New Metrics between Bodies of Evidences

Pascal Djiknavorian, Dominic Grenier
Laval University/Electrical and Computer Engineering, Quebec, Canada
Email: {djikna, Dominic.Grenier}@gel.ulaval.ca

Pierre Valin
Defence R&D Canada Valcartier/Decision Support Systems section, Quebec, Canada
Email: pierre.valin@drdc-rddc.gc.ca

***Abstract*—We address the problem of the computational difficulties occurring by the heavy processing load required by the use of the Dempster-Shafer Theory (DST) in Information Retrieval. Specifically, we focus our efforts on the measure of performance known as the Jousselme distance between two basic probability assignments (or bodies of evidences). We discuss first the extension of the Jousselme distance from the DST to the Dezert-Smarandache Theory, a generalization of the DST. It is followed by an introduction to two new metrics we have developed: a Hamming inspired metric for evidences, and a metric based on the degree of shared uncertainty. The performances of theses metrics are compared one to each other.**

***Index Terms*—Dempster-Shafer, Measure of performance, Evidential Theory, Dezert-Smarandache, Distance**

I. Introduction

Comparing two, or more, bodies of evidences (BOE) in the case of large frame of discernment, in the Dempster-Shafer theory of evidence [1, 2], may not always give intuitive choices from which we can simply choose a proposition the with largest basic probability assignment (BPA) (or mass), or belief. A metric becomes very useful to analyze the behavior of a decision system in order to correct and enhance its performance. It is also useful when trying to evaluate the distance between two systems giving different BOEs. It is also helpful to determine if a source of information regularly gives an answer that is far from other sources, so that this faulty source can be weighted or discarded. Different approaches to deal with conflicting or unreliable sources are proposed in [3, 4, 5].

Although the Dempster-Shafer Theory (DST) has many advantages, such as its ability to deal with uncertainty and ignorance, it has the problem of becoming quickly computationally heavy as it is an NP-hard problem [6]. To alleviate this computational burden, many approximation techniques of belief functions exist [7, 8, 9]. References [10, 11] show implementations and a comparative study of some approximation techniques.

To be able to efficiently evaluate the various approximation techniques, one needs some form of metric. The Jousselme distance between two bodies of evidences [12] is one of them. However, there is a problem with this metric: it requires the computation of the cardinal of a given set, an operation which is very costly computation-wise within the DST. Alternatives to the Jousselme distance are thus needed. This is the objective of the research we present here.

A. *The Dempster-Shafer Theory in Information Retrieval*

The authors of [13] use the DST to combine the visual and textual measures for ranking choosing the best word to use as annotation for an image. The DST is also used in the modeling of uncertainty in Information Retrieval (IR) applied to structured documents. We find in [14] that the use of the DST is due to: (i) it's ability to represent leaf objects; (ii) it's ability to capture uncertainty and the aggregation operator it provides, allowing the expression of uncertainty with respect to aggregated components; and (iii) the properties of the aggregation operator that are compatible with those defined by the logical model developed by [15].

Extensible Markup Language (XML) IR, by contrast to traditional IR, deals with documents that contain structural markups which can be used as hints to assess the relevancy of individual elements instead of the whole document. Reference [16] presents how the DST can be used in the weighting of elements in the document. It is also used to express uncertainty and to combine evidences derived from different inferences, providing relevancy values of all elements of the XML document.

Good mapping algorithms that perform efficient syntactic and semantic mappings between classes and their properties in different ontologies is often required for Question Answering systems. For that purpose, a multi-agent framework was proposed in [17]. In this framework, individual agents perform the mappings, and their beliefs are combined using the DST. In that system, the DST is used to deal with the uncertainty related to the use of different ontologies. The authors also use similarity assessment algorithms between concepts (words) and inherited hypernyms; once using BOE to represent information, metrics between BOE could be used to accomplish this.

As shown in [18], the fundamental issues in IR are the selection of an appropriate scheme/model for document representation and query formulation, and the determination of a ranking function to express the relevance of the document to the query. The authors

doi:10.4304/jetwi.4.3.264-272

compare IR systems based on probability and belief theories, and note a series of advantages and disadvantages with the use of the DST in IR. Putting aside the issue of computational complexity, they come to the conclusion that the DST is the better option, thanks to its ability to deal with uncertainty and ignorance.

The most significant differences between DST and probability theory are the explicit representation of uncertainty and the evidence combination mechanism. This can allow for more effective document processing [19]. It is also reported by [20] that the uncertainty occurring in IR can come from three sources regarding the relation of a document to a query: (i) in the existence of different evidences; (ii) due to unknown number of evidences; and (iii) in the existence of incorrect evidences. There is thus a clear benefit to using a method that can better combine evidences and handle their uncertainty. Interested readers are encouraged to consult [21] for an extensive study of the use of Dempster-Shafer Theory to Information Retrieval.

II. Background

A. Dempster-Shafer Theory of Evidence

Dempster-Shafer Theory (DST) has been in use for over 40 years [1-2]. The theory of evidence or DST has been shown to be a good tool for representing and combining pieces of uncertain information. The DST of evidence offers a powerful approach to manage the uncertainties within the problem of target identity. DST requires no a priori information about the probability distribution of the hypothesis; it can also resolve conflicts and can assign a mathematical meaning to ignorance.

However, traditional DST has the major inconvenience of being an NP-hard problem [6]. As various evidences are combined over time, Dempster-Shafer (DS) combination rules will have a tendency to generate more and more propositions (i.e. focal elements), which in turn will have to be combined with new input evidences. Since this problem increases exponentially, the number of retained solutions must be limited by some approximation schemes, which truncate the number of such propositions in a coherent (but somewhat arbitrary) way. Let Θ be the frame of discernment, i.e. the finite set of n mutually exclusive and exhaustive hypotheses $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$. The power set of Θ, 2^Θ is the set the $2^n - 1$ subsets of Θ, $2^\Theta = \{\emptyset, \{\theta_1\}, \{\theta_2\}, \dots, \{\theta_n\}, \{\theta_1 \cup \theta_2\}, \dots \{\theta_1 \cup \theta_2 \cup \dots \cup \theta_n\}\}$, where $\emptyset$ denotes the empty set.

1) Belief functions:

Based on the information provided by sensor sources and known a priori information (i.e. a knowledge base), a new proposition is built. Then, based on this proposition, a Basic Probability Assignment (BPA or mass function) is generated, taking into account some uncertainty or vagueness. Let us call $m_{t=0}$, the new incoming BPA. The core of the fusion process is the combination of m_t and the BPA at the previous time, m_{t-1}. The resulting BPA at time t, m_t, is then the support for decision making. Using different criteria, the best candidate for identification is selected from the database. On the other hand, m_t must be combined with a new incoming BPA and thus becomes m_{t+1}. However, this step must be preceded by a proposition management step, where m_t is approximated. Indeed, since the combination process is based on intersections of sets, the number of focal elements increases exponentially and rapidly becomes unmanageable. This proposition management step is a crucial one as it can influence the entire identification process.

The Basic Probability Assignment is a function m such that $m: 2^\Theta \to [0,1]$ which satisfies the following conditions:

$$\sum_{\forall A \in 2^\Theta} m(A) = 1 \quad (1)$$

$$m(\emptyset) = 0 \quad (2)$$

Where $m(A)$ is called the mass. It 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. The elements of 2^Θ that have a non-zero mass are called focal elements. Given a BPA m, two functions from 2^Θ to $[0,1]$ are defined: a belief function Bel, and a plausibility function Pl such that

$$\mathrm{Bel}(A) = \sum_{B \subseteq A} m(B) \quad \forall A, B \in 2^\Theta \quad (3)$$

$$\mathrm{Pl}(A) = \sum_{A \cap B \neq \emptyset} m(B) \quad \forall A, B \in 2^\Theta \quad (4)$$

It can also be stated that $\mathrm{Pl}(A) = 1 - \mathrm{Bel}(\overline{A})$, where $\overline{A}$ is the complement of A and $\mathrm{Bel}(A)$ measures the total belief that the object is in A, whereas $\mathrm{Pl}(A)$ measures the total belief that can move into A. The functions m, Bel and Pl are in one-to-one correspondence, so it is equivalent to talk about any one of them or about the corresponding body of evidence.

2) Conflict definition:

The conflict K corresponds to the sum of all masses for which the set intersection yield the null set $\emptyset$. K is called the conflict factor and is defined as:

$$K = \sum_{A \cap B = \emptyset} m_1(A) m_2(B) \quad A, B \subseteq \Theta \quad (5)$$

K measures the degree of conflict between m_1 and m_2: $K = 0$ corresponds to the absence of conflict, whereas $K = 1$ implies a 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 we get $K = 1$ if and only if all the sets resulting from this combination are empty.

3) Dempster-Shafer Combination Formulae:

In DST, a combined or “fused” mass is obtained by combining the previous $m_1(A)$ (presumably the results of previous fusion steps) with a new $m_2(B)$ to obtain a fused result as follows:

$$(m_1 \oplus m_2)(C) = \frac{1}{1-K} m_c(C) \quad \forall C \subseteq \Theta \quad (6)$$

$$m_c(C) = \sum_{A \cap B} m_1(A) m_2(B) \quad (7)$$

TABLE I.
CARDINALITIES FOR DST AND DSMT

Cardinal of Θ	2	3	4	5	6
Cardinal of 2^{Θ}	4	8	16	32	64
Cardinal of D^{Θ}	5	19	167	7,580	7,828,353

The renormalization step using the conflict K, corresponding to the sum of all masses for which the set intersection yields the null set, is a critical feature of the DS combination rule. Formulated as is equation (6), the DS combination rule is associative. Many alternative ways of redistributing the conflict lose this property. The associativity of the DS combination rule is critical when the timestamps of the sensor reports are unreliable. This is because an associative rule of combination is impervious to a change in the order of reports coming in. By contrast, other rules can be extremely sensitive to the order of combination.

B. Dezert-Smarandache Theory

The Dezert-Smarandache Theory (DSmT) [22, 23, 24] encompasses DST as a special case, namely when all intersections are null. Both the DST and the DSmT use the language of masses assigned to each declaration from a sensor. A declaration is a set made up of singletons of the frame of discernment Θ, and all sets that can be made from them through unions are allowed (this is referred to as the power set 2^{Θ}). In DSmT, all unions and intersections are allowed for a declaration, this forming the much larger hyper power set D^{Θ} which follows the Dedekind sequence.

For a case of cardinality 3, $\Theta = \{\theta_1, \theta_2, \theta_3\}$, with $|\Theta| = 3$, D^{Θ} is still of manageable size:

$D^{\Theta}(|\Theta| = 3) \equiv \{\{\theta_1\}, \{\theta_2\}, \{\theta_3\}, \{\theta_1 \cup \theta_2\}, \{\theta_1 \cup \theta3, \theta2 \cup \theta3, \theta1 \cup \theta2 \cup \theta3,$
$\{\theta_1 \cap \theta_2\}, \{\theta_1 \cap \theta_3\}, \{\theta_2 \cap \theta_3\}, \{\theta_1 \cap \theta_2 \cap \theta_3\},$
$\{(\theta_1 \cup \theta_2) \cap \theta_3\}, \{(\theta_1 \cup \theta_3) \cap \theta_2\}, \{(\theta_2 \cup \theta_3) \cap \theta1,$
$\{(\theta_1 \cap \theta_2) \cup \theta_3\}, \{(\theta_1 \cap \theta_3) \cup \theta_2\}, \{(\theta_2 \cap \theta_3) \cup \theta1, (\theta1 \cap \theta2) \cup \theta1 \cap \theta3 \cup (\theta2 \cap \theta3)\}$ (8)

For larger cardinalities, the hyper power set makes computations prohibitively expensive (in CPU time). Table I illustrates the problem with the first few cardinalities of 2^{Θ} and D^{Θ}.

1) Dezert-Smarandache Hybrid Combination Formulae:

In DSmT, the hybrid rule [22, 23, 24] appropriate for constraints turns out to be much more complicated:

$$m_{M(\Theta)}(A) = \phi(A)[S_1(A) + S_2(A) + S_3(A)] \quad \forall A \in D^{\Theta} \quad (9)$$

$$S_1(A) = \sum_{X_1 \cap X_2 = A} m_1(X_1) m_2(X_2) \quad \forall X_1, X_2 \in D^{\Theta} \quad (10)$$

$$S_2(A) = \sum_{\substack{[(u(X_1) \cup u(X_2)) = A] \\ \vee \\ \left[\begin{smallmatrix} ((u(X_1) \cup u(X_2)) \in \emptyset) \\ \wedge \\ (A = I_t) \end{smallmatrix}\right]}} \begin{matrix} m_1(X_1) m_2(X_2) \\ \forall X_1, X_2 \in \emptyset \end{matrix} \quad (11)$$

$$S_3(A) = \sum_{X_1 \cup X_2 = A} m_1(X_1) m_2(X_2) \quad \begin{matrix} \forall X_1, X_2 \in D^{\Theta} \\ X_1 \cap X_2 \in \emptyset \end{matrix} \quad (12)$$

The reader is referred to a series of books on DSmT [22, 23, 24] for lengthy descriptions of the meaning of this formula. A three-step approach is proposed in the second of these books, which is used in this technical report. From now on, the term “hybrid” will be dropped for simplicity.

C. Pignistic Transformation

1) Classical Pignistic Transformation:

One of the most popular transformations is the pignistic transformation proposed by Smets [25] as basis for decision in the evidential theory framework. The decision rule based on a BPA m is:

$$\Pr\{A\} = \sum_{X \in D^{\Theta}} \frac{m(X \cap A)}{|X|} \quad (13)$$

$$\theta^* = \operatorname{Arg}[\max_{\theta \in \Theta} \operatorname{BetP}(\theta, m)] \quad (14)$$

$$\operatorname{BetP}(\theta, m) = \sum_{\theta \in A \subseteq \Theta} \frac{m(A)}{|A|} \quad (15)$$

with θ^* the identified object among the objects in Θ.

This decision presents the main advantage that it takes into account the cardinality of the focal elements.

2) DSm Cardinal:

The Dezert-Smarandache (DSm) cardinal [22, 23, 24] of a set A, noted $C_M(A)$, accounts for the total number of partitions including all intersection subsets. Each of these partitions possesses a numeric weight equal to 1, and thus they are all equal. The DSm cardinal is used in the generalized pignistic transformation equation to redistribute the mass of a set A among all its partitions B such that $B \subseteq A$.

3) Generalized Pignistic Transformation:

The mathematical transformation that lets us go from a representation model of belief functions to a probabilistic model is called a generalized pignistic transformation [22, 23, 24]. The following equation defines the transformation operator.

$$\Pr\{A\} = \sum_{X \in D^{\Theta}} \frac{C_M(X \cap A)}{C_M(X)} m(X) \quad \forall A \in D^{\Theta} \quad (16)$$

D. Jousselme Distance between two BOEs

1) Similarity Properties:

Diaz and al. [26] expects that a good similarity measure should respect the following six properties:

$$S'(A, B) \in [0,1] \quad \text{normalization} \quad (17)$$

$$S'(A, B) = S'(B, A) \quad \text{symmetry} \quad (18)$$

$$\begin{matrix} \text{Both increasing on} & |A \cap B| \\ \text{and decreasing on} & |A - B| \text{ and } |B - A| \end{matrix} \quad (19)$$

$$S'(A, B) = 1 \Leftrightarrow A = B \quad \text{identity of indiscernible} \quad (20)$$

$$S'(A, B) = 0 \Leftrightarrow A \cap B = \emptyset \quad \text{exclusiveness} \quad (21)$$

$$\text{Decreasing on} \quad R = \frac{|A \cup B|}{|\Theta|} \quad (22)$$

2) Jaccard Similarity Measure:

The Jaccard similarity measure [27] is a statistic used for comparing the similarity and diversity of sample sets. It was originally created for species similarity evaluation.

$$S(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (23)$$

3) Distances Properties:

A distance function, also called a distance metric, on a set of points $\mathbb{S}$ is a function $d: \mathbb{S} x \mathbb{S} \rightarrow \mathbb{R}$ with four properties [28, 29]; suppose $x, y \in \mathbb{S}$:

$$d(x,y) \geq 0 \quad \text{non-negativity} \quad (24)$$

$$d(x,y) = 0 \text{ iff } x = y \quad \text{identity of indiscernible} \quad (25)$$

$$d(x,y) = d(y,x) \quad \text{symmetry} \quad (26)$$

$$d(x,z) \leq d(x,y) + d(y,z) \quad \text{triangle inequality} (27)$$

Some authors also require that $\mathbb{S}$ be non empty.

4) Jousselme Distance:

To analyze the performance of approximation algorithm, to compare the proximity to non-approximated versions, or to analyze the performance of the DS fusion algorithm comparing the proximity with the ground truth if available, the Jousselme distance measures can be used [12]. The Jousselme distance is an Euclidean distance between two BPAs. Let m_1 and m_2 be two BPAs defined on the same frame of discernment Θ , the distance between m_1 and m_2 is defined as:

$$d_j(m_1, m_2) = \sqrt{\frac{1}{2}\langle(m_1 - m_2), (m_1 - m_2)\rangle} \quad (28)$$

$$\langle m_1, m_2 \rangle = \sum_{A \subseteq \Theta} \sum_{B \subseteq \Theta} m_1(A) m_2(B) S(A,B) (29)$$

where $S(A,B)$ is the Jaccard similarity measure

III. New Metrics

A. Extension of the Jousselme distance to the DSmT

The Jousselme distance as defined originally in [12] can work without major changes, as it is within the DSm framework. The user simply has to use two BPAs defined over the DSm theory instead of BPAs defined within the DS theory. Boundaries, size, and thus amount of computation will of course be increased. But otherwise, there is no counter indication to using this distance in DSmT. We thus can keep equation (28) as the definition of Jousselme distance within DSmT, with the definition of the DSm Cardinal.

Tables II and III show the bodies of evidences and their distances one-to-another. The example was realized with a discernment frame of size three ($|\Theta| = 3$), so that the cardinal of its hyper power set would be $|D^{\Theta}|$ =19 for the free model, as defined by Dezert and Smarandache [22]. Table II is divided into three sections, each one of them represents data for one BOE. The three columns give the focal sets, associated BPA value, and the cardinal of that set.

TABLE II
FIRST SERIES OF THREE BODIES OF EVIDENCES

	$m_1(\cdot)$	$\lvert\cdot\rvert$
A	0.8	4
$A \cap B$	0.1	2
$A \cup B$	0.1	6

	$m_2(\cdot)$	$\lvert\cdot\rvert$
B	0.8	4
$A \cap B$	0.1	2
$A \cup B$	0.1	6

	$m_3(\cdot)$	$\lvert\cdot\rvert$
$A \cap B$	0.8	2
$A \cup B$	0.1	6
$A \cap C$	0.1	2

Pairwise computation between the different pairs of BOEs took quite some time with all the required calculations by the Jousselme distance of evidences. The results are shown in Table III. The proof of respect of all properties has already been done for the DST in [12].

The difference with the original version of the distance presented in [12] is the allowed presence of intersections which creates the hyper power set from the power set. This difference adds up possibilities of more computations to get to the distance value. More specifically, the cardinal evaluation part of the Jousselme distance is worsened by the hyper power set increase in size when compared to the power set.

B. Hamming-inspired metric on evidences

1) Continuous XOR mathematical operator:

In [30], Weisstein define the standard OR operator noted $\vee$ as a connective in logic which yields true if any one of a sequence conditions is true, and false if all conditions are false.

In [31], Germundsson and Weisstein define the standard XOR logical operator ($\veebar$) as a connective in logic known as the exclusive OR or exclusive disjunction. It yields true if exactly one, but not both, of two conditions are true. This operator is typically designed as symmetric difference in set theory [32]. As such, the authors define it as the union of the complement of A with respect to B and B with respect to A. Figure 1 is a Venn diagram displaying binary XOR operator on numerical discrete values in Figure 1.

TABLE III
EXTENTED JOUSSELME DISTANCE RESULTS

	d_j
$d_j(m_1, m_2)$	0.65
$d_j(m_2, m_3)$	0.52
$d_j(m_1, m_3)$	0.52

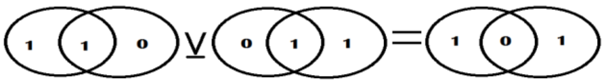

Figure 1. Venn diagram displaying binary XOR operator on numerical discrete values.

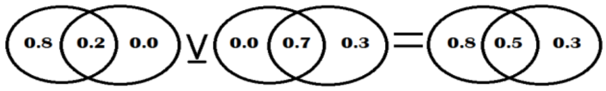

Figure 2. Venn diagram displaying continuous XOR operator.

Starting with the standard XOR logical operator and inspired by the Hamming distance [33] which uses a symmetric difference implicitly, we develop the idea of a continuous XOR operator. Figure 2 shows a simple case similar to that of previous figure but using values from $\mathbb{R}$. We can see that it is working as an absolute value of the difference applied on each partition of the Venn diagrams individually one to another.

2) Metric between evidences based on Hamming distance principle:

The Hamming distance [33] between two strings is the minimum number of substitutions required to change one string to another. In other words, it is defined by the sum of absolute values of differences. From this, with the DSm cardinal [22], and using a continuous XOR mathematical operator, we have developed a new distance, the Hamming Distance of Evidences (HDE). This distance is bounded within normal values, such that $d_{HDE} \in [0,1]$. This new distance also respects the properties of equations (24-27): non-negativity, identity of indiscernibility, symmetry, and the triangle inequality.

The HDE is defined as in equation (30), which uses the $\mathrm{X_i}$ defined in equation (31), and where $\mathrm{S}^{\Theta} \triangleq (\Theta,\cup,\cap,c(.))$ is the super-power set. For example, in the case where we have a discernment frame such as $\Theta = \{\theta_1,\theta_2\}$, we would obtain the following super-power set $\mathrm{S}^{\Theta}(|\Theta| = 2) \equiv \{\{\theta_1\},\{\theta_2\},\{\theta_1 \cup \theta_2\}, \{\theta_1 \cap \theta_2\},\{\emptyset\}, c(\{\theta_1\}), c(\{\theta_2\}), c(\{\theta_1 \cup \theta_2\}), c(\{\theta_1 \cap \theta_2\}), c(\{\emptyset\})\}$.

$$\mathrm{HDE(X,Y)} = \frac{1}{2}\sum\nolimits_{\forall \mathrm{X_i,Y_j}} \mathrm{X_i} \veebar \mathrm{Y_j} \tag{30}$$

$$\sum_{\substack{\mathrm{X_i} \subseteq \mathrm{A} \\ \mathrm{X_i} \in \mathrm{S}^{\Theta}}} \mathrm{m(X_i)} = \frac{\mathrm{m(A)}}{|\mathrm{A}|} \tag{31}$$

The HDE uses the BPA mass distributed among the different parts (sets) in S^{Θ} that composes the BPA from D^{Θ}. This transition from D^{Θ} to S^{Θ} is done using equation (31). Using the super-power set version of the BPA gets us a more refined and precise definition of it.

Once in the super-power set framework, we use an adaptation of the Hamming distance or the continuous XOR operation defined previously. Its implementation is more easily understood as a summation of the absolute of the differences[1] between the BPAs in S^{Θ} divided by 2.

For BOEs defined in Table II in the previous section, without any constrained set, we get the results given in Table IV. Then, we can easily compare relative distances to have a reliable point of reference. The Jousselme distance is considered to be our distance of reference.

[1] This is equivalent to the *symmetric difference* expression used to define XOR operator in literature [32].

C. Metric using a degree of shared uncertainty

1) Similarity coefficient of degree of shared uncertainty:

The idea behind a similarity coefficient of degree of shared uncertainty is to quantify the degree of shared uncertainty that lies behind a pair of sets. We want to avoid the use of cardinal operators. We conceived a decision tree test which will evaluate the degree of shared uncertainty. The following equation shows what the coefficient of similarity between a pair of sets is when using the metric that we suggest.

$$\varphi_{\mathrm{A\times B}} = \begin{cases} 3, & \mathrm{A = B} \\ 2, & \mathrm{A \subset B | A \supset B} \\ & \mathrm{A \cap B \neq \emptyset} \\ 1, & \mathrm{A \not\subset B | A \not\supset B} \\ & \mathrm{A \neq B} \\ 0, & \mathrm{A \cap B = \emptyset} \end{cases} \tag{32}$$

Equation (32) gives a coefficient value of 3 when the pair of sets is equal; the value 2 when one of the sets is included in the other one, and 1 when the sets give a non-empty intersection but none is included in another nor being equal. Finally, the coefficient has a value of 0 when the intersection between the pair of sets is the empty set. The maximum value that the coefficient of similarity $\varphi_{\mathrm{A\times B}}$ has between sets A and B is 3.

2) Metric between evidences based on a degree of shared uncertainty:

From the similarity coefficient of degree of shared uncertainty as defined above, we get the following distance, noted DSU_{init} and defined in equation (33). In that equation, the factor $1/|\mathfrak{F}_\mathrm{i}||\mathfrak{F}_\mathrm{j}| \times \max(\varphi)$ is a normalization factor *N* required to bound of the distance. The summation over $\forall \mathrm{A} \in \mathfrak{F}_\mathrm{i} \times \forall \mathrm{B} \in \mathfrak{F}_\mathrm{j}$ symbolizes a sum going over the matrix of every possible pair of sets from focal elements $\mathfrak{F}_\mathrm{i}, \mathfrak{F}_\mathrm{j}$.

$$\mathrm{DSU}_{init}(\mathrm{X,Y}) = 1 - \frac{1}{|\mathfrak{F}_\mathrm{i}||\mathfrak{F}_\mathrm{j}| \times \max(\varphi)} \sum\nolimits_{\forall \mathrm{A} \in \mathfrak{F}_\mathrm{i} \times \forall \mathrm{B} \in \mathfrak{F}_\mathrm{j}} \varphi_{\mathrm{A\times B}} \tag{33}$$

Even if we consider (33) as the distance using $\varphi_{\mathrm{A\times B}}$ similarity coefficient, we might want to consider the possibility of building one that uses only a triangular matrix out of the matrix-domain of the summation. However, since commutativity is a built-in property, this measure will have a bit of useless redundancy.

TABLE IV
HAMMING DISTANCE ON EVIDENCES RESULTS

	$\mathrm{d_{HDE}}$
$\mathrm{d_{HDE}(m_1, m_2)}$	0.40
$\mathrm{d_{HDE}(m_2, m_3)}$	0.40
$\mathrm{d_{HDE}(m_1, m_3)}$	0.35

Equation (33) could be expressed in the simple form: $\mathrm{DSU}_{init}(\mathrm{X},\mathrm{Y}) = 1 - \mathrm{Z}$, where Z is a similarity factor. Since distances use dissimilarity factors (so that a distance of 0 means that X = Y), a subtraction from 1 is required. However, the idea of a distance solely based on equation (33) isn't enough. One should consider weighting similarities with mass value from BPAs in order to really represent the distance between bodies of evidences and not only a combination of sets. We propose (34) as a final equation for that reason.

$$\mathrm{DSU}(\mathrm{X},\mathrm{Y}) = 1 - \frac{1}{\max(\varphi)} \sum_{\forall A \in \mathfrak{F}_i \times \forall B \in \mathfrak{F}_j} \mathrm{m}_{\mathfrak{F}_i}(\mathrm{A}) \times \mathrm{m}_{\mathfrak{F}_j}(\mathrm{B}) \varphi_{\mathrm{A}\times\mathrm{B}} \quad (34)$$

Table V uses a simple case to show the inner workings of this method. The first matrix shown in the table is a computation matrix with the degree of shared uncertainty $\varphi_{\mathrm{A}\times\mathrm{B}}$, defined in (32), and the product of the masses of the pair of sets $\mathrm{m}_{\mathfrak{F}_i}(\mathrm{A}) \times \mathrm{m}_{\mathfrak{F}_j}(\mathrm{B})$. The second matrix gives the value of the weighted similarity values. Finally, the last table in Table V indicates the sum of the values from within the previous matrix, or the value of the sum in equation (34), the normalization factor and finally the Distance of Shared Uncertainty (DSU).

This distance could be qualified as discrete in the sense that not all values of $\mathbb{R}$ will be possible for DSU in any case of distance measurement. However, that is true only for fixed values of BPA. Since BPA values are continuous in $[0,1]$ then $\mathrm{DSU} \epsilon [0,1]$.

Table VI shows the results of the metric based on the degree of shared uncertainty measurements on the same BOEs described in Table II as previously experimented on at Tables III for Jousselme distance and Table IV for Hamming distance on evidences.

TABLE V
SIMPLE CASE OF METRIC BASED ON SHARED UNCERTAINTY DEGREE

m_1 \ m_2	A 0.8	B 0.2
$A \cap B$ 0.8	**2** 0.64	**2** 0.16
$A \cup B$ 0.2	**2** 0.16	**3** 0.04

m_1 \ m_2	A 0.8	B 0.2
$A \cap B$ 0.8	**1.28**	**0.32**
$A \cup B$ 0.2	**0.32**	**0.12**

$\sum$	2.04
$\sum/N$	0.68
DSU	0.32

TABLE VI
METRIC BASED ON SHARED UNCERTAINTY DEGREE RESULTS

	$\mathrm{d}_{\mathrm{DSU}}$
$\mathrm{d}_{\mathrm{DSU}}(\mathrm{m}_1,\mathrm{m}_2)$	0.513
$\mathrm{d}_{\mathrm{DSU}}(\mathrm{m}_2,\mathrm{m}_3)$	0.333
$\mathrm{d}_{\mathrm{DSU}}(\mathrm{m}_1,\mathrm{m}_3)$	0.307

IV. EXAMPLES AND PERFORMANCES

This section explores the metrics presented in the previous section. Theses metrics will be used as distance measurements. We have implemented a DST, DSmT combination system within Matlab™. The details explaining how DSmT was implemented appear in [34, 35]. Functions have been added in that system for the execution of the computation of various metrics.

A. A few simple examples

1) Exploration case 1:

Using the same bodies of evidences as presented in Table II, we obtained the results and times given in Table VII for the execution, in seconds, for the same inputs given to the three distances presented previously: the d_j, the HDE and DSU. Based only on this data, it is difficult to choose which metric is best. However we can already see, as expected, that the Jousselme distance would be difficult to use in real-time complex cases due to the computation time it requires.

2) Exploration case 2:

This case further explores the behaviors of the distance metrics. We will use two bodies of evidences. The first will be fixed with the following values: $\{\mathrm{m}_1(C) = 0.8, \mathrm{m}_1(A \cup B \cup C) = 0.2\}$. For the second BOE, we will increment successively the mass of one focal element nine times, reducing from the same value the mass of the second focal element such as $\{\mathrm{m}_2(C) = i/10, \mathrm{m}_2(A \cup B \cup C) = (1 - \frac{i}{10})\}$, where $i \in [1:9]$.

The results of this exploration case are given in Table VIII. We can notice from that table that DSU is not able to correctly consider distances with the mass distributions. Obviously, this is an undesirable behavior occurring for the situation with a pair of BOE with identical sets.

We can also see that the HDE and Jousselme distance responds in a symmetric manner to the symmetric mass distribution around equal BOEs. In other words, steps $i = 7$ and $i = 9$ gives equal values, as they should. For step $i = 8$, all metrics gives the proper distance of zero.

TABLE VII
DISTANCE AND TIME OF EXECUTION VALUES FOR CASE 1

	$\mathrm{d}(\mathrm{m}_1,\mathrm{m}_2)$		$\mathrm{d}(\mathrm{m}_2,\mathrm{m}_3)$		$\mathrm{d}(\mathrm{m}_1,\mathrm{m}_3)$	
	Dist.	time	Dist.	time	Dist.	time
d_j	0.653	1.621	0.534	0.090	0.546	0.099
HDE	0.4	0.027	0.35	0.018	0.4	0.018
DSU	0.54	0.101	0.307	0.014	0.333	0.016

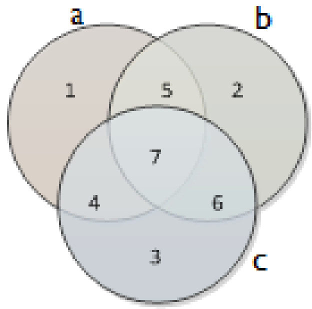

Figure 3. Venn diagram with the 7 partitions of a size 3 case.

Table IX shows that both HDE and DSU demonstrate a clear advantage over Jousselme distance in terms of execution times.

3) Exploration case 3:

Figure 3 shows the 7 possible partitions of a size 3 case. This case proceeds a little differently from the previous two. Instead of keeping identical BOEs with varying masses, the BOEs are now varied. A third and fourth focal elements in some of the BOEs are introduced for that purpose. The first BOE is always the same: $\{\mathrm{m}_1(b) = 0.8, \mathrm{m}_1(a \cup b \cup c) = 0.2\}$. The BOEs used as the second one in the pairwise distances are listed here:

A. $\{\mathrm{m}_2(b) = 0.7, \mathrm{m}_2(b \cup c) = 0.1, \mathrm{m}_2(a \cup b \cup c=0.2\}$
B. $\{\mathrm{m}_2(b) = 0.6, \mathrm{m}_2(b \cup c) = 0.2, \mathrm{m}_2(a \cup b \cup c=0.2\}$
C. $\{\mathrm{m}_2(b) = 0.7, \mathrm{m}_2(b \cap c) = 0.1, \mathrm{m}_2(a \cup b \cup c=0.2\}$
D. $\{\mathrm{m}_2(b) = 0.6, \mathrm{m}_2(b \cap c) = 0.2, \mathrm{m}_2(a \cup b \cup c) = 0.2\}$
E. $\{\mathrm{m}_2(b) = 0.6, \mathrm{m}_2(b \cap c) = 0.1, \mathrm{m}_2(b \cup c) = 0.1, \mathrm{m}_2(a \cup b \cup c) = 0.2\}$
F. $\{\mathrm{m}_2(b) = 0.1, \mathrm{m}_2(b \cap c) = 0.6, \mathrm{m}_2(b \cup c) = 0.1, \mathrm{m}_2(a \cup b \cup c) = 0.2\}$
G. $\{\mathrm{m}_2(b) = 0.1, \mathrm{m}_2(b \cap c) = 0.1, \mathrm{m}_2(b \cup c) = 0.6, \mathrm{m}_2(a \cup b \cup c) = 0.2\}$

The results of this case are given in Table X. As expected, we can observe a *Distance Variation* (Δd) increase for the following pairs: $\Delta d(A \to B)$, $\Delta d(C \to D)$,

TABLE VIII
DISTANCE VALUES FOR CASE 2

i	$\mathrm{d_j}$	HDE	DSU
1	0.458	0.300	0.000
2	0.393	0.257	0.000
3	0.327	0.214	0.000
4	0.236	0.171	0.000
5	0.196	0.128	0.000
6	0.131	0.086	0.000
7	0.065	0.043	0.000
8	0.000	0.000	0.000
9	0.065	0.043	0.000

TABLE IX
EXECUTION TIME VALUES FOR CASE 2

i	$\mathrm{d_j}$	HDE	DSU
1	0.241	0.027	0.078
2	0.043	0.010	0.008
3	0.043	0.010	0.007
4	0.043	0.010	0.007
5	0.043	0.010	0.006
6	0.040	0.010	0.006
7	0.041	0.010	0.007
8	0.040	0.010	0.006
9	0.041	0.010	0.006

$\Delta d(E \to F)$ and $\Delta d(E \to G)$. The notation $\Delta d(X \to Y)$ signifies that the observed distance variation going from case X to Y is increasing.

For the interesting cases F and G, we have $\Delta d(E \to F) > \Delta d(E \to G)$. The difference between F and G is that the mass of $\{b\}$ goes to $\{b \cap c\}$ in F, while in G it mainly goes to $\{b \cup c\}$.

DSU metric for case F is equal to case G, in all the other metrics they give smaller values for case G when compared to case F. Similar conclusions are obtained when comparing metrics for the pairs of cases (A, C), and (B,D): for similar mass redistribution, when giving the mass to a disjunction the resulting distance is smaller than if it were to be distributed to an intersection.

3) Exploration's conclusions:

In general, it is better for identical sets to have lowest distance. Otherwise, a minimal number of sets will minimize the distribution of mass onto unshared partitions. With no identical partitions in common, it is preferable to have a higher mass onto disjunctive sets which have more common partitions. Also, it is better to have disjunctive sets as specific as possible; in other words, of lowest cardinality. Hence, too much mass given to a set that has too many uncommon partitions with the targeted ID or ground truth must be avoided. To get distances values such as $\Delta d(BOE_1 \to BOE_2) < \Delta d(BOE_1 \to BOE_3)$, one needs masses in BOE_2 to be distributed on sets that have a higher ratio of common partitions with BOE_1 than the sets of BOE_3 would have.

Finally the use of either Jousselme (adapted to DSmT) or the DHE, which is much quicker, is recommended.

TABLE X
DISTANCE VALUES FOR CASE 3

	$\mathrm{d_j}$	HDE	DSU
A	0.058	0.033	0.133
B	0.115	0.067	0.160
C	0.071	0.050	0.133
D	0.141	0.100	0.160
E	0.100	0.067	0.160
F	0.440	0.317	0.293
G	0.367	0.200	0.293

V. CONCLUSIONS

This paper introduced two new distances between evidences for both the Dempster-Shafer Theory and Dezert-Smarandache Theory to replace the Jousselme distance.

When the size of the discernment frame gets high: the distance calculation becomes too big to handle in a reasonable amount of time. In time critical systems, it would be better to use the Hamming distance of evidences. For the distance using the degree of shared uncertainty DSU, studies must be done further. A correction may be required to prevent it from considering masses properly when facing identical bodies of evidences.

Future works would include the use of DSmT [22, 23, 24] and its hierarchical information representation abilities in conjunction with approximation of belief functions algorithms in Information Retrieval.

ACKNOWLEDGMENT

The authors wish to thank the reviewers for their comments. This work was carried out as part of Pascal Djiknavorian's doctoral research at Université Laval. Pascal Djiknavorian study was partly funded by RDDC.

REFERENCES

[1] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, USA, 1976.
[2] A. Dempster, "Upper and lower probabilities induced by multivalued mapping", *The Annals of Mathematical Statistics*, vol. 38, pp. 325-339, 1967.
[3] M. C. Florea and E. Bosse, "Dempster-Shafer Theory: combination of information using contextual knowledge", in Proceedings of 12th International Conference on Information Fusion, Seattle, WA, USA, July 6-9, 2009, pp. 522-528.
[4] S. Le Hegarat-Mascle, I. Bloch, and D. Vidal-Madjar, "Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, issue: 4, pp.1018-1031, August 1997.
[5] J. Klein and O. Colot, "Automatic discounting rate computation using a dissent criterion", Proceedings of Workshop on the Theory of Belief Functions, Brest, France, April 1-2, 2010.
[6] P. Orponen, "Dempster's rule of combination is # P-complete", Artificial *Intelligence*, vol. 44, no. 1-2, pp. 245–253, 1990.
[7] B. Tessem, "Approximations for efficient computation in the theory of evidence", *Artificial Intelligence*, vol. 61, pp 315-329, June 1993.
[8] M. Bauer, "Approximation Algorithms and Decision Making in the Dempster-Shafer Theory of Evidence-An Empirical study", *International Journal of Approximate Reasoning*, vol. 17, no. 2-3, pp. 217–237, 1997.
[9] D. Boily, and P. Valin, "Truncated Dempster-Shafer Optimization and Benchmarking", in Proceedings of Sensor Fusion: Architectures, Algorithms, and Applications IV, SPIE Aerosense 2000, Orlando, Florida, April 24-28, 2000, Vol. 4051, pp. 237-246.
[10] P. Djiknavorian, P. Valin and D. Grenier, "Approximations of belief functions for fusion of ESM reports within the DSm framework", Proceedings of the 13th International Conference of Information Fusion, Edinburg, UK, 2010.
[11] P. Djiknavorian, A. Martin, P. Valin and D. Grenier, « Étude comparative d'approximation de fonctions de croyances généralisées / Comparative Study of approximations of generalized beliefs functions », in Proceedings of Logique Floue et ses Applications, LFA2010, Lannion, France, Novembre 2010.
[12] A.-L. Jousselme, D. Grenier, and E. Bosse, "A new distance between two bodies of evidence", *Information Fusion*, vol. 2, no. 2, pp. 91-101, June 2001.
[13] R. Xiaoguang, Y. Nenghau, W. Taifeng, L. Mingjing, "A Search-Based Web Image Annotation Method", Proceedings of IEEE International Conference on Multimedia and Expo, 2007, pp. 655-658.
[14] M. Lalmas, "Dempster-Shafer's theory of evidence applied to structured documents: modelling uncertainty". in Proceedings of the 20th annual international ACM SIGIR, pp. 110-119, Philadelphia, PA, USA. ACM, 1997. DOI:10.1145/258525.258546
[15] Y. Chiaramella, P. Mulhem and F. Fourel, "A Model for Multimedia Information Retrieval", Technical Report, Basic Research Action FERMI 8134, 1996.
[16] F. Raja, M. Rahgozar, F. Oroumchian, "Using Dempster-Shafer Theory in XML Information Retrieval", Proceedings of World Academy of Science, Engineering and Technology, Vol. 14, August 2006.
[17] M. Nagy, M. Vargas-Vera and E. Motta, "Uncertain Reasoning in Multi-agent Ontology Mapping on the Semantic Web", in Proceedings of the Sixth Mexican International Conference on Artificial Intelligence – Special Session, MICAI 2007, November 4-10, 2007. Pp. 221-230. DOI:10.1109/MICAI.2007.11
[18] K.R. Chowdhary and V.S. Bansal, "Information Retrieval using probability and belief theory", in Proceedings of the 2011 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC), 2011, pp. 188-191. DOI:10.1109/ETNCC.2011.5958513
[19] A. Verikas, A. Lipnickas, K. Malmqvist, M. Bacauskiene, and A. Gelzinis, "Soft combination of neural classifiers: A comparative study", *Pattern Recognition Letters*, vol. 20, pp. 429-444, 1999.
[20] A.M. Fard, H. Kamyar, *Intelligent Agent based Grid Data Mining using Game Theory and Soft Computing*, Bachelor of Science Thesis, Ferdowsi University of Mashhad, September 2007.
[21] I. Ruthven and M. Lalmas, "Using Dempster-Shafer's Theory of Evidence to Combine Aspects of Information Use", *Journal of Intelligent Information Systems*, vol. 19 issue 3, pp.267-301, 2002.
[22] Smarandache, F., Dezert, J. editors. *Advances and Applications of DSmT for Information Fusion*, vol. 1, American Research Press, 2004.
[23] Smarandache, F., Dezert, J. editors. *Advances and Applications of DSmT for Information Fusion*, vol. 2, American Research Press, 2006.
[24] Smarandache, F., Dezert, J. editors. *Advances and Applications of DSmT for Information Fusion*, vol. 3, American Research Press, 2009.
[25] Ph. Smets, "Data fusion in the Transferable Belief Model", in Proceedings of the 3rd International Conference on Information Fusion, Fusion 2000, Paris, July 10-13, 2000, pp. PS21-PS33.
[26] J. Diaz, M. Rifqi, B. Bouchon-Meunier, "A Similarity Measure between Basic Belief Assignments", in Proceedings of the 9th International Conference on Information Fusion, Florence, Italy, 10-13 July, 2006.

[27] P. Jaccard, "Étude comparative de la distribution florale dans une portion des Alpes et des Jura" Bulletin de la Société Vaudoise des Sciences Naturelles, vol. 37, pp. 547–579, 1901.
[28] E.W. Weisstein, "Distance" From MathWorld--A Wolfram Web Resource, July 2012. http://mathworld.wolfram.com/Distance.html
[29] M. Fréchet, "Sur quelques points du calcul fonctionnel", *Rendiconti del Circolo Matematico di Palermo*, vol. 22, pp. 1-74, 1906.
[30] E.W. Weisstein, "OR", From MathWorld--A Wolfram Web Resource, July 2012. http://mathworld.wolfram.com/OR.html
[31] R. Germundsson and W. E. Weisstein, "XOR", From MathWorld--A Wolfram Web Resource, July 2012. http://mathworld.wolfram.com/XOR.html
[32] E.W. Weisstein, "Symmetric Difference", From MathWorld--A Wolfram Web Resource, July 2012. http://mathworld.wolfram.com/SymmetricDifference.html
[33] R.W. Hamming, "Error detecting and error correcting codes", Bell System Technical Journal, vol. 29, issue 2, pp. 147–160, 1950.
[34] P. Djiknavorian, D. Grenier, "Reducing DSmT hybrid rule complexity through optimization of the calculation algorithm", in *Advances and Applications of DSmT for Information Fusion*, F. Smarandache and J. Dezert, editors, American Research Press, 2006.
[35] P. Djiknavorian, "Fusion d'informations dans un cadre de raisonnement de Dezert-Smarandache appliqué sur des rapports de capteurs ESM sous le STANAG 1241", Master's Thesis, Université Laval, 2008.

Pascal Djiknavorian received a B.Eng. in computer engineering and a certificate in business administration in 2005 from Laval University. From there, he also completed in 2008 a M.Sc. in electrical engineering on information fusion within the Dezert-Smarandache theory framework applied to ESM reports under STANAG 1241.

He is currently a Ph.D. student in information fusion at Laval University and is supervised by Professor Dominic Grenier and Professor Pierre Valin. He has a dozen publications as book's chapters, in journals, and conference's proceedings. His research interests include evidential theory, Dezert-Smarandache theory, approximation algorithms, and optimization methods.

Mr. Djiknavorian is a graduate student member of the IEEE.

Dominic Grenier received the M.Sc. and Ph.D. degrees in electrical engineering in 1985 and 1989, respectively, from the Université Laval, Quebec City, Canada.

From 1989 to 1990, he was a Postdoctoral Fellow in the radar division of the Defense Research Establishment in Ottawa (DREO), Canada. In 1990, he joined the Department of Electrical Engineering at Université Laval where he is currently a Full Professor since 2000. He was also co-editor for the Canadian Journal on Electrical and Computer Engineering during 6 years.

Recognized by the undergraduate students in electrical and computer engineering at Université Laval as the electromagnetism and RF specialist, his excellence in teaching has resulted in his being awarded the "Best Teacher Award" many times. He obtained in 2009 one special fellowship from the Quebec Minister for education.

His research interests include inverse synthetic aperture radar imaging, signal array processing for high resolution direction of arrivals and data fusion for identification.

Prof. Grenier has 32 publications in refereed journals and 75 more in conference proceedings. In addition, 33 graduate students completed their thesis under his direction since 1992.

Prof. Grenier is a registered professional engineer in the Province of Quebec (OIQ), Canada.

Pierre Valin received a B.Sc. honours physics (1972) and a M.Sc. degree (1974) from McGill University, then a Ph.D. in theoretical high energy physics from Harvard University (1980), under the supervision of the 1979 Nobel Laureate Dr. Sheldon Glashow.

He was a faculty lecturer at McGill and Associate Professor of Physics in New Brunswick at Moncton and Fredericton before joining Lockheed Martin Canada in 1993 (then called Paramax), as a Principal Member of R&D. In 2004, he became a defence scientist at Defence R&D Canada (DRDC) at Valcartier, where he currently leads a research group in Future C2 Concepts & Structures. He is thrust leader for Air Command at DRDC since 2007.

He has been particularly active in the International Society of Information Fusion (ISIF) though the organization of FUSION 2001 and 2007. He has been ISIF board member since 2003, VP-membership since 2004, and was president in 2006. He is also an associate editor for the Journal of Advances in Information Fusion (JAIF).

Dr. Valin's interests focus mainly on the following topics: Multi-Sensor Data Fusion (MSDF) requirements and design, C2 systems, algorithmic benchmarking, use of a priori information databases, imagery classifiers (EO/IR and SAR) and their fusion, neural networks, fuzzy logic, information processing and uncertainty representation, reasoning techniques for recognition and identification (Bayes, Dempster-Shafer, Dezert-Smarandache), SAR image processing, Network Centric Warfare, distributed information fusion, dynamic resource management, as well as theoretical and mathematical physics.