3. The problem of learning discount rates for a single sensor and for a group of sensors, is then addressed in Section 4, and Section 5 concludes the paper.

2 The Transferable Belief Model: basic concepts

Let x be a variable taking values in a finite set $\Omega = \{\omega_1, \ldots, \omega_K\}$, called the frame of discernment. The knowledge held by a rational agent Y, regarding the actual value ω_0 taken by x, given an evidential corpus EC, can be quantified by a basic belief assignment (BBA) $m_Y^\Omega[EC]$, defined as a function from 2^Ω to $[0,1]$ verifying:

$$\sum_{A \subseteq \Omega} m_Y^\Omega[EC](A) = 1 \ .$$

When there is no ambiguity, the full notation $m_Y^\Omega[EC]$ will be simplified to m_Y^Ω, m^Ω, or even m. The vacuous BBA, defined by $m(\Omega) = 1$, represents total ignorance. Note that, in the TBM, BBAs are not required to be normalized, i.e., we may have $m(\emptyset) > 0$. The interpretation of $m(\emptyset)$ is discussed in [9].

The belief and plausibility functions associated with a BBA are defined, respectively, as:

$$bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B),$$

and

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

These functions play a central role in the TBM as they have easy interpretation: $bel(A)$ is interpreted as a degree of justified support given to proposition A by the available evidence, whereas $pl(A)$ is a measure of the maximum potential support that could be given to A, if further evidence became available. Two other functions which do not have such simple interpretation but are technically useful are the implicability function, defined as $b(A) = bel(A) + m(\emptyset)$, $\forall A \subseteq \Omega$, and the commonality function, defined as

$$q(A) = \sum_{B \supseteq A} m(B), \quad \forall A \subseteq \Omega.$$

Two distinct pieces of evidence, quantified by BBAs m_1 and m_2, may be combined using the *conjunctive rule of combination* (CRC) or the *disjunctive rule of combination* (DRC), defined, respectively, $\forall A \subseteq \Omega$ as:

$$\begin{aligned} m_{1 \cap\!\!\!\bigcirc 2}(A) &= \sum_{B \cap C = A} m_1(B) m_2(C) \ , \\ m_{1 \cup\!\!\!\bigcirc 2}(A) &= \sum_{B \cup C = A} m_1(B) m_2(C) \ . \end{aligned}$$

The CRC applies when both sources are known to be reliable, whereas the DRC corresponds to the hypothesis that at least one of the two sources is reliable [10]. These two operations can be very conveniently expressed using the commonality and implicability functions. The commonality function $q_{1 \textcircled{\cap} 2}$ associated with $m_{1 \textcircled{\cap} 2}$ can be obtained from q_1 and q_2, the commonality functions associated with m_1 and m_2, by pointwise multiplication: $q_{1 \textcircled{\cap} 2}(A) = q_1(A)q_2(A)$, for all $A \subseteq \Omega$. Similarly, we have $b_{1 \textcircled{\cup} 2}(A) = b_1(A)b_2(A)$, for all $A \subseteq \Omega$.

3 Discounting

Introduced on intuitive grounds by Shafer in [8, page 252], the classical discounting operation allows to take into account the reliability of a source of information. Its expression was given by:

$$\begin{cases} {}^{\alpha}bel^{\Omega}(\Omega) & = & 1, \\ {}^{\alpha}bel^{\Omega}(A) & = & \beta \, bel_S^{\Omega}(A), \ \forall A \subset \Omega, \end{cases}$$

where β is the degree of trust in the evidence as a whole.

As shown by Smets [10], the discounting operation is not *ad hoc*, but it can be derived from a simple model of sensor reliability. In this model, a source can be in two states: reliable or not. If it is known that the source is reliable (R), then the belief function it provides is accepted without any modification. Else, if it is known that the source is fully unreliable (NR), then the belief function it provides is totally discarded. In practice, we do not know for sure if the source is reliable, but we have some degrees of belief in the reliability of the source. As arguing in the introduction of this paper, the degree of confidence in the reliability of a source can depend on the true value of the variable of interest: each source may be more or less a specialist on the determination of some elements of Ω. So, the Θ-contextual discounting, allowing to take into account hypothesis on the reliability of the source on any coarsening Θ of Ω, has been developped and justified in [6].

The following subsection exposes the general expression of the contextual discounting, called Θ-*contextual discounting*, and its main properties usefull for the expert tuning.

3.1 General expression of the contextual discounting: Θ-contextual discounting

Let $\Theta = \{\theta_1, \ldots, \theta_L\}$ be a coarsening of Ω, which means that $\theta_1, \ldots, \theta_L$ form a partition of Ω.

Let us assume that Y receives a BBA m_S^{Ω} from a source S, describing the source's beliefs regarding the actual value ω_0 of the variable of interest x. Moreover, let us assume that Y has evidence regarding the reliability of a source S, conditionally on each $\theta_\ell \in \Theta$, i.e., in a context where ω_0 is known to be a member of θ_ℓ. We thus have L conditional BBAs $m_Y^{\mathcal{R}}[\theta_\ell]$, $\ell = 1, \ldots, L$, expressed on $\mathcal{R} = \{R, NR\}$, such that:

$$\begin{cases} m_Y^{\mathcal{R}}[\theta_\ell](\{R\}) & = & \beta_\ell, \\ m_Y^{\mathcal{R}}[\theta_\ell](\mathcal{R}) & = & \alpha_\ell, \end{cases} \tag{1}$$

with $\beta_\ell = 1 - \alpha_\ell$. Thus, β_ℓ represents our degree of belief that the source is reliable, when it is known that $\omega_0 \in \theta_\ell$.

For any $A \subseteq \Omega$, let

$$A_* = \bigcup_{\{\theta \in \Theta, \ \theta \subseteq A\}} \theta,$$

$$A^* = \bigcup_{\{\theta \in \Theta, \ \theta \cap A \neq \emptyset\}} \theta,$$

and

$$\mathcal{C} = \{A \subseteq \Omega \mid \exists I \subseteq \{1, \ldots, L\}, A = \bigcup_{i \in I} \theta_i\}.$$

A_* and A^* are thus, respectively, the largest element of $\mathcal{C}$ included in A, and the smallest element of $\mathcal{C}$ that contains A. A_* and A^* are respectively called in [8, page 117], the *inner and outer reduction* of A.

With these notations, the contextually discounted BBA ${}^{\boldsymbol{\alpha}}m_Y^{\Omega}$ with discount rate vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_L) \in [0, 1]^L$, is given, $\forall A \subseteq \Omega$, by:

$$\begin{aligned} &{}^{\boldsymbol{\alpha}}m^{\Omega}(A) \\ &= \prod_{\cup \theta_\ell = \overline{A}^*} \beta_\ell \sum_{\substack{B \subseteq A, \\ \exists C \in \mathcal{C}, \ B \cup C = A}} m_S^{\Omega}(B) \prod_{\cup \theta_k = (A \setminus B)^*} \alpha_k \end{aligned} \tag{2}$$

or, equivalently,

$$^{\boldsymbol{\alpha}}m^{\Omega}(A) = \sum_{B \subseteq A} {}^{\boldsymbol{\alpha}}G(A,B)\, m_S^{\Omega}(B), \qquad (3)$$

with:

$$^{\boldsymbol{\alpha}}G(A,B) = \begin{cases} 1 & \text{if } B = A = \Omega, \\ \displaystyle\prod_{\cup\theta_k = (A \setminus B)^*} \alpha_k \prod_{\cup\theta_\ell = \overline{A}^*} \beta_\ell & \text{if } B \in \mathcal{D}(A), \\ 0 & \text{otherwise,} \end{cases}$$

where, by convention, a product of terms vanishes if the index set is empty, and $\mathcal{D}(A) = \{B \subset \Omega \mid B \subseteq A \text{ and } \exists C \in \mathcal{C}, B \cup C = A\}$.

$G(A,B)$ is equal to the fraction of m_S^{Ω} which is tranferred to A, for each $A \supseteq B$. The whole set of such coefficients define a generalization matrix [13].

Example 1 Let us consider the case $K = 3$, and the following coarsening of Ω: $\Theta = \{\theta_1, \theta_2\}$ with $\theta_1 = \{\omega_1\}$, $\theta_2 = \{\omega_2, \omega_3\}$. Let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$. Assuming the subsets of Ω to be arranged in binary order, as suggested in [13], (e.g., for $K = 3$: $\emptyset$, $\{\omega_1\}$, $\{\omega_2\}$, $\{\omega_1, \omega_2\}$, $\{\omega_3\}$,$\{\omega_1, \omega_3\}$,$\{\omega_2, \omega_3\}$, Ω), the generalization matrix $^{\boldsymbol{\alpha}}\boldsymbol{G}$ has the following form:

$$^{\boldsymbol{\alpha}}\boldsymbol{G} = \begin{pmatrix} \beta_1\beta_2 & & & & & & & \\ \alpha_1\beta_2 & \beta_2 & & & & & & \\ & & \beta_1\beta_2 & & & & & \\ & & \alpha_1\beta_2 & \beta_2 & & & & \\ & & & & \beta_1\beta_2 & & & \\ & & & & \alpha_1\beta_2 & \beta_2 & & \\ \beta_1\alpha_2 & & \beta_1\alpha_2 & & \beta_1\alpha_2 & & \beta_1 & \\ \alpha_1\alpha_2 & \alpha_2 & \alpha_1\alpha_2 & \alpha_2 & \alpha_1\alpha_2 & \alpha_2 & \alpha_1 & 1 \end{pmatrix}$$

Roughly, the interpretation of this discounting can be summarized as follows: the mass on each element A, remains on A proportionally to the source reliabilities for the detection of elements of $\overline{A}$ and is transferred to $A \cup B$ proportionally to the lack of information on the source reliabilities for the detection of elements of B (with $A \cap B = \emptyset$, and $A \cup B \in \mathcal{C}$).

Let $^{\boldsymbol{\alpha}}b_Y^{\Omega}$ and $^{\boldsymbol{\alpha}}pl_Y^{\Omega}$ be, respectively, the implicability and plausibility functions associated with $^{\boldsymbol{\alpha}}m_Y^{\Omega}$. They can be obtained from b_S^{Ω} and pl_S^{Ω}, the corresponding functions associated with m_S^{Ω}, and their expressions are:

$$^{\boldsymbol{\alpha}}b_Y^{\Omega}(A) = b_S^{\Omega}(A) \prod_{\cup\theta_\ell = \overline{A}^*} \beta_\ell \qquad (4)$$

$$^{\boldsymbol{\alpha}}pl_Y^{\Omega}(A) = 1 - (1 - pl_S^{\Omega}(A)) \prod_{\cup\theta_\ell = A^*} \beta_\ell \qquad (5)$$

The cases $\Theta = \Omega$ and $\Theta = \{\Omega\}$, corresponding respectively as the finest and the coarsest knowledge on the conditional reliability of the source, are particularly interesting. The first one expresses the contextual discounting with the maximum of information about the reliability of the source, the second with the minimum.

3.2 Contextual discounting expressed on $\Theta = \Omega$: Ω-contextual discounting

If $\Theta = \Omega = \{\omega_1, \dots, \omega_K\}$, equations (2) and (3) allow to express the contextually discounted BBA $^{\boldsymbol{\alpha}}m_Y^{\Omega}$ with discount rate vector $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K) \in [0,1]^K$ by:

$$^{\boldsymbol{\alpha}}m^{\Omega}(A) = \prod_{\omega_\ell \in \overline{A}} \beta_\ell \sum_{B \subseteq A} \left(\prod_{\omega_k \in (A \setminus B)} \alpha_k \right) m_S^{\Omega}(B) \qquad (6)$$

or, equivalently,

$$^{\boldsymbol{\alpha}}m^{\Omega}(A) = \sum_{B \subseteq A} {}^{\boldsymbol{\alpha}}G(A,B) m_S^{\Omega}(B), \qquad (7)$$

with:

$$^{\boldsymbol{\alpha}}G(A,B) = \begin{cases} 1 & \text{if } B = A = \Omega, \\ \displaystyle\prod_{\omega_k = (A \setminus B)} \alpha_k \prod_{\omega_\ell = \overline{A}} \beta_\ell & \text{if } B \subseteq A, \\ 0 & \text{otherwise.} \end{cases}$$

where, by convention, a product of terms vanishes if the index set is empty.

Example 2 As in Example 1, let us consider the case $K = 3$, the generalization matrix $^{\boldsymbol{\alpha}}\boldsymbol{G}$ associated with the Ω-contextual discounting is defined by

$$
{}^{\boldsymbol{\alpha}}\boldsymbol{G} = \begin{pmatrix}
\beta_1\beta_2\beta_3 & & & & & & & \\
\alpha_1\beta_2\beta_3 & \beta_2\beta_3 & & & & & & \\
\beta_1\alpha_2\beta_3 & & \beta_1\beta_3 & & & & & \\
\alpha_1\alpha_2\beta_3 & \alpha_2\beta_3 & \alpha_1\beta_3 & \beta_3 & & & & \\
\beta_1\beta_2\alpha_3 & & & & \beta_1\beta_2 & & & \\
\alpha_1\beta_2\alpha_3 & \beta_2\alpha_3 & & & \alpha_1\beta_2 & \beta_2 & & \\
\beta_1\alpha_2\alpha_3 & & \beta_1\alpha_3 & & \beta_1\alpha_2 & & \beta_1 & \\
\alpha_1\alpha_2\alpha_3 & \alpha_2\alpha_3 & \alpha_1\alpha_3 & \alpha_3 & \alpha_1\alpha_2 & \alpha_2 & \alpha_1 & 1
\end{pmatrix}
$$

The expressions of ${}^{\boldsymbol{\alpha}}b_Y^{\Omega}$ and ${}^{\boldsymbol{\alpha}}pl_Y^{\Omega}$ becomes:

$$
{}^{\boldsymbol{\alpha}}b_Y^{\Omega}(A) = b_S^{\Omega}(A) \prod_{\omega_k \in \overline{A}} \beta_k \tag{8}
$$

$$
{}^{\boldsymbol{\alpha}}pl_Y^{\Omega}(A) = 1 - (1 - pl_S^{\Omega}(A)) \prod_{\omega_k \in A} \beta_k \tag{9}
$$

3.3 Contextual discounting expressed on $\Theta = \{\Omega\}$

If $\Theta = \{\Omega\}$, equations (2) and (3) allow to express the contextually discounted BBA ${}^{\boldsymbol{\alpha}}m_Y^{\Omega}$ with only one discount rate $\boldsymbol{\alpha} = (\alpha) \in [0, 1]$ by:

$$
{}^{\boldsymbol{\alpha}}m^{\Omega}(A) = \begin{cases} \beta m_S^{\Omega}(A) & \text{if } A \subset \Omega, \\ \beta m_S^{\Omega}(\Omega) + \alpha & \text{if } A = \Omega, \end{cases}
$$

which is the classical expression of the discounting expressed by Shafer. It can also be expressed using the generalization matrix:

$$
{}^{\boldsymbol{\alpha}}m^{\Omega}(A) = \sum_{B \subseteq A} {}^{\boldsymbol{\alpha}}G(A, B) m_S^{\Omega}(B), \tag{10}
$$

with:

$$
{}^{\boldsymbol{\alpha}}G(A, B) = \begin{cases} 1 & \text{if } B = A = \Omega, \\ \beta & \text{if } B = A \neq \Omega, \\ \alpha & \text{if } B \subset A \text{ and } A = \Omega. \\ 0 & \text{otherwise.} \end{cases}
$$

Example 3 As in Example 1, let us consider the case $K = 3$, the generalization matrix ${}^{\boldsymbol{\alpha}}\boldsymbol{G}$ associated with the $\{\Omega\}$-contextual discounting is defined by

$$
{}^{\boldsymbol{\alpha}}\boldsymbol{G} = \begin{pmatrix}
\beta & & & & & & & \\
& \beta & & & & & & \\
& & \beta & & & & & \\
& & & \beta & & & & \\
& & & & \beta & & & \\
& & & & & \beta & & \\
& & & & & & \beta & \\
\alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & 1
\end{pmatrix}
$$

The classical discounting operation is then a particular case of the contextual discounting where the reliability of the source is only known conditionnally to the fact that x takes its values in Ω. So, when a single global reliability of the source is only available, the lake of knowledge of contextual reliabilities leads to the classical discounting.

The expressions of generalization matrices in Examples 1, 2 and 3 allows to enhance the mass transfers more or less fine according to the granularity of the chosen coarsening Θ. Therefore, instead of learning a global reliability rate as the classical discounting does (i.e. a generalization matrix like Example 3), it would be convenient to be able to learn more precise pieces of information about any sensor or expert reliability (like the generalization matrices of Examples 1 and 2). This is made possible by the next section.

4 Learning Discount Rates

4.1 Learning the reliability of a single sensor

Following [3], let us assume that we have a training set related to n objects $o_1, \ldots, o_n$. Each object o_i belongs to a class in $\Omega = \{\omega_1, \ldots, \omega_K\}$. The class of object o_i is encoded by K binary indicator variables $\delta_{i,k}$, with $\delta_{i,k} = 1$ if object o_i belongs to class ω_k, and $\delta_{i,k} = 0$ otherwise.

Assume that a sensor or expert S provides, for each object o_i, a BBA $m_S^{\Omega}\{o_i\}$, quantifying its belief concerning the class of object o_i. In the expert tuning approach, a scalar discount rate α is found so as to minimize the following measure of discrepancy between beliefs and observations:

$$
\begin{aligned}
& E_{bet}(\alpha) \\
&= \sum_{i=1}^{n} \sum_{k=1}^{K} ({}^{\alpha}BetP^{\Omega}\{o_i\}(\omega_k) - \delta_{i,k})^2.
\end{aligned} \tag{11}
$$

where ${}^{\alpha}BetP^{\Omega}\{o_i\}$ is the pignistic probability distribution [14] associated with ${}^{\alpha}m^{\Omega}$, defined as:

$$\begin{aligned} & {}^{\alpha}BetP^{\Omega}\{o_i\}(\omega_k) \qquad (12)\\ = & \sum_{\{A\subseteq\Omega,\omega_k\in A\}} \frac{{}^{\alpha}m^{\Omega}\{o_i\}(A)}{(1-m^{\Omega}\{o_i\}(\emptyset))|A|}. \end{aligned}$$

The minimization of $E_{bet}(\alpha)$ with respect to $\alpha \in [0,1]$, is a constrained nonlinear programming problem. When all BBAs are normalized, i.e., $m^{\Omega}\{o_i\}(\emptyset) = 0$ for all i, then $E_{bet}(\alpha)$ is a quadratic function of α, whose minimum can be found analytically, as shown in [3].

The above expert tuning approach can easily be generalized to contextual discounting. In this section, only Ω-contextual discounting will be considered. Indeed, the same approach can be extended to Θ-contextual discounting, once a coarsening Θ has been chosen. Let $E_{bet}(\boldsymbol{\alpha})$ be the same error function as defined in (11), where contextual discounting with discount rate vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$ is used in place of classical discounting. This is now a nonlinear K-dimensional function, whose minimization with respect to $\boldsymbol{\alpha}$ under the constraints $0 \le \alpha_k \le 1$ can be achieved using a constrained nonlinear minimization procedure. Eq (11) has the drawback to penalize in the same way an expert expressing his belief as a Bayesian bba (uniform probability) and an expert using the vacuous belief function. An alternative error function which leads to a simpler optimization problem could be:

$$\begin{aligned} & E_{pl}(\boldsymbol{\alpha}) \qquad (13)\\ = & \sum_{i=1}^{n}\sum_{k=1}^{K} ({}^{\boldsymbol{\alpha}}pl^{\Omega}\{o_i\}(\{\omega_k\}) - \delta_{i,k})^2. \end{aligned}$$

Indeed, from (9), we can see that ${}^{\boldsymbol{\alpha}}pl^{\Omega}\{o_i\}(\{\omega_k\})$ is an affine function of α_k:

$$\begin{aligned} & {}^{\boldsymbol{\alpha}}pl^{\Omega}\{o_i\}(\{\omega_k\})\\ = & 1 - (1 - pl^{\Omega}\{o_i\}(\{\omega_k\}))\beta_k\\ = & \alpha_k(1 - pl^{\Omega}\{o_i\}(\{\omega_k\})) + pl^{\Omega}\{o_i\}(\{\omega_k\}). \end{aligned}$$

Equation (13) can be written in matrix form as follows. Let

$$\boldsymbol{pl}_i = (pl^{\Omega}\{o_i\}(\{\omega_1\}), \ldots, pl^{\Omega}\{o_i\}(\{\omega_K\}))^T$$

denote the column vector containing the plausibilites of the singletons for object o_i, $\mathrm{diag}(1-\boldsymbol{pl}_i)$ the diagonal matrix whose diagonal elements are the complements to one of the components of $\boldsymbol{pl}_i$, and $\boldsymbol{\delta}_i = (\delta_{i,1}, \ldots, \delta_{i,K})^T$ the column vector of 0-1 class indicator variables for object o_i. Let

$$\boldsymbol{Q} = \begin{bmatrix} \mathrm{diag}(1-\boldsymbol{pl}_1)\\ \vdots \\ \mathrm{diag}(1-\boldsymbol{pl}_K) \end{bmatrix}, \ \boldsymbol{d} = \begin{bmatrix} \boldsymbol{\delta}_1 - \boldsymbol{pl}_1\\ \vdots \\ \boldsymbol{\delta}_K - \boldsymbol{pl}_K \end{bmatrix}.$$

Then (13) can be writen as

$$E_{pl}(\boldsymbol{\alpha}) = \|\boldsymbol{Q}\boldsymbol{\alpha} - \boldsymbol{d}\|^2. \qquad (14)$$

It is clear from the above formulation that the minimization of $E_{pl}(\boldsymbol{\alpha})$ is a constrained least-squares problem, for which efficient algorithms exist.

Note that a same weight is given in (13) to opposite situations: when a plausibility of 0 is assigned to the true class and when a plausibility of 1 is assigned to a wrong class. According to the application, this symetric behavior of the error function could be undesirable. A manner of correcting this behaviour consists in modifying the indicators $\delta_{i,k}$ between 0 and 1. These values can depend on i and k, so indicators $\delta_{i,k}$ play the same role as costs of error dependent on the application (some errors can cost less than others). In the sequel, only eq (13) will be considered.

Example 4 As an example, consider the data of Table 1, taken from [3]. This is a simplified aerial target recognition problem, in which we have three classes: airplane (a), helicopter (h) and rocket (r). Let $\Omega = \{a, h, r\}$. Four objects with known classification have been classified by two sensors S_1 and S_2. Each sensor has provided, for each objet, a BBA on Ω, as shown in Table 1. As reported in [3], the expert tuning approach (with a single discount rate and minimization of (11)) yields $\alpha = 0.66$ for sensor S_1, and $\alpha = 0.52$ for sensor S_2. With our approach (three contextual discount rates, $\Theta = \Omega$ like Example 2, and minimization of (14)), we obtain $\boldsymbol{\alpha} = (0.24, 0, 0)$ for sensor S_1, and $\boldsymbol{\alpha} = (0.26, 0, 0)$ for sensor S_2. Thus, we are able to learn

Table 1: Data of Example 4, taken from [3].

	a	h	r	$\{a,h\}$	$\{a,r\}$	$\{h,r\}$	Ω	truth
$m_{S_1}\{o_1\}$	0	0	0.5	0	0	0.3	0.2	a
$m_{S_1}\{o_2\}$	0	0.5	0.2	0	0	0	0.3	h
$m_{S_1}\{o_3\}$	0	0.4	0	0	0.6	0	0	a
$m_{S_1}\{o_4\}$	0	0	0	0	0.6	0.4	0	r
$m_{S_2}\{o_1\}$	0	0	0	0.7	0	0	0.3	a
$m_{S_2}\{o_2\}$	0.3	0	0	0.4	0	0	0.3	h
$m_{S_2}\{o_3\}$	0.2	0	0	0	0	0.6	0.2	a
$m_{S_2}\{o_4\}$	0	0	0	0	0	1	0	r

richer information on sensors' reliabilities: we have learnt their reliabilities for each category of aerial target. With a larger training set, such a result would indicates troubles with the recognition of airplanes for both sensors, and that S_1 is slightly more reliable that S_2 (smaller values of α_k correspond to higher reliability). However, it is clear that such an interpretation is of limited validity with such a small training set, which is only used here as an illustration.

4.2 Learning the reliability of a group of sensors

When we have several sensors providing independent measurements, a usual strategy is to discount each sensor, and then combine them conjunctively using the CRC. In such a case, it seems preferable to optimize the performance of the combination, instead of optimizing the performance of each sensor individually.

For the sake of simplicity, assume that we have two sensors S_1 and S_2. Let ${}^{\boldsymbol{\alpha}_1}pl^{\Omega}_{S_1}\{o_i\}$ be the plausibility provided by sensor S_1 discounted with rates $\boldsymbol{\alpha}_1$, and ${}^{\boldsymbol{\alpha}_2}pl^{\Omega}_{S_2}\{o_i\}$ the plausibility provided by sensor S_2 discounted with rates $\boldsymbol{\alpha}_2$. After conjunctive combination, the plausibility of each singleton is equal to the product of the plausibilities given by the two sources (this results from the fact that the plausibility of a singleton is equal to its commonality). The plausibility of $\{\omega_k\}$ is thus equal to

$$\begin{aligned}&{}^{\boldsymbol{\alpha}_{12}}pl^{\Omega}\{o_i\}(\{\omega_k\})\\ =\ &{}^{\boldsymbol{\alpha}_1}pl^{\Omega}_{S_1}\{o_i\}(\{\omega_k\}) \times {}^{\boldsymbol{\alpha}_2}\, pl^{\Omega}_{S_2}\{o_i\}(\{\omega_k\}).\end{aligned}$$

Discount rate vectors $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ can thus be determined so as to minimize the following error function:

$$\begin{aligned}&E_{pl}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2)\\ =\ &\sum_{i=1}^{n}\sum_{k=1}^{K}\left({}^{\boldsymbol{\alpha}_{12}}pl^{\Omega}\{o_i\}(\{\omega_k\})-\delta_{i,k}\right)^2.\end{aligned}$$

Note that this criterion is no longer quadratic. It can be minimized using a standard constrained nonlinear optimization procedure.

Example 5 With the data of Table 1, we obtain $\boldsymbol{\alpha}_1 = (0.45, 0, 0)$ and $\boldsymbol{\alpha}_2 = (0.39, 1, 0)$. Note that, as remarked in [3], this result should not be compared to the previous one: in the first case, we optimized the performance of each sensor individually, whereas in this second case we optimize the performances of the two sensors used in combination. Thus we do not compare the reliability between sensor S_1 and sensor S_2 too. Coefficients $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ correspond to the best coefficients for the best combination of this two sensors. With a larger training set, such a result would indicate that in the combination, sensor S_2 is not used for the detection of helicopters as the second component of vector $\boldsymbol{\alpha}_2$ indicates no abilities for the detection of helicopters. Whereas the information on the detection of helicopters and rockets for sensor S_1 will be taken into account with no discount in the combination. Actually, learning these discount rates is equivalent to learn a strategy of combination.

5 Conclusion

In this paper, a refined sensor tuning has been developped using the Θ-contextual discounting. Once Θ has been fixed, we have shown that discount rates from the contextual discounting operation can be learnt automatically from labeled data by minimizing an error function.

Applied to a single sensor, the result of this new sensor tuning is a richer learning process allowing to learn the reliability of a sensor for different situations to be recognized, and the possibility to compare sensors on each situation. Applied to improve the combination of a group of sensors, it allows to create some refined combination by automatically detecting strenghs and weaknesses of each sensor.

The automatic determination of Θ from data, however, has been left for further study.

Acknowledgements

The authors would like to thank the anonymous referee for his useful comments, in particular those relating to error functions.

References

[1] A. Appriou (1998). Uncertain data aggregation in classification and tracking processes. In *Aggregation and Fusion of imperfect information*, B. Bouchon-Meunier (Ed.), Physica-Verlag, Heidelberg, pages 231-260, 1998.

[2] S. Le Hégarat-Mascle, I. Bloch and D. Vidal-Madjar (1998). Introduction of Neighborhood Information in Evidence Theory and Application to Data Fusion of Radar and Optical Images with Partial Cloud Cover. In *Pattern recognition*, volume 31(11), pages 1811-1823.

[3] Z. Elouedi, K. Mellouli and Ph. Smets (2004). Assessing sensor reliability for multisensor data fusion with the transferable belief model. In *IEEE Transactions on Systems, Man and Cybernetics B*, volume 34, pages 782-787, 2004.

[4] S. Fabre, A. Appriou and X. Briottet (2001). Presentation and Description of Two Classification Methods Using Data Fusion Based on Sensor Management. In *Information Fusion*, volume 2, pages 49-71, 2001.

[5] H. Kim and P. H. Swain (1995). Evidential reasoning approach to multisource-data classification in remote sensing. In *IEEE Transactions on Systems, Man and Cybernetics*, volume 25, pages 1257-1265, 1995.

[6] D. Mercier, B. Quost and T. Denœux (2005). Contextual Discounting of Belief Functions. In *Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2005)*, Lluis Godo (Ed.), Springer Verlag, Barcelona, Spain, pages 552-562, 2005.

[7] N. Milisavljević, I. Bloch, S. van den Broek and M. Acheroy. Improving mine recognition through processing and Dempster-Shafer fusion of ground-penetrating radar data. In *Pattern Recognition*, volume 36, pages 1233-1250, 2003.

[8] G. Shafer (1976). A mathematical theory of evidence. Princeton University Press, Princeton, N.J., 1976.

[9] Ph. Smets (1992). The nature of the unnormalized beliefs encountered in the transferable belief model. In *Proceedings of the 8th Conference on Uncertainty in Artificial Intelligence*, D. Dubois, M. P. Wellman and B. D'Ambrosio (Ed.), Morgan Kaufmann, San Mateo, CA, pages 292-297, 1992.

[10] Ph. Smets (1993). Belief functions: the disjunctive rule of combination and the generalized bayesian theorem. In *International Journal of Approximate Reasoning*, volume 9, pages 1-35, 1993.

[11] Ph. Smets (1994). What is Dempster-Shafer's model? In *Advances in the Dempster-Shafer theory of evidence*, R. R. Yager, M. Fedrizzi and J. Kacprzyk (Ed.), Wiley, New-York, pages 5-34, 1994.

[12] Ph. Smets (2000). Data Fusion in the Transferable Belief Model. In *Proceedings of the 3rd International Conference on Information Fusion, FUSION 2000*, Paris, France, pages PS21-33, 2000.

[13] Ph. Smets (2002). The application of the matrix calculus to belief functions. In *International Journal of Approximate Reasoning*, volume 31, pages 1-30, 2002.

[14] Ph. Smets and R. Kennes (1994). The Transferable Belief Model. In *Artificial Intelligence*, volume 66, pages 191-243, 1994.