Information Fusion 2 (2001) 91–101

INFORMATION FUSION

www.elsevier.nl/locate/inffus

A new distance between two bodies of evidence

Anne-Laure Jousselme [a], Dominic Grenier [a], Éloi Bossé [b,*]

[a] *Laboratoire de Radiocommunication et de Traitement du Signal, Université Laval, Ste-Foy (Qc), Canada G1K 7P4*

[b] *Centre de Recherche pour la Défense, Valcartier, 2459 Pie-XI Blvd. North, P.O. Box 8800, Val-Bélair (Qc), Canada G3J 1X5*

Received 19 July 2000; received in revised form 7 November 2000; accepted 10 November 2000

Abstract

We present a measure of performance (MOP) for identification algorithms based on the evidential theory of Dempster–Shafer. As an MOP, we introduce a principled distance between two basic probability assignments (BPAs) (or two bodies of evidence) based on a quantification of the similarity between sets. We give a geometrical interpretation of BPA and show that the proposed distance satisfies all the requirements for a metric. We also show the link with the quantification of Dempster's weight of conflict proposed by George and Pal. We compare this MOP to that described by Fixsen and Mahler and illustrate the behaviors of the two MOPs with numerical examples.

Keywords: Fusion systems; Evidential theory; Measure of performance

1. Introduction

Interest in data fusion has increased sharply over the last decade, especially for military applications. An important subset of this domain is the identity information fusion process, which fuses the attribute information and identity declarations from various sources and identifies objects from a finite set of possible entities. Among the methods used to address this classification problem, the theory of evidence first presented by Shafer [1] has attracted considerable attention for its effectiveness in combining information and dealing with imprecise evidence.

Many classification algorithms are based on the theory of evidence. In them, a new body of evidence is created by conditioning rules and combined with another provided by an observation. The conditioning rules follow from those first described by Dempster in [2]. The body of evidence then represents the state of beliefs about the situation, given the observations. How can one know if some body of evidence is "far" from the solution or "close" to it? Once this "distance" to the solution is quantified, one is able to observe the progress the algorithm makes as it converges on a solution. In particular, one is able to compare two algorithms for efficiency, according to selected criteria. This is our objective.

So, a very important aspect of the classification problem is scoring how well an identity information fusion process classifies objects. While much research is being carried out to develop and apply new data-fusion algorithms and techniques, little work has been performed to determine how well such methods work, or to compare alternate methods using a common problem. To the authors knowledge, the only measure of performance (MOP) as such in this context, is the one proposed by Fixsen and Mahler [3]. It appears useful to define a MOP that can quantify the effectiveness of a classification algorithm in a given context. Although a variety of criteria might be taken into account, we concentrate on the definition of a "distance" between two basic probability assignments (BPAs). Such a distance can then describe the evolution of an algorithm in its convergence to the solution, which we assume to be known, at least partially.

Previous related work addressing distances between BPAs deserves to be mentioned here. Zouhal and Denoeux [4] introduced a distance based on the mean square error between pignistic probabilities to improve (i.e., optimize some parameters) a classification algorithm based on the *k*-nearest neighbor rule and Dempster–Shafer's theory. An equivalent idea has been used by Tessem [5]: the error due to an approximation of BPAs is quantified by the maximal deviation in the

* Corresponding author. Tel.: +1-418-844-4000 ext. 4478; fax: +1-418-844-4538.

E-mail address: eloi.bosse@drev.dnd.ca (E. Bossé).

1566-2535/01/$ - see front matter
PII: S1566-2535(01)00026-4

pignistic probabilities before and after approximation. Bauer [6], on the same problem, introduced two other measures of error to reflect the quality of a decision based on the pignistic probability distribution after approximation. All these criteria could be used as measures of performance for algorithms based on the theory of evidence. For that, the unavoidable step is the pignistic transformation of the current BPA. Since there is no bijection between BPAs and pignistic probabilities (transformation from the power set to the set), these criteria might not use all the information of the BPAs. The idea to define a measure on the power set of Θ has been used by Petit-Renaud [7], where an error criterion between two belief structures based on the generalized Hausdorf distance has been defined. Our purpose is not to study and compare these criteria, but to define a new distance between two BPAs taking the maximum advantage of the information contained in the BPA, in other words, working on the power set rather than on the set itself.

This paper is organized as follows. In the first part of the paper, we review the concepts and principles of the theory of evidence applied to the classification problem. In the second part, we propose an MOP based on the distance between two BPAs, and show how it could be used to evaluate the classification capability. In the third part, we compare the result with the MOP presented by Fixsen and Mahler [3] and illustrate the behavior of the two MOPs in a simple numerical example. We conclude by discussing the opportunities given by this MOP and especially by its adaptability to different contexts.

2. Review of the theory of evidence

In this section, we review the basic concepts of theory of evidence and introduce related functions and notations. The theory of evidence (also called Dempster–Shafer's theory) has been originally developed by Dempster [2] in his work on upper and lower probabilities, and later on by Shafer [1]. The theory of evidence is often interpreted as an extension of the Bayesian theory of probabilities, however, it has also inspired several models of reasoning under uncertainty, which do not require the probabilistic view. Amongst these models is the more subjective view of the transferable belief model (TBM), introduced by Smets and Kennes [8].

2.1. Introduction

We consider the problem of the identification of an object which can take N possible values. Faced with the usual lack of information on the measures and with the difficulty in quantifying their accuracy, uncertainty must be taken into account. The theory of evidence offers the advantage of focusing on each of the subsets of the set composed of the N objects, rather than on each of the individual objects (as the Bayesian theory does). The functions defined in the theory of evidence allow one to quantify the confidence that a particular event could be the one observed. Then while new information arrives, the identification system integrates it using conditioning rules to provide a representation of the obviousness of the situation.

In Section 2.2, we define the terminology of the Dempster–Shafer theory of evidence and the notation used in this paper.

2.2. Terminology

Let Θ be the set of N elements corresponding to the N identifiable objects, i.e., the finite set of N mutually exclusive and exhaustive hypotheses. This set is called the *frame of discernment*

$$\Theta = \{1, 2, 3, \ldots, N\}. \tag{1}$$

The power set of Θ is the set containing all the possible subsets of Θ, represented by $P(\Theta)$. This set consists of 2^N elements A, each representing the event "the object is in A"

$$P(\Theta) = \{\emptyset, 1, \ldots, N, (1,2), (1,3), \ldots, (N-1, N), (1,2,3), \ldots \Theta\}, \tag{2}$$

where $\emptyset$ denotes the empty set. The N subsets containing only one element are called singletons. A *BPA* (or more judiciously a *Basic Belief Assignment*) is a function from $P(\Theta)$ to $[0,1]$ defined by:

$$m: \quad P(\Theta) \to [0,1], \quad A \mapsto m(A), \tag{3}$$

and which satisfies the following conditions:

$$\sum_{A \in P(\Theta)} m(A) = 1, \tag{4}$$

$$m(\emptyset) = 0. \tag{5}$$

Note that the condition $m(\emptyset) = 0$ is not necessarily required, and if it is not, this corresponds to the "open-world assumption", (and to "closed-world assumption" if $m(\emptyset) = 0$ is specified) [9]. In other words, this assumption of open-world allows the event "the object is none of the N proposed" to have a non-zero mass. We will however consider the closed-world assumption for the rest of the paper.

Because the theory of evidence considers the subsets of the set of N objects, it could then be stated: "$m(A)$ represents our confidence in the fact that all we know is that the object belongs to A". In other words, $m(A)$ is a measure of the belief attributed exactly to A, and to none of the subsets of A. So $m(A) = 1$ expresses the certainty that the object is in (A), but this alone does not allow to distinguish among the elements of (A) and identify the object. To obtain total certainty of an object's identity, the following two conditions must hold: (1) $m(A) = 1$ and (2) $|A| = 1$, where $|\cdot|$ denotes the cardinality function. Only then, it can be stated with certainty that $A = \theta^*$, where θ^* is the object to be identified. On the other had, $m(A) = 1$ and $|A| = N$ correspond to total uncertainty.

The elements of $P(\Theta)$ that have a non-zero mass are called *focal elements*, and the union of all the focal elements is called the *core* of the m-function.

A body of evidence (BOE) is the set of all the focal elements, each with its BPN:

$$(\mathscr{B}, m) = \{[A, m(A)];\ A \in P(\Theta) \text{ and } m(A) > 0\}. \tag{6}$$

Given a BPA m, a belief function Bel is defined as:

$$\begin{aligned} \mathrm{Bel}: \quad & P(\Theta) \to [0,1], \\ & A \mapsto \mathrm{Bel}(A) = \sum_{B \subseteq A} m(B). \end{aligned} \tag{7}$$

$\mathrm{Bel}(A)$ measures the total belief that the object is in A. In particular, we have $\mathrm{Bel}(\emptyset) = 0$ and $\mathrm{Bel}(\Theta) = 1$.

Given a belief function, a plausibility function $P1$ is defined as:

$$\begin{aligned} P1: \quad & P(\Theta) \to [0,1], \\ & A \mapsto P1(A) = \sum_{A \cap B \neq \emptyset} m(B). \end{aligned} \tag{8}$$

It can also be stated that $P1(A) = 1 - \mathrm{Bel}(A^c)$, where A^c is the complement of A. $P1(A)$ measures the total belief that can move into A. In particular, we have $P1(\emptyset) = 0$ and $P1(\Theta) = 1$.

Because the functions m, Bel and $P1$ are one-to-one corresponding, it is equivalent to talking about one of them, or also about the corresponding body of evidence.

2.3. Combination of evidence

Two BPAs m_1 and m_2 (i.e., two bodies of evidence or two belief functions) can be combined to yield a new BPA m, by a combination rule. The more classical one is the Dempster's rule of combination [2], also called *orthogonal sum*, noted by $m = m_1 \oplus m_2$ and defined by:

$$m(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - K} \tag{9}$$

with

$$K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C), \tag{10}$$

where K is a normalization constant, called *conflict* because it measures the degree of conflict between m_1 and m_2. $K = 0$ corresponds to the absence of conflict between m_1 and m_2, whereas $K = 1$ implies complete contradiction between m_1 and m_2. Indeed $K = 0$, if and only if no empty set is created when m_1 and m_2 are combined. On the other hand, $K = 1$ if and only if all the sets resulting from this combination are empty.

The Dempster's rule of combination is not the only rule to combine two BPAs. Smets [10] proposed the *conjunctive* and *disjunctive* rules, which are also unnormalized rules, and by that allow the empty set to have a non-null mass (open-world assumption).

2.4. Identification process

A BPA (or a belief function, or a BOE) represents our belief in a situation at a given time, and the identification algorithm is the process through which the belief function representing our belief at each time is built, according to the combination rules. One of the two BPAs comes from the combination at the previous observation time; the other from external observation. The aim of the fusion algorithm is to determine, given specific observations, which object among N is observed, or at least to determine some attributes of it. Successive combinations cannot provide the answer directly; decision rules must be applied. Their purpose is to select the most likely object, given a state of combination. Since the theory of evidence considers only the agglomerated subsets, decision rules quantify each object individually in order to choose at each time among the N objects. The decision rules can be based on maximum of plausibility, maximum of pignistic probability, maximum of expected utility interval, etc.

3. Measures of performance

3.1. Geometrical interpretation for BPAs

A body of evidence can be seen as a discrete random variable whose values are $P(\Theta)$ with a probability distribution m. This interpretation is very useful in many common situations as it allows Bayesian rules to be used in the framework of random sets to express the relations of the theory of evidence [3,11]. However, for a given BOE $(\mathscr{B}, m)$, $\mathscr{B}$ is a subset of $P(\Theta)$, and each of $A \in \mathscr{B}$ has a fixed value $m(A)$. We can then neglect the random aspect of $(\mathscr{B}, m)$ and use the following "geometric" interpretation.

Let us call $\mathscr{E}_{P(\Theta)}$, the space generated by the elements of $P(\Theta)$. $\mathscr{E}_{P(\Theta)}$ is a *vector space* if any linear combination of the objects of $\mathscr{E}_{P(\Theta)}$ is in $\mathscr{E}_{P(\Theta)}$. That means

$$\vec{V} = \sum_{i=1}^{2^N} \alpha_i A_i \in \mathscr{E}_{P(\Theta)}, \tag{11}$$

where A_i is an element of $P(\Theta)$ and $\alpha_i \in \Re$, $\Re$ being the real set. Then, $A_1, \ldots, A_2^N$ forms a base for $\mathscr{E}_{P(\Theta)}$. Because condition (11) is easily satisfied, we can state that $\mathscr{E}_{P(\Theta)}$ is a vector space on $\Re$ (we can also note that $\mathscr{E}_{P(\Theta)}$ is an isomorphism of $\Re^{2^N}$). Thus, this will lead to the definition of a BPA as a special case of vectors of $\mathscr{E}_{P(\Theta)}$.

Definition 1. Let Θ, be a frame of discernment containing N mutually exclusive and exhaustive hypotheses, and let $\mathscr{E}_{P(\Theta)}$ be the space generated by all the subsets of Θ. A BPA is a vector $\vec{m}$ of $\mathscr{E}_{P(\Theta)}$ with coordinates $m(A_i)$ such that

$$\sum_{i=1}^{2^N} m(A_i) = 1 \quad \text{and} \quad m(A_i) \geqslant 0, \qquad i = 1, \ldots, 2^N. \tag{12}$$

$A_i \in P(\Theta)$.

In this definition, we do not need to impose $m(\varnothing) = 0$. In order to keep maximum of generality, the empty set will be considered in the following discussion. Restricting or not its mass to zero will not change anything.

3.2. Metric spaces

Since the aim of this paper is to define a meaningful metric distance for BPAs, let us introduce some basic concepts for metric spaces and distances.

Definition 2. A metric distance defined on the set $\mathscr{E}$ is a function

$$\begin{aligned} d: \quad & \mathscr{E} \times \mathscr{E} \to \Re, \\ & (A, B) \mapsto d(A, B), \end{aligned}$$

that satisfies the following requirements for any A and B of $\mathscr{E}$:

1. *Nonnegativity*: $d(A, B) \geqslant 0$.
2. *Nondegeneracy*: $d(A, B) = 0 \Longleftrightarrow A = B$.
3. *Symmetry*: $d(A, B) = d(B, A)$.
4. *Triangle inequality*: $d(A, B) \leqslant d(A, C) + d(C, B)$ $\forall C \in \mathscr{E}$.

Definition 3. A metric space is a set $\mathscr{E}$ with a real-valued function $d(A, B)$ defined for all $A, B \in \mathscr{E}$ that satisfies Definition 2.

If one is able to find in $\mathscr{E}_{P(\Theta)}$ a function d that is a metric distance, then $\mathscr{E}_{P(\Theta)}$ will be a metric space. We then propose to look for a distance between two BPAs m_1 and m_2 to be of the form

$$d(m_1, m_2) = (\vec{m}_1 - \vec{m}_2)^{\mathrm{T}} \underline{\underline{D}} (\vec{m}_1 - \vec{m}_2). \tag{13}$$

In order to satisfy the axioms of a distance (Definition 2), $\underline{\underline{D}}$ must be a $2^N \times 2^N$ positively defined matrix. The problem rests in the definition of a d-function i.e., in the construction of the matrix $\underline{\underline{D}}$.

3.3. Similarity between sets

A simple definition of distance between BPAs, such as a Euclidean distance ($\underline{\underline{D}} = \underline{\underline{I}}$), is not appropriate. Indeed, $d_I(m_1, m_2) = \|\vec{m}_1 - \vec{m}_2\|$, with $\|\vec{m}\|^2 = \vec{m}^{\mathrm{T}}\vec{m}$, gives an equivalent weight to all the subsets of Θ, without taking into account similarities among them. For example, the subset $\{1, 2, 3\}$ is "closer" to $\{1, 2, 3, 4\}$ than is $\{5, 6, 7\}$. [1] In particular, if we consider three bodies of evidence:

$$\begin{aligned} (\mathscr{B}, m) &= \{[(1, 2, 3), 0.8],\ [\Theta, 0.2]\}, \\ (\mathscr{C}_1, n_1) &= \{[(1, 2, 3, 4), 0.8],\ [\Theta, 0.2]\}, \\ (\mathscr{C}_2, n_2) &= \{[(5, 6, 7), 0.8],\ [\Theta, 0.2]\}, \end{aligned}$$

then we have $d_I(m, n_1) = d_I(m, n_2)$, even if we expect $d_I(m, n_1) < d_I(m, n_2)$. Regarding this aspect of BOEs, a matrix must be defined that describes the "similarity" (or likeness) between the subsets of Θ.

The requirements for $\underline{\underline{D}}$ are then:

- $\underline{\underline{D}}$ must define a metric distance.
- $\underline{\underline{D}}$ must take into account the similarity among the subsets of Θ.
- $\underline{\underline{D}}$ must satisfy $d(m, n_1) < d(m, n_2)$, if n_1 is "closer" to m than n_2 is.

Let $D(A, B)$ be a coefficient of $\underline{\underline{D}}$. The previous requirements imply the following conditions for all A and B of $P(\Theta)$:

(a) $D(A, B) \leqslant 1$ and $D(A, B) = 1$ if and only if $A = B$.
(b) The "closer" to each other A and B are, the nearer $D(A, B)$ must be to unity.
(c) The "farther" from each other A and B are, the nearer $D(A, B)$ must be to zero.

In other words, $D(A, B)$ must express a sort of "inverse-conflict" between A and B.

If we suppose that the N objects of Θ are indiscernible and unorderable, the only measure distinguishing any two subsets of is their cardinality. We therefore contend that a distance between two subsets A and B of Θ can be a function only of $|A|$, $|B|$, $|A \cap B|$, and $|A \cup B|$, where $|A|$ denotes the cardinality of A.

$|A \cap B|$ is a measure of the conflict between A and B. $|A \cap B| = 0$ means that A and B have no object in common, so they are in great conflict. This measure satisfies condition (c). To satisfy conditions (a) and (b), a denominator always greater than $|A \cap B|$ must be introduced:

$$\begin{aligned} A = B &\Longleftrightarrow A \subseteq B \quad \text{and} \quad B \subseteq A, \\ &\Longleftrightarrow |A \cap B| = |A| \quad \text{and} \quad |A \cup B| = |A|, \\ &\Longleftrightarrow |A \cap B| = |A \cup B|. \end{aligned}$$

[1] The numbers represent object indices.

We propose then

$$D(A,B) = \frac{|A \cap B|}{|A \cup B|}, \tag{14}$$

which satisfies conditions (a)–(c). Furthermore, matrix $\underline{\underline{D}}$ satisfies all the requirements for a metric (i.e., $\underline{\underline{D}}$ is definite positive).

This measure of similarity between sets is very natural and has been yet used for example in [12] to simplify belief functions or in [13] to detect similar fuzzy sets and also in [14] to quantify the conflict in Dempster–Shafer's theory of evidence. This last aspect will be detailed in Section 4.1.2.

This leads then to the following definition.

Definition 4. Let m_1 and m_2 be two BPAs on the same frame of discernment Θ, containing N mutually exclusive and exhaustive hypotheses. The distance between m_1 and m_2 is:

$$d_{\mathrm{BPA}}(m_1, m_2) = \sqrt{\frac{1}{2}(\vec{m}_1 - \vec{m}_2)^{\mathrm{T}} \underline{\underline{D}} (\vec{m}_1 - \vec{m}_2)}, \tag{15}$$

where $\vec{m}_1$ and $\vec{m}_2$ are the BPAs according to Definition 1 and $\underline{\underline{D}}$ is an $2^N \times 2^N$ matrix whose elements are

$$D(A,B) = \frac{|A \cap B|}{|A \cup B|}.$$

$$A, B \in P(\Theta).$$

Note that a factor 1/2 is needed in Eq. (15) to normalize d_{BPA} and to guarantee that $0 \leqslant d_{\mathrm{BPA}}(m_1, m_2) \leqslant 1$.

From Definition 4, another way to write d_{BPA} is:

$$d_{\mathrm{BPA}}(m_1, m_2) = \sqrt{\frac{1}{2}(\|\vec{m}_1\|^2 + \|\vec{m}_2\|^2 - 2\langle \vec{m}_1, \vec{m}_2 \rangle)}, \tag{16}$$

where $\langle \vec{m}_1, \vec{m}_2 \rangle$ is the scalar product defined by

$$\langle \vec{m}_1, \vec{m}_2 \rangle = \sum_{i=1}^{2^N} \sum_{j=1}^{2^N} m_1(A_i) m_2(A_j) \frac{|A_i \cap A_j|}{|A_i \cup A_j|}, \tag{17}$$

with A_i, $A_j \in P(\Theta)$ for $i, j = 1, \ldots, 2^N$. $\|\vec{m}\|^2$ is then the square norm of $\vec{m}$:

$$\|\vec{m}\|^2 = \langle \vec{m}, \vec{m} \rangle. \tag{18}$$

Having defined the distance between any two BPAs, we will focus now on the convergence.

3.4. Convergence to the solution

Measuring the performance of an algorithm, and especially determining its accuracy, presupposes knowledge (complete or partial) of the solution. In the case of an object identification algorithm, the solution is expected to be a singleton (solution completely known) that represents the object identified. However, the solution could also be a set of several objects (solution partially known) that cannot be distinguished by the system. Consider, for example, two targets that differ only in name. If the fusion system lacks an appropriate sensor for this particular parameter, is it really relevant to know how close the result of the algorithm is to the real object? Is it not more appropriate to know how close the result is to these two objects considered as one? We could simply also want to know how close the result of the algorithm is to a group of elements with the same attribute. For example, a question of interest could be "what is our confidence in the fact that the target is a frigate?" The definition of distance proposed here allows a choice of solutions among all the subsets of Θ, not only among the elements of Θ. The MOP can thus readily be adapted to the limitations of the fusion system.

Consequently, we will call *solution* any body of evidence (or BPA) representing *the best expectation* of the user. Ideally, the solution will take the form of a body of evidence consisting of only one focal element, thus the BNP is 1:

$$(\mathscr{B}^*, m^*) = \{A^*, 1\}. \tag{19}$$

In other words, this means that we are absolutely certain that the object observed is in A^*, A^* being any subset of Θ. m^* is then *the simple support function focused on* A^*, with a degree of support of 1.

However, the solution could also be a BPA focused on more than one subset of Θ. The mass given to Θ is sometimes constrained in practice to be greater than 0, to avoid problems due to a convergence of the algorithm to a wrong identification. This occurs for example when the same report (which moreover is a wrong information) is combined many times which makes the mass of ignorance decreasing to 0. In this case, it is difficult to account for these successive countermeasures except if the mass of Θ is imposed to a certain threshold. This constraint must be included in the analysis of convergence to the solution. For instance, the solution could be set to

$$(\mathscr{B}^*, m^*) = \{[A^*, 0.9], [\Theta, 0.1]\} \tag{20}$$

if 0.1 were the minimum threshold imposed on $m(\Theta)$ in the algorithm under investigation (for an example see [15]).

Definition 5. Let $m_1, m_2, \ldots$ be a sequence of BPAs coming from successive combinations, and let m^* be the solution of the algorithm. The sequence is said to converge to m^* if for any positive real number ε there exists an integer $n(\varepsilon)$ such that

$$d_{\mathrm{BPA}}(m_i, m^*) < \varepsilon \quad \forall i \geqslant n(\varepsilon).$$

m^* is then called the limit of the sequence.

Given the distance between a BPA and a solution, one is able to compute $d_{\mathrm{BPA}}(m_i, m^*)$ at each time, and then analyze the convergence of the algorithm. This will be the aim of Section 4.1.1. But before that, we will recall the definition of the Bayesian percent attribute miss, introduced by Fixsen and Mahler.

3.5. The classification "miss-distance" metric (BPAM)

Fixsen and Mahler [3] have defined a principled classification miss-distance metric called BPAM, for Bayesian percent attribute miss, to measure the performance of an algorithm based on the modified-Dempster–Shafer theory. Let $(\mathscr{B}_1, m_1)$ and $(\mathscr{B}_2, m_2)$ be two bodies of evidence. Then they define

$$d_\alpha^2(\mathscr{B}_1,\mathscr{B}_2) = \alpha_q(\mathscr{B}_1,\mathscr{B}_1) - 2\alpha_q(\mathscr{B}_1,\mathscr{B}_2) + \alpha_q(\mathscr{B}_2,\mathscr{B}_2), \tag{21}$$

where $\alpha_q(\mathscr{B}_1,\mathscr{B}_2)$ is the scalar product defined by:

$$\alpha_q(\mathscr{B}_1,\mathscr{B}_2) = \sum_{A\in\mathscr{B}_1}\sum_{B\in\mathscr{B}_2} m_1(A)m_2(B)\frac{q(A\cap B)}{q(A)q(B)}. \tag{22}$$

q is a Bayesian a priori distribution on Θ. Assuming q is uniform ($q(A) = |A|/|\Theta|$, for all A, subset of Θ), i.e., giving equal weight to each object in Θ, (22) reduces to

$$\alpha_q(\mathscr{B}_1,\mathscr{B}_2) = \sum_{A\in\mathscr{B}_1}\sum_{B\in\mathscr{B}_2} m_1(A)m_2(B)|\Theta|\frac{(A\cap B)}{|A|\cdot|B|}. \tag{23}$$

In order to have values between 0 and 1 (and then independent on $|\Theta|$), they used in [16] $\alpha(\mathscr{B}_1,\mathscr{B}_2) = \alpha_q(\mathscr{B}_1,\mathscr{B}_2)/|\Theta|$ instead of $\alpha_q(\mathscr{B}_1,\mathscr{B}_2)$. Given the geometrical interpretation of Section 3.1, matrix $\underline{\underline{D}}_\alpha$ defining the BPAM is then given by its elements

$$D_\alpha(A,B) = \frac{|A\cap B|}{|A||B|} \tag{24}$$

$\forall A,B \in P(\Theta)$.The result is a pseudo-metric, since the condition of nondegeneracy is not respected. This means that $(\mathscr{B}_1,m_1) \neq (\mathscr{B}_2,m_2)$ exists such that $d_\alpha(\mathscr{B}_1,\mathscr{B}_2) = 0$. We choose this measure as a comparison for our distance and we will, later in Section 4.2, detail its behavior.

4. Special cases and illustrative examples

As mentioned in Section 3.4, the solution is represented by a simple support BPA m^* focused on $A^* \in P(\Theta)$, ideally with a degree of support equal to 1. Let us call m_t the BPA arising from the combination at time t. In this section, particular values of $d_{\text{BPA}}(m^*, m_t)$ are given in cases of correct and incorrect convergences (Sections 4.1.1 and 4.1.2, respectively), and these results are illustrated by fictitious convergences on a simple case.

We consider the simple case of six objects, $\Theta = \{1,2,3,4,5,6\}$. $P(\Theta)$ is then composed of the 64 subsets of Θ. Since the number of elements is small, it is easy to explore the behavior of the MOP. To test the MOP d_{BPA} independently of the algorithm used (since it is based on the Dempster–Shafer theory), i.e., without considering the fact that BPAs originate from conditioning rules, the convergence is "forced" to one or the other of the elements of $P(\Theta)$ and the reaction of the MOP in this context is observed. The spurious convergence of the algorithm to the BPA m^* is forced by transferring the mass linearly from all the subsets of $P(\Theta)$ to A^*, until 1 is attained. We start with an equal distribution of the mass over the 63 non-empty subsets (the mass of the empty set will be set to 0 for the rest of the discussion); that is, m_0 is a uniform distribution on $P(\Theta) \setminus \{\varnothing\}$. At each step of the transfer, an increase of Δ for $m(A^*)$ implies a decrease of $\Delta/62$ for each of the other subsets, so that the sum of the BPNs remains 1. For the simulations Δ has been fixed to 0.02.

Figs. 1–5 show the MOP d_{BPA} in several situations:

- The algorithm converges to the right solution:
  - The solution is a singleton (Fig. 1).
  - The solution consists of several objects (Fig. 2).
- The algorithm converges to an incorrect BPA:
  - The solution is a singleton and the algorithm converges to a different singleton (Fig. 3).
  - The solution is a singleton and the algorithm converges to a subset containing the object (Fig. 4).
  - The solution is a singleton and the algorithm converges to total ignorance (Fig. 5).

These figures illustrate the behavior of the equations developed in the following section.

4.1. Analysis of particular cases

4.1.1. Distance to the solution

In the general case, the distance from any BPA m_t to m^* is given, following (16), by:

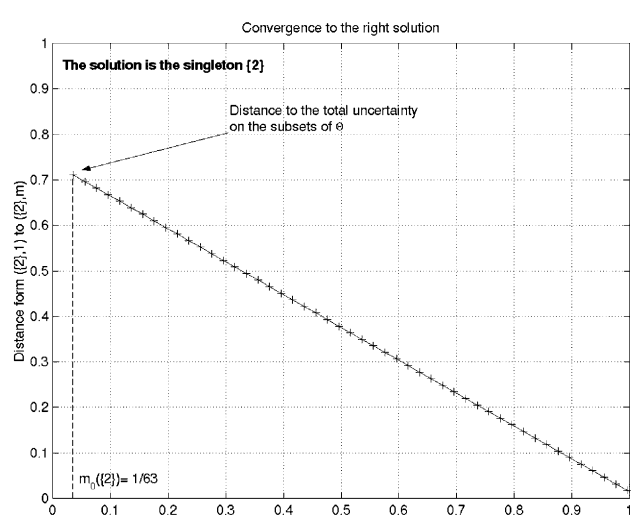

Fig. 1. Convergence to the right solution, which is a singleton.

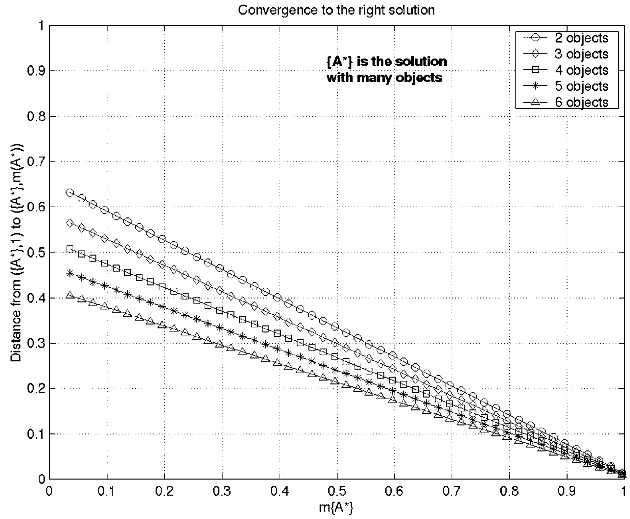

Fig. 2. Convergence to the right solution, when it contains several objects.

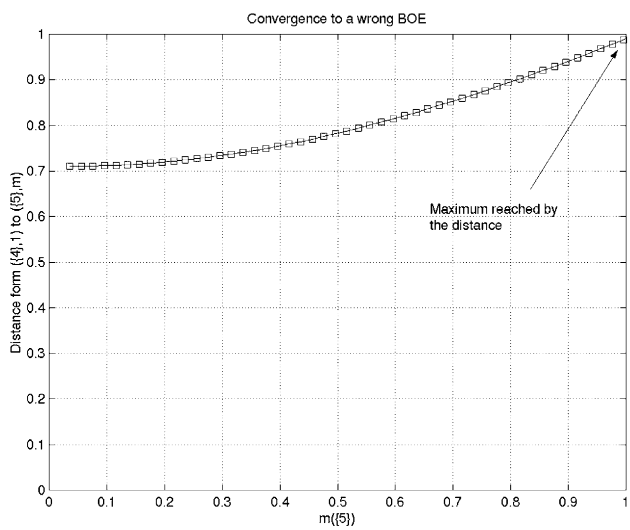

Fig. 3. Convergence to the wrong BPA. Solution is a singleton and the point of convergence is another singleton.

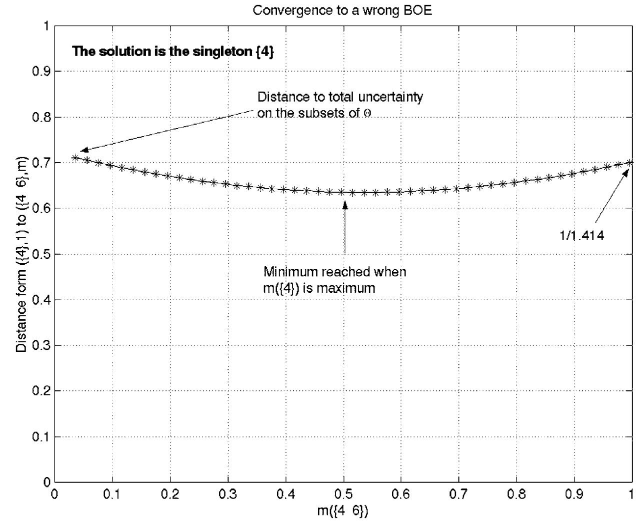

Fig. 4. Convergence to the wrong BPA. Solution is a singleton and the point of convergence is a subset containing the singleton solution.

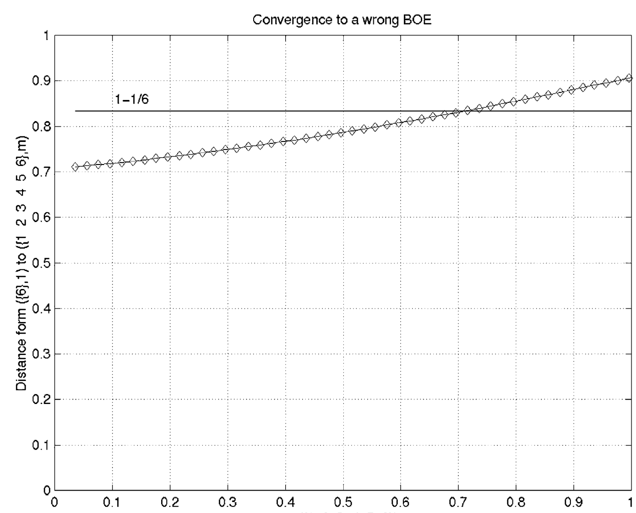

Fig. 5. Convergence to the wrong BPA. Solution is a singleton and the point of convergence is the ignorance $\{\Theta, 1\}$.

$$d^2_{\mathrm{BPA}}(m^*, m_t) = \frac{1}{2}\left[1 + \|\vec{m}_t\|^2 - 2\sum_{B\subseteq\Theta} m(B)\frac{|A^*\cap B|}{|A^*\cup B|}\right]. \tag{25}$$

When the solution is known to be a singleton ($|A^*| = 1$ and $A^* = \theta^*$), (25) becomes:

$$d^2_{\mathrm{BPA}}(m^*, m_t) = \frac{1}{2}\left[1 + \|\vec{m}_t\|^2 - 2\,\mathrm{Bet}\,P(\theta^*)\right], \tag{26}$$

where

$$\mathrm{Bet}\,P(\theta^*) = \langle\vec{\theta}^*, \vec{m}_t\rangle = \sum_{\theta^*\in B\subseteq\Theta}\frac{m(B)}{|B|} \tag{27}$$

is the pignistic probability of θ^* defined by Smets [17].

Thus when m_t tends to $\{A^*, 1\}$, $d_{\mathrm{BPA}}(m^*, m_t)$ tends to 0. Fig. 1 shows the linear convergence of the sequence m_t to the singleton $\{2\}$. Fig. 2 shows the equivalent convergence when the solution is a set A^* that contains several objects.

Expression (26) must be related to the square error introduced by Zouhal and Denoeux [4]. They proposed to compute the square error between the pignistic probability distribution associated to a BPA m_t, and a vector of binary indicator representing the solution. This expresses the discrepancy between an output BPA m_t and a target value θ^*. Within our formalism, the square error E_x they defined is expressed as

$$E_x(m^*, m_t) = 1 + \sum_{i=1}^{N}\mathrm{Bet}\,P(\theta_i)^2 - 2\,\mathrm{Bet}\,P(\theta^*), \tag{28}$$

with m^* being the simple support function focused on θ^* with a degree of support equal to 1.

If we perform the pignistic transformation on m_t in Eq. (26), $\vec{m}_t$ will give a vector whose components are

$$m_{\mathrm{PT}}(A) = \begin{cases} \mathrm{Bet}\,P(\theta_i) & \text{if } |A| = 1 (A = \theta_i), \\ 0 & \text{otherwise,} \end{cases} \tag{29}$$

and $\underline{\underline{D}}$ gives a matrix whose elements are

$$D_{\mathrm{PT}}(A,B) = \begin{cases} 1 & \text{if } |A| = 1 \text{ and } |B| = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{30}$$

Pruning for simplicity all the components of $\vec{m}_{\mathrm{PT}}$ such that $|A| \neq 1$ leads to an $N \times 1$-vector $\vec{\theta}_t$. So

$$\|\vec{m}_{\mathrm{PT}}\|^2 = \vec{\theta}^{\mathrm{T}}\vec{\theta} = \sum_{i=1}^{N} \mathrm{Bet}\,P(\theta_i)^2. \tag{31}$$

Finally, with this pignistic transformation, (26) becomes

$$d^2_{\mathrm{BPA}}(m^*, m_t) = \frac{1}{2}\left[1 + \sum_{i=1}^{N} \mathrm{Bet}P(\theta_i)^2 - 2\,\mathrm{Bet}\,P(\theta^*)\right]. \tag{32}$$

We thus can state that

$$E_x(m^*, m_t) = 2d^2_{\mathrm{BPA}}(m^*, m_{\mathrm{PT}}), \tag{33}$$

where m_{PT} is the pignistic transformation of m_t.

4.1.2. Convergence to an incorrect BPA

Let us call m' the BPA corresponding to the limit of the algorithm, and suppose m' is the simple support function focused on A' with 1 as degree of support, where A' can be any element of $P(\Theta), A' \neq A^*$. The algorithm then converges to a BPA that *is not* the solution, and the sequence m_t tends to m'. Then (25) becomes

$$d^2_{\mathrm{BPA}}(m^*, m') = \Pi(A^*, A'), \tag{34}$$

where

$$\Pi(A^*, A') = 1 - \frac{|A^* \cap A'|}{|A^* \cup A'|} \tag{35}$$

is the conflict between A^* and A' defined by George and Pal [14]. The authors have also defined the total conflict in a BOE $(\mathscr{B}, m)$ by

$$\mathrm{TC}(\mathscr{B}, m) = \sum_{A,B \in \mathscr{B}} m(A)m(B)\Pi(A,B), \tag{36}$$

$$= 1 - \|\vec{m}\|^2, \tag{37}$$

where $\|\vec{m}\|^2$ is the square norm of $\|\vec{m}\|^2$ defined in Definition 4.

If $A^* \subseteq A'$ (i.e., the algorithm converges to a BPA whose only focal element contains the solution), then

$$d^2_{\mathrm{BPA}}(m^*, m') = 1 - \frac{|A^*|}{|A'|}. \tag{38}$$

In particular, if $|A'| = 2$, which means that the algorithm is unable to distinguish between two elements but has come to the conclusion that the object is among them, the $d^2_{\mathrm{BPA}}(m^*, m') = 1/2$. Fig. 4 illustrate this behavior.

Finally, if the solution is a singleton θ^* and if the algorithm converges to $\{A', 1\}$, two cases are possible:

$$d^2_{\mathrm{BPA}}(m^*, m') = \begin{cases} 1 - \frac{1}{|A'|} & \text{if } \theta^* \in A', \\ 1 & \text{if } \theta^* \notin A'. \end{cases} \tag{39}$$

Note that 1 is the maximum value that can be reached by d_{BPA}, and this occurs in computing the distance between $\{A^*, 1\}$ and $\{A', 1\}$ only when these two sets have no object in common.

4.1.3. Vacuous belief function

The *vacuous belief function* is defined by a BPA (noted m_Θ) whose only focal element is Θ. It corresponds to *total ignorance* of the elements of Θ. Its distance to any other BPA is then:

$$d^2_{\mathrm{BPA}}(m_t, m_\Theta) = \frac{1}{2}\left[1 + \|\vec{m}_t\|^2 - \frac{2}{N}\sum_{B \subseteq \Theta} m(B)|B|\right], \tag{40}$$

with $N = |\Theta|$. If the solution is $\{A^*, 1\}$ and the algorithm converges to $\{\Theta, 1\}$, one obtains:

$$d^2_{\mathrm{BPA}}(m^*, m_\Theta) = 1 - \frac{|A^*|}{N}. \tag{41}$$

The closer $|A^*|$ becomes to N (i.e., the greater the number of elements in the solution), the closer to 0 the distance is, which agrees with intuition. Moreover, if A^* is a singleton, then:

$$d^2_{\mathrm{BPA}}(m^*, m_\Theta) = 1 - \frac{1}{N} \tag{42}$$

or as the number of elements in Θ increases, so does the distance between total ignorance and the solution. Note that Eqs. (41) and (42) are special cases of Eq. (38).

4.1.4. Uniform distribution on $P(\Theta)$

We consider now a uniform distribution m_{U} on $P(\Theta)$. This implies that $m_{\mathrm{U}}(A) = 1/(2^N - 1)$ for all $A \in P(\Theta)$, $m_{\mathrm{U}}(\emptyset) = 0$. This corresponds to a total uncertainty on the subsets of Θ, so that any set could be the solution, whereas the vacuous belief function signifies a total uncertainty on the elements of Θ, so that any object in Θ could be the solution. Knowing with certainty that the object is in Θ is not the same as believing that the subsets of $P(\Theta)$ are equally likely to contain the object. From a classical probabilistic point of view, if the question of interest is "which is the object?" these two statements are equivalent, since both lead to Prob(θ^* is the object) $= 1/N$. Nevertheless, if the concern is to know which set contains the object with total certainty, there is a difference between m_{U} and m_Θ. Computing the distance d_{BPA} from m_{U} to any BPA m_t gives

$$d^2_{\mathrm{BPA}}(m_t, m_{\mathrm{U}}) = \frac{1}{2}\left[\frac{K_{\mathrm{U}}}{2^N - 1} + \|\vec{m}_t\|^2 - \frac{2}{2^N - 1}\sum_{A\subseteq\Theta}\sum_{B\subseteq\Theta} m(B)\frac{|A\cap B|}{|A\cup B|}\right], \qquad (43)$$

where

$$K_{\mathrm{U}} = \sum_{A\subseteq\Theta}\sum_{B\subseteq\Theta}\frac{|A\cap B|}{|A\cup B|} \qquad (44)$$

is a constant for a given Θ. In particular, the distance to m_Θ is then:

$$d^2_{\mathrm{BPA}}(m_\Theta, m_{\mathrm{U}}) = \frac{1}{2}\left[\frac{K_{\mathrm{U}}}{(2^N - 1)^2} + 1 - \frac{2K_A}{(2^N - 1)}\right], \qquad (45)$$

where

$$K_A = \sum_{A\subseteq\Theta}|A| \qquad (46)$$

is also a constant for a given Θ. It follows from (45) that the distance between m_Θ and m_{U} depends only on N and differs from 0. It will be seen in Section 4.2 that the situation is not the same for the BPAM for which the two BPAs are equal.

4.1.5. Bayesian functions

We conclude with the case of two Bayesian functions m_1 and m_2: that is, two BPAs having only singletons as focal elements. d_{BPA} thus is reduced to the Euclidean distance:

$$d^2_{\mathrm{BPA}}(m_1, m_2) = \|\vec{m}_1 - \vec{m}_2\|^2 \qquad (47)$$

with

$$\|\vec{m}\|^2 = \sum_{i=1}^{N} m(\theta_i)^2. \qquad (48)$$

This last expression is then equivalent to (31), which represents a square norm of a pignistic probability distribution.

4.2. Comparison with BPAM

The main difference between d_{BPA} and d_α, the distance described in Section 3.5, lies in the nondegeneracy, which is not respected by d_α. An example is obtained by computing the distance between m_Θ and m_{U}, as was done in Section 4.1.4 for d_{BPA}. We have then

$$d^2_\alpha(m_{\mathrm{U}}, m_\Theta) = 0, \qquad (49)$$

which, in addition to illustrating the degeneracy of d_α, means that this distance confuses "knowing with certainty that the object is in Θ" with "having an equal belief that each subset of $P(\Theta)$ contains the object".

For another comparison between the two MOPs that highlights the pseudo-distance feature of the BPAM, consider the following examples:

Example 1. Let $(\mathscr{A}_t, m_t)$ be the body of evidence defined by

$$(\mathscr{A}_t, m_t) = \{(\Theta, 0.1), ([2,3,4], 0.05), ([7], 0.05), (A_t, 0.8)\}, \qquad (50)$$

where A_t is a variable set taking the values given by Table 1. Let $(\mathscr{A}^*, m^*)$ be a second BOE representing the solution and defined by

$$(\mathscr{A}^*, m^*) = \{[1,2,3,4,5], 1\}. \qquad (51)$$

As we can see, the set A_t tends to the solution $A^* = [1,2,3,4,5]$ from steps 1 to 4, attains the solution at step 5, and then departs from it from steps 6 to 20. The comparative behavior of the two MOPs is shown in Fig. 6. Both distances reach their minimum at step 5. However, whereas $d_{\mathrm{BPA}}(m^*, m_t)$ still increases as A_t departs from A^*, $d_\alpha(m^*, m_t)$ increases towards a threshold, meaning that the set A_{20} is closer to A^* than A_2 is. But $A_{20} = [1,2,\ldots,20]$ and $A_2 = [1,2]$, and we thus expect A_2 to be closer to A^* than A_{20}.

Example 2. Let $(\mathscr{B}_t, m_t)$ be now the body of evidence defined by

Table 1
Variable set A_t

Step	A_t
1	$[1]$
2	$[1,2]$
3	$[1,2,3]$
4	$[1,2,3,4]$
5	$[\mathbf{1},\mathbf{2},\mathbf{3},\mathbf{4},\mathbf{5}]$
6	$[1,2,3,4,5,6]$
7	$[1,2,3,4,5,6,7]$
⋮	⋮
20	$[1,2,3,4,\ldots,20]$

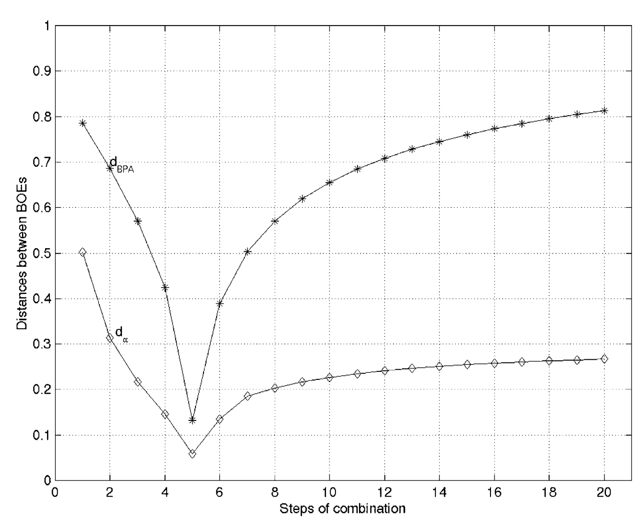

Fig. 6. Comparison of $d_{\mathrm{BPA}}(m_1, m_2)$ with BPAM – Example 1.

Table 2
Variable set B_t

Step	B_t
1–10	A_t
11	[2, 3, 4, 5, 6, 7, 8, 9, 10]
12	[3, 4, 5, 6, 7, 8, 9, 10]
13	[4, 5, 6, 7, 8, 9, 10]
14	[5, 6, 7, 8, 9, 10]
⋮	⋮
19	[10]

$$(\mathscr{B}_t, m_t) = \{(\Theta, 0.1), ([2,3,4], 0.05), ([7], 0.05), (B_t, 0.8)\}, \tag{52}$$

where B_t is a variable set taking the values given by Table 2. Let $(\mathscr{B}^*, m^*)$ he a second BOE representing the solution and defined simply by

$$(\mathscr{B}^*, m^*) = \{[10], 1\}. \tag{53}$$

According to Table 2, the set B_t adopts the same behavior as the set A_t did in example 1 from steps 1 to 10. Then, from steps 11 to 19, B_t is pruned from its first element until attaining the singleton solution $B^* = \{10\}$ at last step. We observe the comparative behaviors of the two MOPs in Fig. 7.

The solution is now the singleton $\{10\}$ and both distances have the same values at step 1. But from steps 1 to 9, $d_{\mathrm{BPA}}(m^*, m_t)$ increases whereas $d_\alpha(m^*, m_t)$ decreases. The first behavior is expected to be the correct one because during this 10 first steps of combination, the set B_t really departs from the solution, because the singleton $\{10\}$ is still not included in the variable set while increasing in the number of elements. The distance $d_{\mathrm{BPA}}(m^*, m_t)$ starts to decrease while the singleton $\{10\}$ appears in B_t, at step 10. Then, both MOPs have the same behavior from steps 11 to 19: they decrease sharply

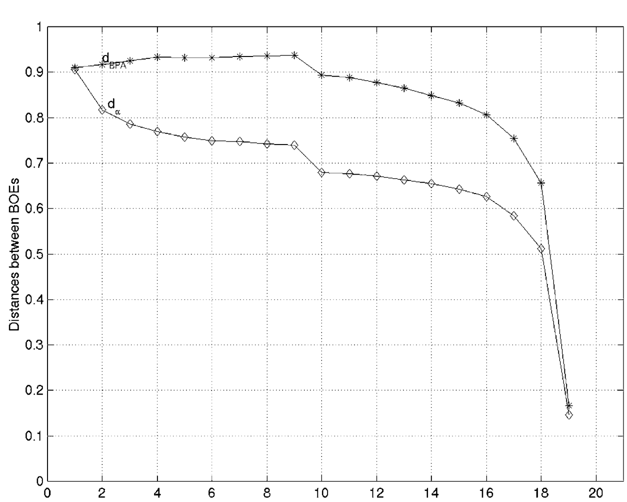

Fig. 7. Comparison of $d_{\mathrm{BPA}}(m_1, m_2)$ with BPAM – Example 2.

to reach their respective minimum values when $B_t = \{10\}$.

These two examples have shown some comparative behaviors of our distance and the BPAM. They essentially point out the pseudo-distance feature of the BPAM.

5. Conclusion

This paper has presented a view of the power set of the frame of discernment $P(\Theta)$ in the theory of evidence as a 2^N-linear space where a distance and vectors are defined with the BPA as a particular case of vectors. A distance (we called d_{BPA}) between two BPAs is proposed as an MOP for identification algorithms and the properties of this distance are analyzed. The function d_{BPA} respects all the properties expected of a distance and is an appropriate measure of the difference-or the lack of similarity-between any two BPAs. Because d_{BPA} is not a pseudo-distance $P(\Theta)$, this allows any two BPAs to be accounted for the computation of their distance. Thus, the BPA solution is not necessarily focused on a singleton with a degree of support equal to 1, but may take more general value m^*. This flexibility allows to take into account the limitations of sensors (if, in particular, the available sensors cannot distinguish among two or more objects in the database).

The distance d_{BPA} is compared to the pseudo-distance (BPAM) defined by Fixsen and Mahler. Because BPAM is a pseudo-distance, it can only compute a real distance to a body of evidence that is in the form of $\{\theta, 1\}$, θ being a singleton of $P(\Theta)$. According to the BPAM measure the two BPAs m_Θ (corresponding to the vacuous belief function) and m_{U} (uniform distribution on the elements of $P(\Theta) \setminus \emptyset$) are equal, which confuses total certainty about the ignorance with total uncertainty about all the subsets of Θ.

When the solution is a singleton, a link between d_{BPA} and the pignistic probability of Smets has been established. In particular, by performing a pignistic transformation on the output BPA, we found the same form of that proposed by Zouhal and Denoeux. We also concluded that the distance between two different BOEs in the form of $\{A, 1\}$ is the conflict defined by George and Pal.

To illustrate the behavior of d_{BPA} in different situations, we simulated the convergence of a particular algorithm to different varieties of BPAs (solutions that are or are not singletons, subsets containing the solution, etc.). We concluded in particular, that $d_{\mathrm{BPA}}(m_1, m_2)$ equals 1 if the BPAs have focal elements with no objects in common, and that $d_{\mathrm{BPA}}(m_1, m_2)$ equals 0 if the two focal elements are the same. Between these two bounds, d_{BPA} takes values representing how far (or close) both BPAs are from each other.

We believe that the new distance proposed provides an easy and effective way to test and compare algorithms based on the theory of evidence.

References

[1] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.
[2] A. Dempster, Upper and lower probabilities induced by multivalued mapping, Ann. Math. Stat. 38 (1967) 325–339.
[3] D. Fixsen, R.P.S. Mahler, The modified Dempster–Shafer approach to classification, IEEE Trans. Syst. Man Cybern. A 27 (1) (1997) 96–104.
[4] L.M. Zouhal, T. Denoeux, An evidence-theoric *k*-NN rule with parameter optimization, IEEE Trans. Syst. Man Cybern. C 28 (2) (1998) 263–271.
[5] B. Tessem, Approximations for efficient computation in the theory of evidence, Artif. Intell. 61 (2) (1993) 315–329.
[6] M. Bauer, Approximation algorithms and decision making in the Dempster–Shafer theory of evidence – an empirical study, Int. J. Approx. Reason. 17 (2–3) (1997) 217.
[7] S. Petit-Renaud, Application de la théorie des croyances et des systèmes flous à l'estimation fonctionnelle en présence d'informations incertaines ou imprécises, Ph.D. thesis, Université de Technologie de Compiègne, December 1999.
[8] P. Smets, R. Kennes, The transferable belief model, Artif. Intell. 66 (1994) 191–234.
[9] P. Smets, The combination of evidence in the transferable belief model, IEEE Trans. Pattern Anal. Mach. Intell. 12 (5) (1990) 447–458.
[10] P. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, Int. J. Approx. Reason. (1999) 1–35.
[11] H.T. Ngyuen, On random sets and belief functions, J. Math. Anal. Appl. 65 (1978) 531–542.
[12] S. Petit-Renaud, T. Denoeux, Handling different forms of uncertainty in regression analysis: a fuzzy belief structure approach, in: A. Hunter, S. Pearsons (Eds.), Symbolic and Qualitative Approaches Reasoning and Uncertainty (ECSQARU'99), London, June 1999, pp. 340–352.
[13] M. Setnes, R. Babuška, U. Kaymak, H.R. van Nauta Lemke, Similarity measures in fuzzy rule base simplification, IEEE Trans. Syst. Man Cybern. B 28 (3) (1998) 376–386.
[14] T. George, N.R. Pal, Quantification of conflict in Dempster–Shafer framework: a new approach, Int. J. Gen. Syst. 24 (4) (1996) 407–423.
[15] É. Bossé, M.-A. Simard, Managing evidential reasoning for identity information fusion, Soc. Photo-Optical Instrum. Eng. 37 (2) (1998) 391–400.
[16] D. Fixsen, R. Mahler, A Dempster–Shafer approach to Bayesian classification, in: Proceedings of the Fifth National Symposium on Sensor Fusion, vol. I, 21–23 April, Orlando, FL, 1992, pp. 213–231.
[17] P. Smets, Constructing the pignistic probability function in a context of uncertainty, Uncertainty in Artificial Intelligence, vol. 5, Elsevier, Amsterdam, 1990, pp. 29–39.