Distances between belief functions

Theory and applications

Anne-Laure Jousselme

October, 1st 2015

Purpose of the lecture

1. Provide an overview about distance measures between belief functions
2. Review basic geometry concepts and detail their application to the belief space
3. Discuss some properties and illustrate comparative behaviours
4. Present selected practical uses of distance measures and provide some hints to the criteria for distance choice

"Distance"?

We will discuss here measures which quantify how much Bel_1 and Bel_2 are:

- distant (far from each other)
- different
- dissimilar
- inconsistent
- conflicting
- orthogonal
- discrepant
- in disagreement
- ...

Outline

Preamble

Distances between belief functions

About distances properties

Some applications

Outline

Preamble
- Belief space
- Interaction between sets
- Linear transformations

Distances between belief functions
- Inner product and norm
- Metric distances
- Minkowski family

About distances properties
- Some distances and their properties
- Which distance to use?

Some applications
- Quality assessment
- Decision criterion

Reminder

Belief functions extend both:

1. classical sets
   - A categorical belief function is such that $m(A) = 1$ for some $A \subseteq X$ and defines the classical subset A

2. probabilities
   - A Bayesian belief function is such that $m(A) > 0$ only for $|A| = 1$ (singleton elements) and defines a probability distribution over X

Notations

- X is the frame of discernment of cardinality N:

$$X = \{x_1, \ldots, x_N\}$$

- $\mathcal{P}(X)$ is its power set of cardinality 2^N:

$$\mathcal{P}(X) = \{\varnothing, x_1, \ldots, x_N, (x_1, x_2), \ldots, X\}$$

- x is an element of X: $x \in X$
- A is a subset of X, element of $\mathcal{P}(X)$: $A \subseteq X$, $A \in \mathcal{P}(X)$
- $|A|$ is the cardinality of A
- $\overline{A}$ is the complement of A relatively to X: $\overline{A} = X \backslash A$
- $A \cap B$ denotes the intersection of A and B
- $A \cup B$ denotes the union of A and B

Outline

Preamble
- Belief space
- Interaction between sets
- Linear transformations

Vector space

Vector (or linear) space

A *vector (or linear) space* $\mathcal{V}$ over the field $\mathbb{F}$ is a set that is closed under the two operations of:

- finite vector addition: $\mathbf{v} + \mathbf{u} \in \mathcal{V}$, $\forall (\mathbf{u}, \mathbf{v}) \in \mathcal{V}^2$
- scalar multiplication: $\alpha \mathbf{v} \in \mathcal{V}$, $\forall \alpha \in \mathbb{F}$ and $\forall \mathbf{v} \in \mathcal{V}$

- Elements of $\mathcal{V}$ are called *vectors*
- $\mathbf{v}$ is a column vector: $\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$
- v_i is the i^{th} coordinate of $\mathbf{v}$
- $\mathbf{v}'$ is the transpose vector of $\mathbf{v}$: $\mathbf{v}' = \begin{pmatrix} v_1 \dots v_n \end{pmatrix}$

Remarkable vector spaces (1)

$\mathbb{R}^n$ vector space

The vector space $\mathbb{R}^n$ is the n-dimensional space where $\mathbb{F} = \mathbb{R}$ and

- vectors are represented by a list of n real numbers
- $\{e_1, \ldots, e_n\}$ forms as basis for $\mathbb{R}^n$, where e_i is the i^{th} column of the identity matrix

- vector addition is $\mathbf{v} + \mathbf{u} = \begin{pmatrix} v_1 + u_1 \\ \vdots \\ v_n + u_n \end{pmatrix}$
- scalar multiplication is defined by $\alpha.\mathbf{v} = \begin{pmatrix} \alpha v_1 \\ \vdots \\ \alpha v_n \end{pmatrix}$

Remarkable vector spaces (2)

$\mathcal{P}(\mathcal{X})$ vector space

The power set of X, $\mathcal{P}(\mathcal{X})$, forms a vector space over the two-element field $\{0, 1\}$ with $A, B \in \mathcal{P}(\mathcal{X})$:

- vectors are represented by a list of N binary numbers: $A = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}$
- $\{x_1, \ldots, x_N\}$ is as basis for $\mathcal{P}(\mathcal{X})$, where x_i is the i^{th} element of X

- vector addition is the symmetric difference $A + B = (A \cup B) \backslash (A \cap B)$,
- scalar multiplication is defined by $1.A = A$ and $0.A = \varnothing$

Belief space (1)

Belief space

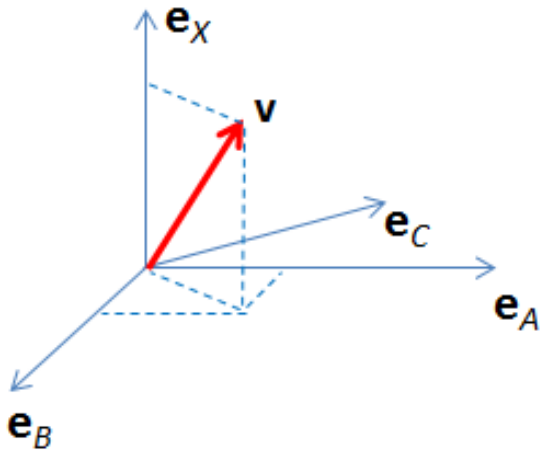
We call *belief space* and denote $\mathcal{E}_X$, the 2^N-dimensional vector space spanned by the set of vectors $\{\mathbf{e}_A, A \subseteq X\}$ over the field $\mathbb{R}$

- $\{\mathbf{e}_A, A \subseteq X\}$ defines a basis for $\mathcal{E}_X$
- $\mathbf{e}_A$ corresponds to a categorical belief function focused on A, $m(A) = 1$ and defines the classical set A

- Any vector $\mathbf{v}$ of $\mathcal{E}_X$ can be then written as:

$$\mathbf{v} = \sum_{A \subseteq X} \alpha_A \mathbf{e}_A$$

- $\alpha_A \in \mathbb{R}$ is the *coordinate* of $\mathbf{v}$ along the dimension $\mathbf{e}_A$



Belief space (2)

Basic Probability Assignment (BPA) vector

The *BPA vector* or *mass vector* $\mathbf{m}$ is the 2^N-dimensional vector whose coordinates $m(A)$ are such that:

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

$$0 \leq m(A) \leq 1$$

$m(\varnothing) = 0$ under the closed-world assumption

We can write

$$\mathbf{m} = \sum_{A \subseteq X} m(A)\mathbf{e}_A$$

Bel Pl q Betp vectors

Belief vector

$$\mathbf{Bel} = \sum_{A \subseteq X} Bel(A)\mathbf{e}_A$$

Plausibility vector

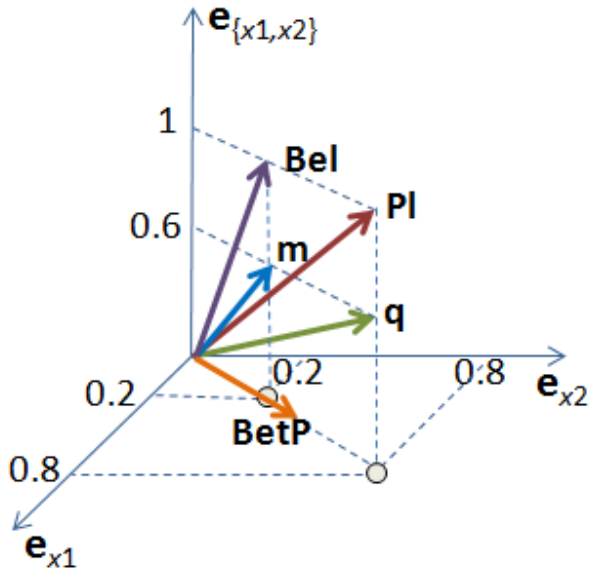
$$\mathbf{Pl} = \sum_{A \subseteq X} Pl(A)\mathbf{e}_A$$

Commonality vector

$$\mathbf{q} = \sum_{A \subseteq X} q(A)\mathbf{e}_A$$

Pignistic vector

$$\mathbf{BetP} = \sum_{A \subseteq X} Betp(A)\mathbf{e}_A$$



Outline

Preamble
- Belief space
- Interaction between sets
- Linear transformations

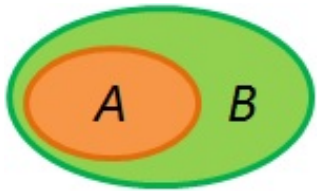
Inclusion

Let us introduce the following indexes between two subsets A and B of X:

Inclusion index

$$Inc(A,B) = \begin{cases} 1 & \text{if } A \subseteq B \\ 0 & \text{otherwise} \end{cases}$$

- Inc is not symmetric



Intersection

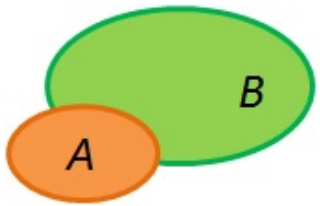
Intersection index

$$Int(A,B) = \begin{cases} 1 & \text{if } A \cap B \neq \varnothing \\ 0 & \text{otherwise} \end{cases}$$

The dual index of Int is $1 - Int$ which is such that:

$$1 - Int(A,B) = \begin{cases} 1 & \text{if } A \cap B = \varnothing \\ 0 & \text{otherwise} \end{cases}$$

- Int and $1 - Int$ are symmetric



Pignistic

Pignistic index

$$Bet(A, B) = \frac{|A \cap B|}{|B|}$$

- *Bet* is not symmetric
- If $|A| = 1$, $Bet(A, B)$ becomes

$$Bet(x, B) = \begin{cases} \frac{1}{|B|} \text{ if } x \in B \\ 0 \text{ else} \end{cases}$$

Jaccard

- Any operation between two subsets A and B can be assigned its corresponding index

Jaccard index

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

- Jac is a similarity measure between sets
- Jac is symmetric, positive
- $Jac(A,B) = 0$ if and only if $A = B$

Other interaction indexes

- $S(A, B) =$ any similarity index between sets
  For instance, Sørensen-Dice:

$$\frac{2|A \cap B|}{|A| + |B|}$$

- $S_p(A, B)$ the specialization matrix
- Fixsen-Mahler:

$$P(A, B) = \frac{\rho(A \cap B)}{\rho(A)\rho(B)}$$

  where $\rho(A) = 1$ if $A \neq \varnothing$ and $\rho(\varnothing) = 0$ is the "trivial" plausibility measure
  ρ can be replaced by p, a probability measure over X

Index writing (1)

$m, Bel, Pl, q, BetP : \mathcal{P}(X) \longrightarrow [0,1]$

Basic Probability Assignment

$$\sum_{A \subseteq X} m(A) = 1 \text{ and } 0 \leq m(A) \leq 1, \forall A \subseteq X$$

$m(\varnothing) = 0$ under the closed-world assumption

Belief

$$Bel(A) = \sum_{B \subseteq A} m(B) \Longleftrightarrow Bel(A) = \sum_{B \subseteq X} m(B) Inc(B, A)$$

Plausibility

$$Pl(A) = \sum_{A \cap B \neq \varnothing} m(B) \Longleftrightarrow Pl(A) = \sum_{B \subseteq X} m(B) Int(A, B)$$

Index writing (2)

Contour function

$$Pl(\{x\}) = \sum_{x \cap B \neq \varnothing} m(B) \iff Pl(\{x\}) = \sum_{B \subseteq X} m(B) Int(x, B)$$

Commonality

$$q(A) = \sum_{A \subseteq B} m(B) \iff q(A) = \sum_{B \subseteq X} m(B) Inc(A, B)$$

Pignistic probability

$$BetP(A) = \sum_{B \subseteq X} m(B) \frac{|A \cap B|}{|B|} \iff BetP(A) = \sum_{B \subseteq X} m(B) Bet(A, B)$$

Index writing (3)

For two BPAs, m_1 and m_2,

Dempster's conflict

$$m_{12}(\varnothing) = \sum_{A \cap B = \varnothing} m_1(A) m_2(B) \iff$$

$$m_{12}(\varnothing) = \sum_{A,B \subseteq X} m_1(A) m_2(B)(1 - Int(A,B))$$

Dempster's agreement is:

$$\sum_{A,B \subseteq X} m_1(A) m_2(B) Int(A,B)$$

Outline

Preamble
- Belief space
- Interaction between sets
- Linear transformations

Linear transformations

A *matrix* $\mathbf{W}$ is a table of real numbers.

Linear transformation

When applied to a vector of $\mathcal{V}$ it defines a *linear transformation* of $\mathbf{v}$:

$$\mathbf{u} = \mathbf{W}.\mathbf{v}$$

- $\mathbf{W}$ is a matrix and $\mathbf{W}'$ and is its transpose
- Eigenvalues λ_i of $\mathbf{W}$ are such that

$$det(\mathbf{W} - \lambda\mathbf{I}) = 0$$

  where $\mathbf{I}$ is the identity matrix
- $\mathbf{W}$ is *positive definite* iff all its eigenvalues are strictly positive ($\lambda_i > 0$)
- $\mathbf{W}$ is *positive semidefinite* iff its eigenvalues are positive ($\lambda_i \geq 0$)

Index matrices (1)

To each index defined previously, we associate the matrix whose elements are the corresponding indexes between two subsets A and B of X, *i.e.*two dimensions of $\mathcal{E}_X$:

$$\begin{array}{c|cccccc} & \{x_1\} & \{x_2\} & \dots & A & B & \dots \\ \{x_1\} & 1 & 0 & \dots & 1 & \frac{1}{2} & \vdots \\ \{x_2\} & 0 & 1 & \dots & \frac{1}{2} & 0 & \vdots \\ \vdots & \dots & \dots & \dots & \dots & \dots & \dots \\ A & \frac{1}{2} & \frac{1}{2} & 0 & 1 & \frac{1}{3} & \vdots \\ B & \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{3} & 1 & \vdots \\ \vdots & \dots & \dots & \frac{1}{2} & \frac{1}{3} & 1 & \vdots \end{array}$$

Index matrices (2)

For example, for $N = 2$ and omitting the $\varnothing$ dimension, we have:

$$\mathbf{Inc} = \begin{array}{c} \\ \{x_1\} \\ \{x_2\} \\ \{x_1, x_2\} \end{array} \begin{array}{c} \begin{array}{ccc} \{x_1\} & \{x_2\} & \{x_1, x_2\} \end{array} \\ \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \end{array}$$

$$\mathbf{Int} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad \mathbf{Bet} = \begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ 1 & 1 & 1 \end{pmatrix} \quad \mathbf{Jac} = \begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 \end{pmatrix} \quad \ldots$$

- Note that $\mathbf{Int}' = \mathbf{Int}$, $\mathbf{Jac}' = \mathbf{Jac}$, ... for the symmetric matrices

Index matrices (3)

Also, we can define rectangular matrices such as:

$$\mathbf{Bet}_x = \begin{array}{c} \\ \{x_1\} \\ \{x_2\} \end{array} \begin{array}{c} \begin{array}{ccc} \{x_1\} & \{x_2\} & \{x_1, x_2\} \end{array} \\ \begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \end{pmatrix} \end{array}$$

$$\mathbf{Int}_x = \begin{array}{c} \\ \{x_1\} \\ \{x_2\} \end{array} \begin{array}{c} \begin{array}{ccc} \{x_1\} & \{x_2\} & \{x_1, x_2\} \end{array} \\ \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \end{array}$$

- They act as projections over $\mathcal{E}_x$, the subspace of $\mathcal{E}_X$ built from the singleton elements

Linear transformations in the belief space

If we take m as the basic function, then the standard other functions can be retrieved by linear transformations of $\mathbf{m}$:

$$\mathbf{Bel} = \mathbf{Inc}'.\mathbf{m}$$

$$\mathbf{Pl} = \mathbf{Int}.\mathbf{m}$$

$$\mathbf{Pl}_x = \mathbf{Int}_x.\mathbf{m}$$

$$\mathbf{q} = \mathbf{Inc}.\mathbf{m}$$

$$\mathbf{BetP} = \mathbf{Bet}.\mathbf{m}$$

$$\mathbf{BetP}_x = \mathbf{Bet}_x.\mathbf{m}$$

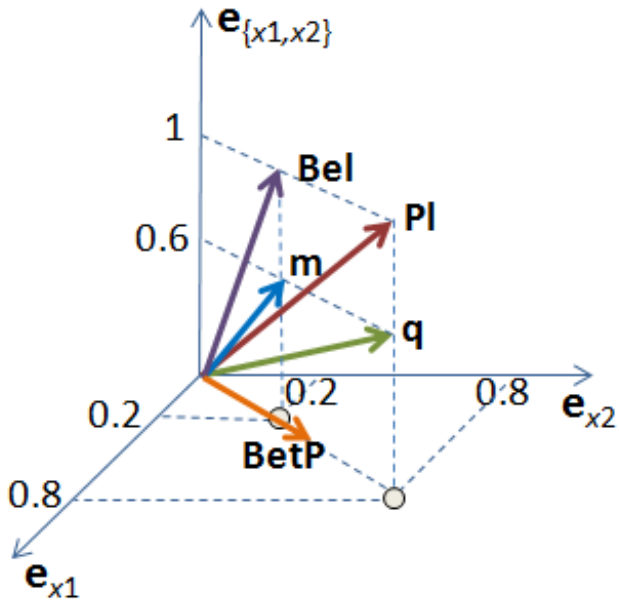
Example

$$X = \{x_1, x_2\}$$

$$\mathbf{m} = \begin{pmatrix} 0.2 \\ 0.2 \\ 0.6 \end{pmatrix}$$

$$\mathbf{Bel} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0.2 \\ 0.2 \\ 0.6 \end{pmatrix} = \begin{pmatrix} 0.2 \\ 0.2 \\ 1 \end{pmatrix}$$

$$\mathbf{Pl} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0.2 \\ 0.2 \\ 0.6 \end{pmatrix} = \begin{pmatrix} 0.8 \\ 0.8 \\ 1 \end{pmatrix}$$



Outline

Preamble
- Belief space
- Interaction between sets
- Linear transformations

Distances between belief functions
- Inner product and norm
- Metric distances
- Minkowski family

About distances properties
- Some distances and their properties
- Which distance to use?

Some applications
- Quality assessment
- Decision criterion

Observations about distances between belief functions

Let $X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$

$m^*(\{x_1, x_2, x_3\}) = 0.8$ $m^*(X) = 0.2$

$m_1(\{x_1, x_2, x_3, x_4\}) = 0.8$ $m_1(X) = 0.2$

$m_2(\{x_5, x_6, x_7\}) = 0.8$ $m_2(X) = 0.2$

- Because $\{x_1, x_2, x_3\} \subset \{x_1, x_2, x_3, x_4\}$ and $\{x_1, x_2, x_3\} \cap \{x_5, x_6, x_6\} = \varnothing$, we expect m^* to be closer to m_1 than to m_2:

$$d(m^*, m_1) < d(m^*, m_2)$$

- However, neither m_1 nor m_2 share any focal element with m^*
- The nature of belief functions requires that the interaction between focal elements is considered in the distance measure

Outline

Distances between belief functions
- Inner product and norm
- Metric distances
- Minkowski family

Inner product definition

Inner product

An *inner product* $\otimes$ over a vector space $\mathcal{V}$ is a mapping

$$\begin{aligned} \otimes : \quad & \mathcal{V} \times \mathcal{V} \longrightarrow \mathbb{R} \\ & (\mathbf{v}_1, \mathbf{v}_2) \longmapsto \otimes(\mathbf{v}_1, \mathbf{v}_2) = \alpha \end{aligned}$$

which must satisfy the 3 following axioms for all vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ of $\mathcal{V}$ and all scalars $a, b \in \mathbb{R}$:

(ip1) Symmetry: $\otimes(\mathbf{v}_1, \mathbf{v}_2) = \otimes(\mathbf{v}_2, \mathbf{v}_1)$,

(ip2) Linearity: $\otimes(a\mathbf{v}_1 + b\mathbf{v}_2, \mathbf{v}_3) = a \otimes (\mathbf{v}_1, \mathbf{v}_3) + b \otimes (\mathbf{v}_2, \mathbf{v}_3)$,

(ip3) Positive-definiteness: $\otimes(\mathbf{v}, \mathbf{v}) \geq 0$ with equality only for $\mathbf{v} = \mathbf{0}$.

- A vector space $\mathcal{V}$ endowed with an inner product $\otimes$ is called an *inner product space*

Inner product in $\mathcal{E}_X$

Inner product belief space

Let us consider the inner product between two BPAs m_1 and m_2 in $\mathcal{E}_X$ of the following general form:

$$\begin{aligned} \otimes_W(\mathbf{m}_1, \mathbf{m}_2) &= \mathbf{m}_1' \mathbf{W} \mathbf{m}_2 \\ &= (\mathbf{U}\mathbf{m}_1)'(\mathbf{U}\mathbf{m}_2) \end{aligned}$$

$\mathcal{E}_X$ endowed with $\otimes_W$ is an inner product space

- $\mathbf{W} = \mathbf{U}'\mathbf{U}$ is a weighting matrix, symmetric and positive definite
- $\mathbf{U}$ is a upper triangle matrix, for instance
- When $\mathbf{W} = \mathbf{I}$, $\otimes_I$ is the standard dot product

Example: Fishing vessel identification

Example

We observe a vessel and we would like to detect if it is a fishing vessel

$$X = \{F, \neg F\}$$

Several sources are at our disposal which after processing give the following BPAs:

$$\text{Radar–speed} \rightarrow \quad \mathbf{m}_1 = \begin{pmatrix} 0.3 \\ 0.1 \\ 0.6 \end{pmatrix} \qquad \text{SAR–length} \rightarrow \quad \mathbf{m}_2 = \begin{pmatrix} 0.2 \\ 0 \\ 0.8 \end{pmatrix}$$

$$\text{AIS–type} \rightarrow \quad \mathbf{m}_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \qquad \text{AIS–length} \rightarrow \quad \mathbf{m}_4 = \begin{pmatrix} 0 \\ 0.7 \\ 0.3 \end{pmatrix}$$

AIS = Automatic Identification System

Example: Dot product (1)

Example

Speed (Radar) and Length (AIS):

$$\mathbf{m}_1 = \begin{pmatrix} 0.3 \\ 0.1 \\ 0.6 \end{pmatrix} \quad \mathbf{m}_4 = \begin{pmatrix} 0 \\ 0.7 \\ 0.3 \end{pmatrix}$$

Whenever $\mathbf{W} = \mathbf{I}$ is the identity matrix, we have:

$$\begin{aligned} \otimes_I(\mathbf{m}_1, \mathbf{m}_4) &= \mathbf{m}_1' \mathbf{m}_4 \\ &= \begin{pmatrix} 0.3 & 0.1 & 0.6 \end{pmatrix} . \begin{pmatrix} 0 \\ 0.7 \\ 0.3 \end{pmatrix} \\ &= 0.3 \times 0 + 0.1 \times 0.7 + 0.6 \times 0.3 \\ &= 0 + 0.07 + 0.18 \\ &= 0.25 \end{aligned}$$

Example: Plausibility (1)

Example

Whenever $\mathbf{U} = \mathbf{Int}$, we have:

$$\begin{aligned}
\otimes_{Int}(\mathbf{m}_1, \mathbf{m}_4) &= \mathbf{Pl}_1' \mathbf{Pl}_4 \\
&= (\mathbf{Intm}_1)' \mathbf{Intm}_4 \\
&= \mathbf{m}_1' \mathbf{Int}' \mathbf{Intm}_4 \\
&= \begin{pmatrix} 0.3 & 0.1 & 0.6 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0.7 \\ 0.3 \end{pmatrix} \\
&= 1.97
\end{aligned}$$

- $\otimes_I(\mathbf{m}_1, \mathbf{m}_4) < \otimes(\mathbf{Pl}_1, \mathbf{Pl}_4)$

Orthogonality

Orthogonality

Two vectors of $\mathcal{E}_X$ (*i.e.*two belief functions) are *orthogonal*, and we note $\mathbf{m}_1 \perp_W \mathbf{m}_2$, iff their inner product is null:

$$\mathbf{m}_1 \perp_W \mathbf{m}_2 \Leftrightarrow \otimes_W(\mathbf{m}_1, \mathbf{m}_2) = 0$$

- This is way to quantify how much $\mathbf{m}_1$ differs from $\mathbf{m}_2$
- $\otimes_W(\mathbf{m}_1, \mathbf{m}_2)$ quantifies a notion of "agreement" : The higher $\otimes_W(\mathbf{m}_1, \mathbf{m}_2)$, the more in agreement $\mathbf{m}_1$ and $\mathbf{m}_2$
- $f\left(\otimes_W(\mathbf{m}_1, \mathbf{m}_2)\right)$ quantifies a notion of "disagreement" for any decreasing function f

Dot product (2)

Example

Length (SAR) and type (AIS):

$$\mathbf{m}_2 = \begin{pmatrix} 0.2 \\ 0 \\ 0.8 \end{pmatrix} \quad \mathbf{m}_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$\begin{aligned} \otimes_I(\mathbf{m}_2, \mathbf{m}_3) &= \mathbf{m}_2' \mathbf{m}_3 \\ &= \begin{pmatrix} 0.2 & 0 & 0.8 \end{pmatrix} . \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = 0 \end{aligned}$$

- Fully orthogonal because no positive mass on the same focal element
- However, they agree on $\neg F$ so m_2 and m_3 should not be fully orthogonal

Plausibility (2)

Example

Whenever $\mathbf{U} = \mathbf{Int}$, we have:

$$
\begin{aligned}
\otimes_{Int2}(\mathbf{m}_2, \mathbf{m}_3) &= \mathbf{Pl}_2' \mathbf{Pl}_3 \\
&= \begin{pmatrix} 0.2 & 0 & 0.8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0.8 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \\
&= 0 + 0.8 + 1 \\
&= 1.8
\end{aligned}
$$

- The plausibility inner product is not null expressing that m_2 and m_3 are not fully orthogonal

Examples of inner products in the belief space

$$\otimes_W(\mathbf{m}_1, \mathbf{m}_2) = \mathbf{m}_1' \mathbf{W} \mathbf{m}_2$$

Notation	$\mathbf{W}$	Def.
$\otimes^s_I$	$\mathbf{I}$	$I(A,B) = 1$ iff $A = B$
$\otimes^s_{Inc}$	$\mathbf{Inc}\mathbf{Inc}'$	$Inc(A,B) = 1$ iff $A \subseteq B$
$\otimes^s_{Int2}$	$\mathbf{Int}'\mathbf{Int}$	$Int(A,B) = 1$ iff $A \cap B \neq \varnothing$
$\otimes^s_{Intx}$	$\mathbf{Int}_x'\mathbf{Int}_x$	$Int_x(x,B) = 1$ iff $x \in B$
$\otimes^s_{Bet}$	$\mathbf{Bet}'\mathbf{Bet}$	$Bet(A,B) = \frac{\lvert A \cap B \rvert}{\lvert B \rvert}$
$\otimes^s_{Betx}$	$\mathbf{Bet}_x'\mathbf{Bet}_x$	$Bet_x(x,A) = \frac{\lvert x \cap A \rvert}{\lvert A \rvert}$
$\otimes^s_P$	$\mathbf{P}$	$P(A,B) = \frac{p(A \cap B)}{p(A)p(B)}$
$\otimes^s_J$	$\mathbf{Jac}$	$J(A,B) = \frac{\lvert A \cap B \rvert}{\lvert A \cup B \rvert}$
$\otimes^s_S$	$\mathbf{S}$	$S(A,B)$ any similarity measure
$\otimes^s_{F(S)}$	$F(\mathbf{S}, R)$	F reward-penalty function

Meaning of W: Interaction between focal elements

The matrix $\mathbf{W}$ weights the basic coordinates of $\mathbf{m}_1$ and $\mathbf{m}_2$ according to the interaction of pairs of their focal elements

- For $\mathbf{W} = \mathbf{I}$, no interaction is considered
- For $\mathbf{W} = \mathbf{InclInc}'$ and $\mathbf{W} = \mathbf{Int}'\mathbf{Int}$, a binary interaction is considered
  - $\mathbf{W} = \mathbf{InclInc}'$ describes an inclusion interaction (stronger)
  - $\mathbf{W} = \mathbf{Int}'\mathbf{Int}$ describes an intersection interaction (weaker)
- ...

Dempster's conflict (1)

Dempster's conflict can be put under the form of an inner product as:

$$\otimes^d_{Int}(m_1, m_2) = \mathbf{m}'_1(1 - \mathbf{Int})\,\mathbf{m}_2$$

- **Int** is the matrix of intersection indexes
- $\otimes^d_{Int}(m_1, m_2)$ should thus not be called an inner product because $(1 - \mathbf{Int})$ is neither positive nor definite

We can also write:

$$\begin{aligned}\otimes^d_{Int}(m_1, m_2) &= 1 - \mathbf{m}'_1\mathbf{Int}\,\mathbf{m}_2 \\ &= 1 - \otimes^s_{Int}(m_1, m_2)\end{aligned}$$

where $\otimes^s_{Int}(m_1, m_2)$ is Dempster's agreement

Dempster's agreement

Example

Length (SAR) and type (AIS): $\mathbf{m}_2 = \begin{pmatrix} 0.2 \\ 0 \\ 0.8 \end{pmatrix}$ $\mathbf{m}_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$

$$\begin{aligned} \otimes^{s}_{Int}(\mathbf{m}_2, \mathbf{m}_3) &= \mathbf{m}_2' \mathbf{Int} \mathbf{m}_3 \\ &= \begin{pmatrix} 0.2 & 0 & 0.8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0.8 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = 0.8 \end{aligned}$$

- The agreement is not maximum (or the conflict is not null) because $m_2(F)$ and $m_3(\neg F)$ are not null AND $Int(F, \neg F) = 0$

Dempster's conflict (2)

- **Int** defines a similarity measure over $\mathcal{P}(X)$
- $1 -$ **Int** defines a dissimilarity measure over $\mathcal{P}(X)$
- Unfortunately, this does not imply that $\otimes^d_{Int}$ is a dissimilarity measure in $\mathcal{E}_X$
- In particular, $\otimes^d_{Int}(\mathbf{m}, \mathbf{m}) = 0$ is not satisfied (reflexivity property)
- Any increasing function of $\otimes^d_{Int}$ can be used to define a "distance". For instance [Ristic, Smets 2006]:

$$d_{RS}(m_1, m_2) = -\log(1 - \otimes^d_{Int}(m_1, m_2))$$

Norm

The norm of $\mathbf{v}$ measures the length of $\mathbf{v}$

Norm

Given a vector space $\mathcal{V}$ over a field $\mathbb{F}$, a norm on $\mathcal{V}$ is a function $\|.\| : \mathcal{V} \to \mathbb{R}$ with the following properties, $\forall \alpha \in \mathbb{F}$ and $\forall \mathbf{v}, \mathbf{u} \in \mathcal{V}$:

(n1) Positivity: $\|\mathbf{v}\| \geq 0$ and $\|\mathbf{v}\| = 0$ iff $\mathbf{v} = \mathbf{0}$

(n2) Homogeneity: $\|\alpha\mathbf{v}\| = |\alpha|.\|\mathbf{v}\|$

(n3) Subadditivity: $\|\mathbf{v} + \mathbf{u}\| \leq \|\mathbf{v}\| + \|\mathbf{u}\|$

Norm induced by an inner product

Norm induced by an inner product

An inner product $\otimes_W$ over $\mathcal{V}$ induces a norm $\|\mathbf{v}\|_W$ over $\mathcal{V}$ defined as:

$$\|\mathbf{v}\|_W = \sqrt{\otimes_W(\mathbf{v}, \mathbf{v})}$$

- Not every norm arises from an inner product
- Only the norms satisfying the parallelogram law
- For instance, L_2-norm is induced by an inner product while L_1-norm is not

Cosine family

The cosine measure defines a normalized inner product measure of similarity between belief functions

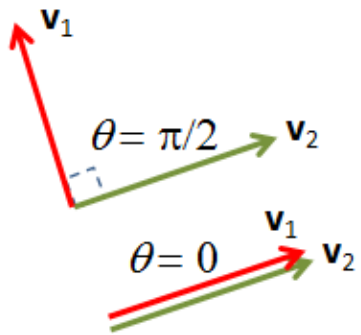
Cosine between belief functions

The general formulation for a cosine measure in $\mathcal{E}_X$ is:

$$\cos_W(m_1, m_2) = \frac{\mathbf{m}_1' \mathbf{W}\, \mathbf{m}_2}{\|\mathbf{m}_1\|_W \cdot \|\mathbf{m}_2\|_W}$$

- Because $m(A) \geq 0$ for all $A \subseteq X$, for all $(m_1, m_2) \in \mathcal{E}_X$, we have $0 \leq \theta \leq \frac{\pi}{2} \Rightarrow 0 \leq \cos_W(\theta) \leq 1$ with:
  - $\cos_W(m_1, m_2) = 0$ iff $\mathbf{m}_1 \perp_W \mathbf{m}_2$ (orthogonal)
  - $\cos_W(m_1, m_2) = 1$ iff $\mathbf{m}_1 = \alpha \mathbf{m}_2$ (collinear)
- A dissimilarity measure in $\mathcal{E}_X$ can be obtained by:

$$\cos_W^d(m_1, m_2) = 1 - \cos_W(m_1, m_2)$$



Cosine plausibility

Example

Length (SAR) and type (AIS):

$$\mathbf{m}_2 = \begin{pmatrix} 0.2 \\ 0 \\ 0.8 \end{pmatrix} \quad \mathbf{m}_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

We had: $\otimes_I(Pl_2, Pl_3) = 1.8$. The cosine measure is:

$$\begin{aligned} \cos_{Int2}(m_2, m_3) &= \frac{\mathbf{Pl}_2' \mathbf{Pl}_3}{\|\mathbf{Pl}_2\|.\|\mathbf{Pl}_3\|} \\ &= \frac{1.8}{\sqrt{2.64 \times 2}} \\ &= 0.7833 \end{aligned}$$

- We obtain a normalized value

Angle

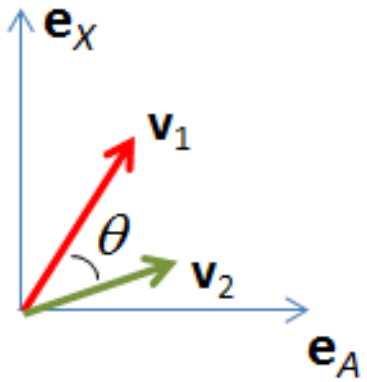
The inner product allows define the angle between two vectors

Angle

The *angle* between $\mathbf{m}_1$ and $\mathbf{m}_2$ is given by:

$$\theta = \arccos\left(\frac{\otimes_W(\mathbf{m}_1, \mathbf{m}_2)}{\|\mathbf{m}_1\|_W . \|\mathbf{m}_2\|_W}\right)$$

where $\|\mathbf{m}\|_W$ is the norm of $\mathbf{v}$



The angle exists and is unique

Outline

Distances between belief functions
- Inner product and norm
- Metric distances
- Minkowski family

Metric distance definition

Metric

A function $d : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ is a *(full) metric* if and only if d satisfies the following properties for all $(y, z, t) \in \mathcal{V}^3$:

(d1) Positivity: $d(y, z) \geq 0$,

(d2) Symmetry: $d(y, z) = d(z, y)$,

(d3) Definiteness: $d(y, z) = 0 \Leftrightarrow y = z$,

- (d3)' Reflexivity (or identity): $d(y, y) = 0$
- (d3)" Separability: $d(y, z) = 0 \Rightarrow y = z$,

(d4) Triangle inequality: $d(y, z) \leq d(y, t) + d(z, t)$

(d1) together with (d3) define positive definiteness

Metric properties

			Metric	Semi-metric	Quasi-metric	Pseudo-metric	Semi-pseudo-metric	Pre-metric
(d1)	**Positivity**	$d(y,z) \geq 0$	×	×	×	×	×	×
(d2)	**Symmetry**	$d(y,z) = d(z,y)$	×	×		×	×	
(d3)	**Definiteness**	$d(y,z) = 0 \Leftrightarrow y = z$	×	×	×			
(d3)'	Reflexivity	$d(y,y) = 0$	×	×	×	×	×	×
(d3)"	Separability	$d(y,z) = 0 \Rightarrow y = z$	×	×	×			
(d4)	**Triangle inequ.**	$d(y,z) \leq d(y,t) + d(t,z)$	×		×	×		

Distance induced by a norm

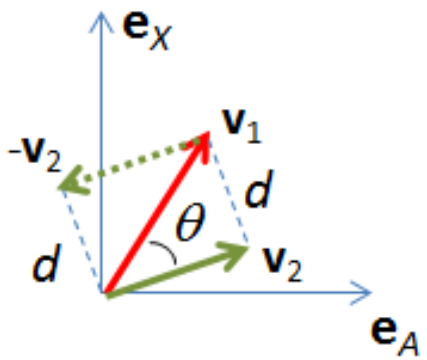
Given a normed space $(\mathcal{V}, \|.\|)$ the (metric) distance between $\mathbf{v}_1$ and $\mathbf{v}_2$ can be defined as the norm of their difference

Distance

A norm $\|.\|_W$ defined over $\mathcal{E}_X$ induces a *distance* on $\mathcal{E}_X$ by:

$$d_W(\mathbf{m}_1, \mathbf{m}_2) = \|\mathbf{m}_1 - \mathbf{m}_2\|_W$$

- A norm may exist without inner product



Outline

Distances between belief functions
- Inner product and norm
- Metric distances
- Minkowski family

Minkowski family

The Minkowski L_p norm is defined as

$$\|\mathbf{m}\|_W^{(p)} = \left(\left[(\mathbf{Um})^{\frac{p}{2}} \right]' \left[(\mathbf{Um})^{\frac{p}{2}} \right] \right)^{\frac{1}{p}}$$

L_p family

The Minkowski (or L_p) family of distances between two belief functions induced by Minkowski norm can be written under the following general form:

$$d_W^{(p)}(m_1, m_2) = \left(\left[(\mathbf{Um}_1 - \mathbf{Um}_2)^{\frac{p}{2}} \right]' \left[(\mathbf{Um}_1 - \mathbf{Um}_2)^{\frac{p}{2}} \right] \right)^{\frac{1}{p}}$$

- $\mathbf{U}$ is the upper triangular matrix such that $\mathbf{W} = \mathbf{U}'\mathbf{U}$
- p is an integer higher than 1
- A normalization constant should be added . . . that will be omitted in the following

Typical cases of interest are obtained with

- $p = 1$: Manhattan, city-block, taxicab
- $p = 2$: Euclidean
- $p = \infty$: Chebyshev

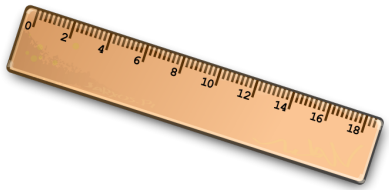
L_p interpretation

L_1 distance
Taxicab norm or Manhattan distance



The distance a car would drive in a city laid out in square blocks

L_2 distance
Euclidean distance



The distance obtained if measured with a ruler: the "intuitive" idea of distance

L_∞ distance
Chebyshev distance



The minimum number of moves kings require to travel between two squares on a chessboard

Ref. Wikipedia

L_2 - **Euclidean distances**

L_2 family

When $p = 2$, $d_W^{(p)}(m_1, m_2)$ becomes:

$$\begin{aligned} d_W^{(2)}(m_1, m_2) &= \sqrt{[\mathbf{U}(\mathbf{m}_1 - \mathbf{m}_2)]'[\mathbf{U}(\mathbf{m}_1 - \mathbf{m}_2)]} \\ &= \sqrt{(\mathbf{m}_1 - \mathbf{m}_2)'\mathbf{W}(\mathbf{m}_1 - \mathbf{m}_2)} \end{aligned}$$

where $\mathbf{W} = \mathbf{U}'\mathbf{U}$ is a positive semidefinite matrix.

- $d_W^{(2)}(m_1, m_2)$ is induced by the inner product $\otimes_W(m_1, m_2) = \mathbf{m}_1'\mathbf{W}\mathbf{m}_2$
- The most "intuitive" notion of distance: Length of a straight line between two points
- Generalization of Pythagore theorem



Classical L_2 distances in $\mathcal{E}_X$ (1)

- Euclidean distance between BPAs: $\mathbf{W} = \mathbf{I}$

$$d_I^{(2)}(m_1, m_2) = \sqrt{(\mathbf{m}_1 - \mathbf{m}_2)'(\mathbf{m}_1 - \mathbf{m}_2)}$$

- Euclidean distance between belief functions: $\mathbf{W} = \mathbf{InclInc}' = \mathbf{Inc}'\mathbf{2}$

$$d_{Inc'2}^{(2)}(m_1, m_2) = \sqrt{(\mathbf{Bel}_1 - \mathbf{Bel}_2)'(\mathbf{Bel}_1 - \mathbf{Bel}_2)}$$

- Euclidean distance between plausibility functions: $\mathbf{W} = \mathbf{Int}'\mathbf{Int} = \mathbf{Int2}$

$$d_{Int2}^{(2)}(m_1, m_2) = \sqrt{(\mathbf{Pl}_1 - \mathbf{Pl}_2)'(\mathbf{Pl}_1 - \mathbf{Pl}_2)}$$

- Because $Pl(A) = 1 - Bel(\overline{A})$, we have

$$d_{Int2}^{(2)}(m_1, m_2) = d_{Inc'2}^{(2)}(m_1, m_2)$$

- $d_I^{(2)}$, $d_{Int2}^{(2)}$ and obviously $d_{Inc'2}^{(2)}$ are full metric distances

Classical L_2 distances in $\mathcal{E}_X$ (2)

- Euclidean distance between pignistic probabilities:
  $\mathbf{W} = \mathbf{Bet}'\mathbf{Bet} = \mathbf{Bet2}$

$$d^{(2)}_{Bet2}(m_1, m_2) = \sqrt{(\mathbf{BetP}_1 - \mathbf{BetP}_2)'(\mathbf{BetP}_1 - \mathbf{BetP}_2)}$$

- Euclidean distance between pignistic probabilities of singletons:
  $\mathbf{W} = \mathbf{Bet}'_x\mathbf{Bet}_x = \mathbf{Betx2}$

$$d^{(2)}_{Betx2}(m_1, m_2) = \sqrt{(\mathbf{BetP}_{x1} - \mathbf{BetP}_{x2})'(\mathbf{BeP}_{x1} - \mathbf{BetP}_{x2})}$$

- $d^{(2)}_{Bet2}$ and $d^{(2)}_{Betx2}$ are pseudo-metric distances (separability non-respected)
- Other functions of **Um** can be considered as well (see [Elouedi etal, 2001])

Jaccard L_2 Distance

Jaccard L_2 distance

When $\mathbf{W} = \mathbf{Jac}$:

$$d_J^{(2)}(m_1, m_2) = \sqrt{(\mathbf{m}_1 - \mathbf{m}_2)' \mathbf{Jac} (\mathbf{m}_1 - \mathbf{m}_2)}$$

where

$$\mathbf{Jac} = \frac{|A \cap B|}{|A \cup B|}$$

- **Jac** quantifies the similarity between pairs of focal elements of m_1 and m_2
- **Jac** is positive definite
- $d_J^{(2)}$ is a full metric, guarantying that $d_J^{(2)}(m_1, m_2) = 0 \Rightarrow m_1 = m_2$

Example: L_2 distances

Example

Speed (Radar) and Length (SAR)

$$\mathbf{m}_1 = \begin{pmatrix} 0.3 \\ 0.1 \\ 0.6 \end{pmatrix} \quad \mathbf{m}_2 = \begin{pmatrix} 0.2 \\ 0 \\ 0.8 \end{pmatrix}$$

We have then:

$$d_I^{(2)}(m_1, m_2) = 0.25$$
$$d_J^{(2)}(m_1, m_2) = 0.14$$

- We observe that $d_J < d_I$

Extensions of $d_J^{(2)}$

1. Any similarity between sets and reward function [Diaz etal, 2006]

$$d^{(2)}_{F(J)}(m_1, m_2) = \sqrt{(\mathbf{m}_1 - \mathbf{m}_2)' F(\mathbf{S}, R)(\mathbf{m}_1 - \mathbf{m}_2)}$$

2. Ordered sets using Hausdorff distance [Sunberg & Rogers, 2013]

$$d^{(2)}_{Haus}(m_1, m_2) = \sqrt{(\mathbf{m}_1 - \mathbf{m}_2)' \mathbf{D}_H (\mathbf{m}_1 - \mathbf{m}_2)}$$

3. Continuous belief functions using intervals similarity [Attiaoui etal, 2013]

$$d^{(2)}_{\delta}(m_1, m_2) = \sqrt{\|f_1\|^2 + \|f_2\|^2 - 2 \otimes (f_1, f_2)}$$

Manhattan distances (1)

L_1 family

When $p = 1$, $d_W^{(p)}(m_1, m_2)$ becomes:

$$d_W^{(1)}(m_1, m_2) = \left[(\mathbf{U}\mathbf{m}_1 - \mathbf{U}\mathbf{m}_2)^{\frac{1}{2}}\right]' \left[(\mathbf{U}\mathbf{m}_1 - \mathbf{U}\mathbf{m}_2)^{\frac{1}{2}}\right]$$

- Manhattan distance is induced by the L_1 norm:

$$\|\mathbf{m}\|_W^{(1)} = \left[(\mathbf{U}\mathbf{m})^{\frac{1}{2}}\right]' \left[(\mathbf{U}\mathbf{m})^{\frac{1}{2}}\right]$$

- The L_1 norm is not induced by an inner product



Classical Manhattan distances in $\mathcal{E}_X$

- Manhattan distance between BPAs:

$$d_I^{(1)}(m_1, m_2) = \sum_{A \subseteq X} |m_1(A) - m_2(A)|$$

- Manhattan distance between belief functions

$$d_{Inc'}^{(1)}(m_1, m_2) = \sum_{A \subseteq X} |Bel_1(A) - Bel_2(A)| = d_{Int}^{(1)}(m_1, m_2)$$

- Manhattan distance between contour functions

$$d_{Intx}^{(1)}(m_1, m_2) = \sum_{x \in X} |Pl_1(x) - Pl_2(x)|$$

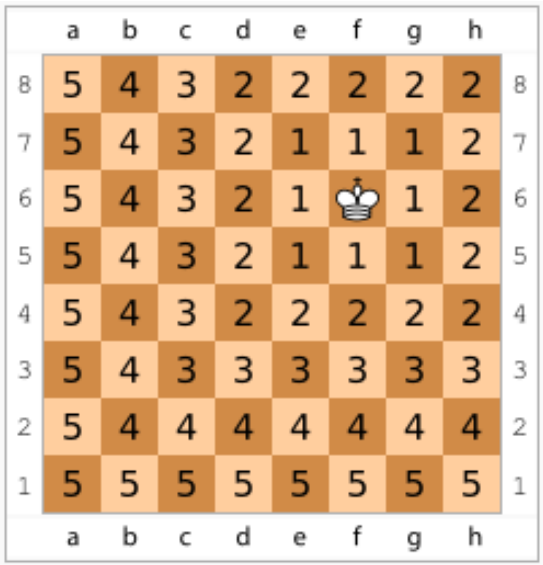
Chebyshev distances

L_∞ family

When $p = \infty$, $d_W^{(p)}(m_1, m_2)$ becomes:

$$d_W^{(\infty)}(m_1, m_2) = \max_{A \subseteq X} \{|(\mathbf{U}\,\mathbf{m}_1)'\mathbf{e}_A - (\mathbf{U}\,\mathbf{m}_2)'\,\mathbf{e}_A|\}$$

- Different Chebyshev distances are obtained by changing the weighting matrix $\mathbf{U}$



Classical Chebyshev distances in $\mathcal{E}_X$

- Chebyshev distance between pignistic probabilities:

$$d_{Bet}^{(\infty)}(m_1, m_2) = \max_{A \subseteq X} \{|(\mathbf{Bet}\ \mathbf{m}_1)'\ \mathbf{e}_A - (\mathbf{Bet}\ \mathbf{m}_2)'\ \mathbf{e}_A|\}$$

- Chebyshev distance between plausibility functions:

$$d_{Int}^{(\infty)}(m_1, m_2) = \max_{A \subseteq X} \left\{|\mathbf{Pl}_1'\ \mathbf{e}_A - \mathbf{Pl}_2'\ \mathbf{e}_A|\right\}$$

- Chebyshev distance between the contour functions:

$$d_{Intx}^{(\infty)}(m_1, m_2) = \max_{x \in X} \left\{|\mathbf{Pl}_1'\ \mathbf{e}_x - \mathbf{Pl}_2'\ \mathbf{e}_x|\right\}$$

Other families of distance

Helinger distance

$$d_W^{(H)}(m_1, m_2) = \left(1 - \otimes_W^{\frac{1}{2}}(m_1, m_2)\right)^{\frac{1}{2}}$$

- Helinger distance is a full metric

Information-based distances

$$d_U(m_1, m_2) = |U(m_1) - U(m_2)|$$

where U is any uncertainty measure defined for belief functions

For instance [Denoeux, 2001], for $GC(m) = \sum_{A \subseteq X} m(A)|A|$, the generalized cardinality of a belief function:

$$d_{GC}(m_1, m_2) = |(\mathbf{m}_1 - \mathbf{m}_2)' \mathbf{c}_A|$$

where $\mathbf{c}_A$ is the column vector of cardinality

Outline

Preamble
- Belief space
- Interaction between sets
- Linear transformations

Distances between belief functions
- Inner product and norm
- Metric distances
- Minkowski family

About distances properties
- Some distances and their properties
- Which distance to use?

Some applications
- Quality assessment
- Decision criterion

Outline

About distances properties
Some distances and their properties
Which distance to use?

Metric properties

			Metric	Semi-metric	Quasi-metric	Pseudo-metric	Semi-pseudo-metric	Pre-metric
(d1)	**Positivity**	$d(y,z) \geq 0$	×	×	×	×	×	×
(d2)	**Symmetry**	$d(y,z) = d(z,y)$	×	×		×	×	
(d3)	**Definiteness**	$d(y,z) = 0 \Leftrightarrow y = z$	×	×	×			
(d3)'	Reflexivity	$d(y,y) = 0$	×	×	×	×	×	×
(d3)"	Separability	$d(y,z) = 0 \Rightarrow y = z$	×	×	×			
(d4)	**Triangle inequ.**	$d(y,z) \leq d(y,t) + d(t,z)$	×		×	×		

Reflexivity in $\mathcal{E}_X$

Reflexivity

$$d(m, m) = 0$$

means that we expect that the "distance" from a given belief function to itself is null

- This property is not satisfied by Dempster's conflict since we may have some m such that

$$\otimes^d_{Int}(m, m) \neq 0$$

For instance if $m' = \begin{pmatrix} 0.2 & 0.2 & 0.6 \end{pmatrix}$, $\otimes^d_{Int}(m, m) = 0.08$

Separability in $\mathcal{E}_X$

Separability

$$d(m_1, m_2) = 0 \Rightarrow m_1 = m_2$$

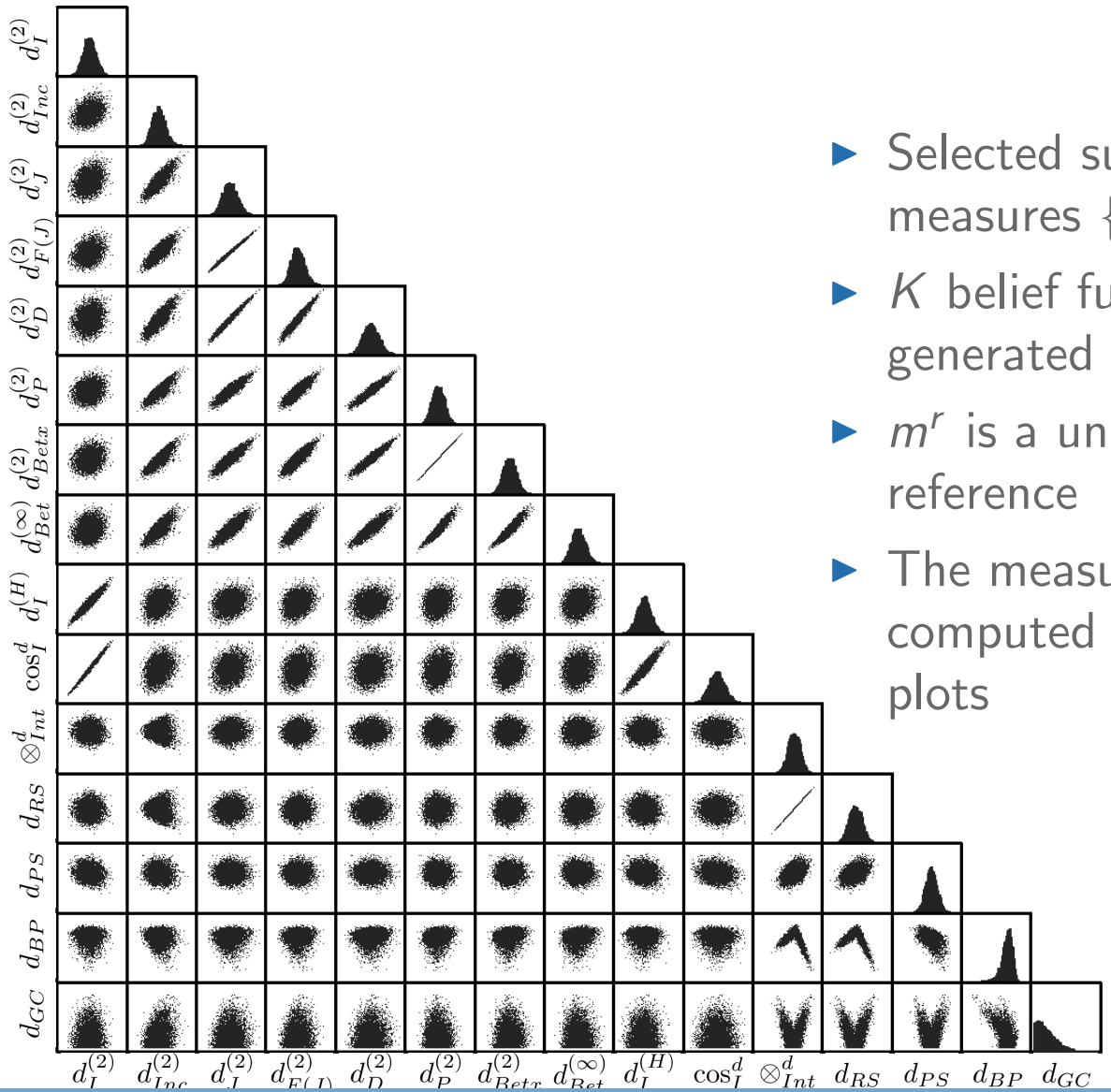
means that we expect that a null "distance" between two belief functions implies that these functions are equal

- This property is satisfied by (full) metric measures
- This property is not satisfied by pseudo-metric measures

Some distances and their properties

Distance	Metric (d1)	(d2)	(d3)'	(d3)"	(d4)	Set interaction (si)
$d_I^{(H)}$	×	×	×	×	×	
$d_I^{(2)}$	×	×	×	×	×	
$d_{Inc}^{(2)}$	×	×	×	×	×	×
$d_J^{(2)}$	×	×	×	×	×	×
$d_{Betx}^{(2)}$	×	×	×		×	×
$d_{Bet}^{(\infty)}$	×	×	×		×	×
$d_P^{(2)}$	×	×	×		×	×
d_{GC}	×	×	×			×
$\cos_I^d$	×	×	×			
$\otimes_{Int}^d$	×	×				×
d_{RS}	×	×				×

Comparative behaviour



- Selected subset of distance measures $\{d_i\}_{i=1}^{15}$
- K belief functions are randomly generated $\{Bel^k\}_{k=1}^{K} \equiv \{m^k\}_{k=1}^{K}$
- m^r is a unique belief function of reference
- The measures $d_i(m^r, m^k)$ are computed and displayed as scatter plots

Two-dimensional measures

Observing that some measures are lowly correlated it is justified to consider pairs of measures to quantify the discrepancy between belief functions

Example of a 2D measure [Liu, 2006]

A two-dimensional "distance" measure between belief functions may have the following form:

$$d_L^{2D} = \left(\otimes_{Int}^{d}(m_1, m_2); d_{Bet}^{(\infty)}(m_1, m_2) \right)$$

where

- $\otimes_{Int}^{d}(m_1, m_2)$ is Dempster's conflict, and
- $d_{Bet}^{(\infty)}(m_1, m_2)$ is the L_∞ distance between pignistic probabilities

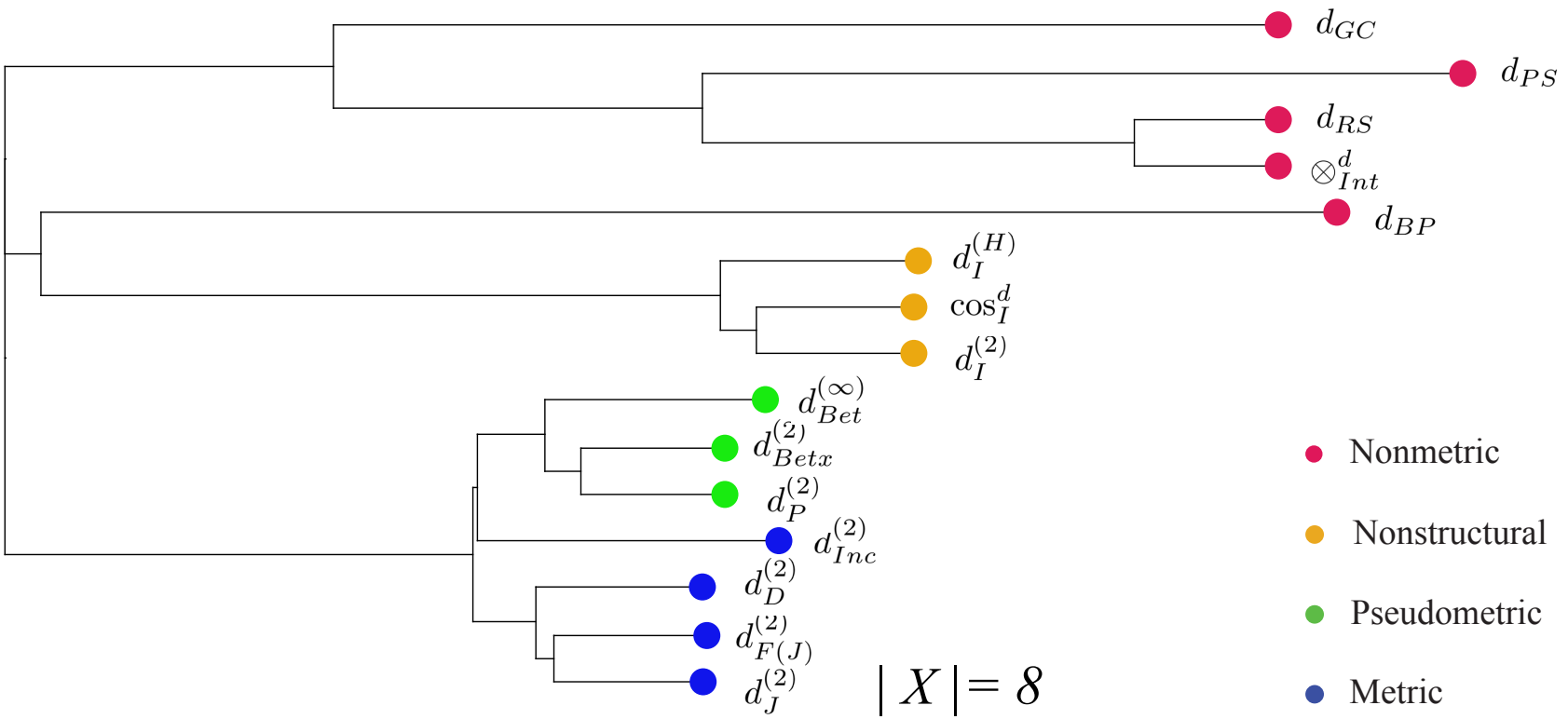
More generally, we can consider measures of the form:

$$d_{(W,V)}^{2D}(m_1, m_2) = \left(\otimes_{W}^{d}(m_1, m_2); d_V(m_1, m_2) \right)$$

Additive tree of some measures (1)

- For each pair of measures (d_i, d_j), we computed a correlation coefficient (Pearson), and built the corresponding matrix whose elements are $c(d_i, d_j)$
- If $c(d_i, d_j) = 0$, the two distances are uncorrelated while the distances are all the more correlated when $c(d_i, d_j)$ is close to 1 (or -1)
- We built the additive trees (dendrograms) based on distances correlation coefficients

Additive tree of some measure (2)



- For a higher cardinality of X, the measures naturally group by types

It seems than similar behaviours depend on (1) the metric properties and (2) the weighting matrices

Outline

About distances properties
Some distances and their properties
Which distance to use?

Which distance to use?

To select the proper measure, we have the following degrees of freedom:

- The desirable metric properties (separability, reflexivity, etc)
- The interaction between focal elements (weaker such as **Int**, **Inc**, or stronger such as **Jac**) and their meaning
- The meaning of the measure:
  - value of p in Minkowski family
  - angle versus distance, conflict versus distance, ...

The choice should be guided by the purpose of the measure within application at hand. For instance:

- Quality assessment vs decision/optimization criterion
- Nature of the quality dimension measured (accuracy, credibility, etc)
- Link/consistency with the decision rule if any
- Consistency properties with classical sets and probabilities measures
- Other properties that remain to be defined and studied!

Outline

Preamble
- Belief space
- Interaction between sets
- Linear transformations

Distances between belief functions
- Inner product and norm
- Metric distances
- Minkowski family

About distances properties
- Some distances and their properties
- Which distance to use?

Some applications
- Quality assessment
- Decision criterion

Applications making use of "distance" measures

1. Quality assessment
   - 1.1 Accuracy
   - 1.2 Credibility
   - 1.3 Information loss

   These quality measures are either used as final measures of performance or reintroduced in the algorithm for further processing

2. Decision criterion
   - 2.1 Classification/identification solution
   - 2.2 Belief function approximation
   - 2.3 Evidential pattern matching
   - 2.4 Evidential information retrieval
   - 2.5 ...

Outline

Some applications
Quality assessment
Decision criterion

Accuracy assessment

- Accuracy is used here to refer to a distance to a reference BPA , denoted as m^*
- m^* can either represent the "truth" or any target BPA

Accuracy assessment with distance

The accuracy of the algorithm Alg with output m relatively to a reference m^* is assessed as

$$\text{Acc}(\text{Alg}) = d(m, m^*)$$

where d is a distance measure between belief functions and m^* is a categorical BPA

- m^* does not contain any uncertainty, but may contain some imprecision (focused on A rather than on x)

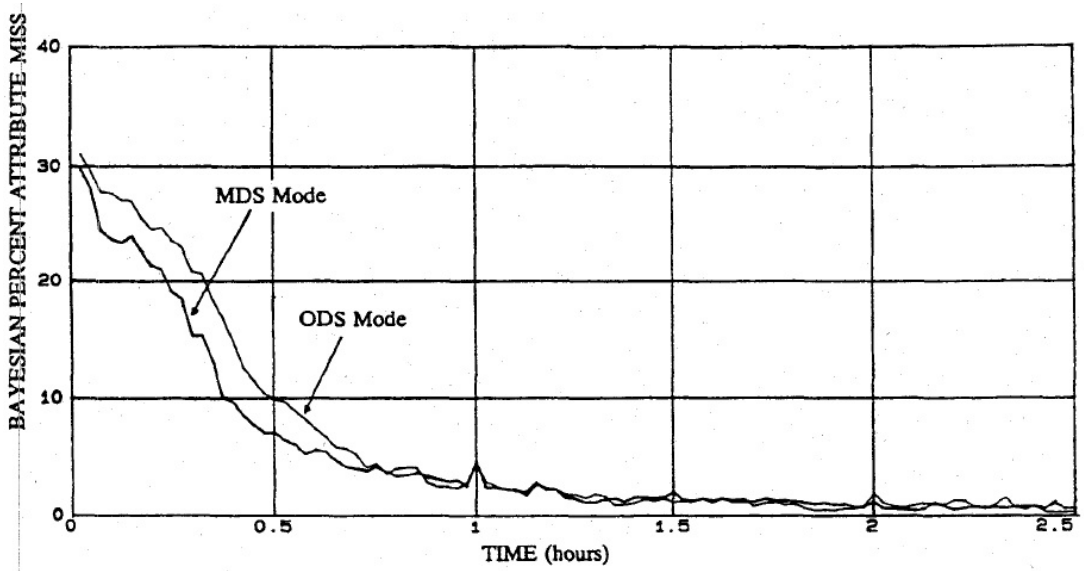
Accuracy assessment - Classification with two combination models

See [Fixsen & Mahler, 2006]

$$\mathrm{Acc}(\mathrm{Alg}_t^{(k)}) = d_P^{(2)}(m_t^{(k)}, m_x^*)$$

with

- $\mathrm{Alg}^{(k)}$ is a fusion algorithm
- Ordinary ($k = 1$) vs modified ($k = 2$) Dempster's model of combination
- $d_P^{(2)}$ is the L_2 pseudo-metric based on $P(A, B) = \frac{p(A \cap B)}{p(A)p(B)}$



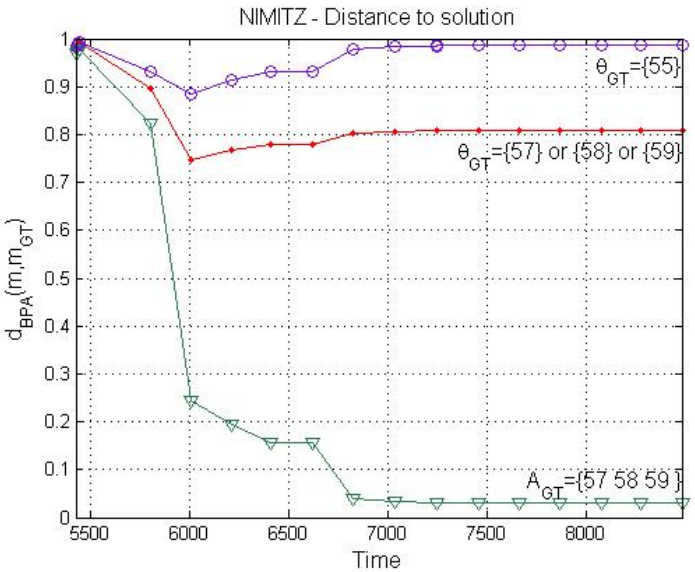
Accuracy assessment - Distance to different solutions

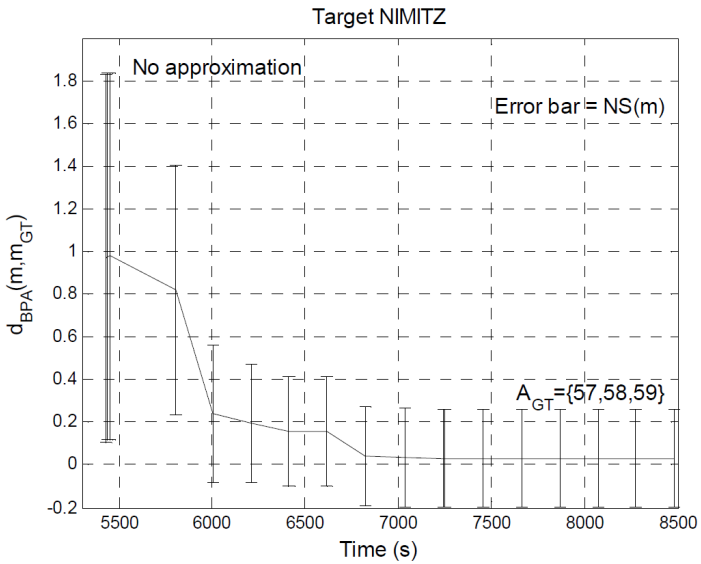
See [Jousselme etal, 2003]
Problem: Identification of individual objects belonging to a group of 5

$$\mathrm{Acc}(\mathrm{Alg}, A) = d_J^{(2)}(m, m_A^*)$$

with

- $m = \oplus_t m_t$ is the sequential combination of a series of BPAs $\{m_t\}$
- $d_J^{(2)}$ is the L_2 metric based on $Jac(A, B) = \frac{|A \cap B|}{|A \cup B|}$





- Can be used as a decision rule as $A* = \arg\min_{A \subseteq X} d(m, m_A)$

Accuracy assessment - Parameters optimization

Problem: Select the best set of parameters Γ of a supervised classifier
If we denote by Γ the set of internal parameters of a classifier Clf_Γ which classification output for sample $\mathbf{z}$ (belonging to the training set Z) is $m_z^{(\Gamma)}$. Then, the optimal set of parameters is:

$$\Gamma^* = \arg\min_{\Gamma} \sum_{\mathbf{z} \in Z} d(m_z^{(\Gamma)}, m_z^*)$$

where m_z^* is the categorical BPA focused on the singleton class of $\mathbf{z}$
The parameters are either used in

- the m_z definition (*e.g.*, evidential k_{NN} [Zouhal & Denoeux, 1998])
- source correction mechanism, *e.g.*discounting (see David's talk)

Adequate measure of accuracy?

- Reflexivity, $d(m, m) = 0$, is a desirable property for assessing accuracy
- For instance, using Dempster's conflict would not make sense
- Separability, $d(m_1, m_2) = 0 \Rightarrow m_1 = m_2$, may be required
- Separability may be satisfied in $\mathcal{E}_x$ while not in $\mathcal{E}_X$ $(d(m, m_x) = 0 \Rightarrow m = m^*)$
- The choice of the weighting matrix may be linked to the decision rule

Credibility assessment

Credibility refers here to how much a given BPA is consistent within a group of other BPAs (without any access to ground truth)

- Principle: The farthest (or the more conflicting) a BPAs from (with) the others, the lower its credibility

The credibility of $m \in G_K$ (relatively to the group G_K of BPAs) can be assessed for instance as:

- (Martin etal, 2008)

$$\mathrm{Cre}_1(m) = \left([1 - d(m, m_\oplus)]^\lambda\right)^{\frac{1}{\lambda}}$$

  where $m_\oplus = \oplus_{k=1, m \neq m_k}^{K} m_k$ and $\lambda > 0$

- (Florea etal, 2009)

$$\mathrm{Cre}_2(m) = \frac{|\{m_k \in G_K, m \neq m_k | d(m, m_k) \leq \tau\}|}{K - 1}$$

  where τ is a maximum threshold on the "distance" measure

- ...

Credibility assessment - Practical uses

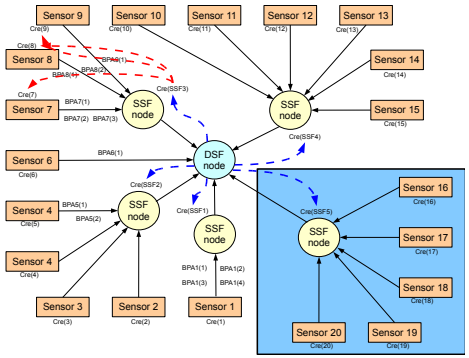
Such a measure of credibility can be used to:

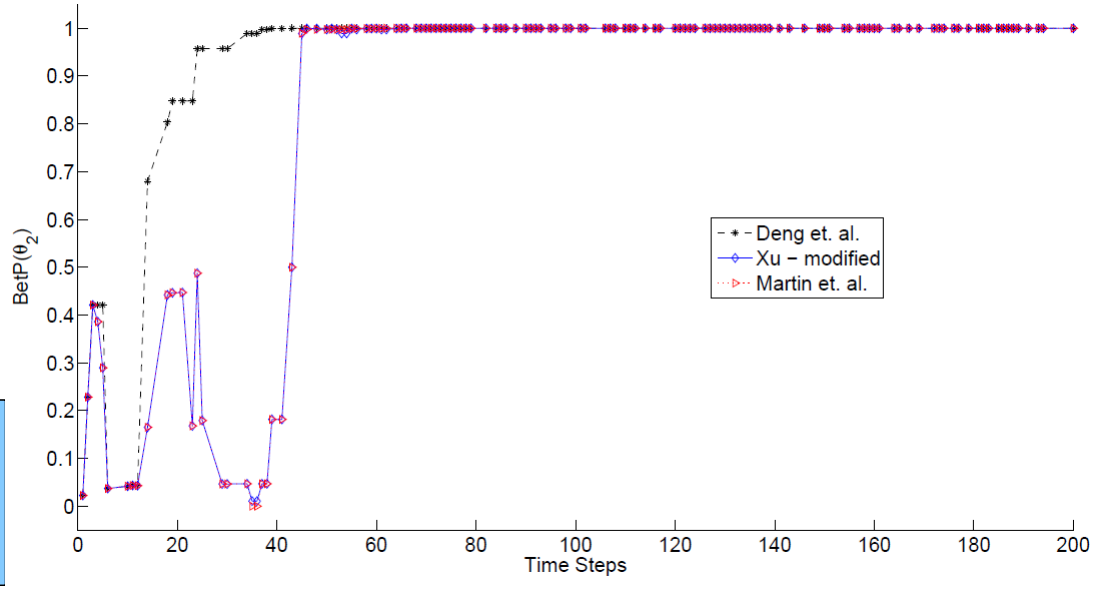
1. discount the BPAs with low credibility assuming a poor reliability of the source
2. discard the BPAs with low credibility assuming an association problem (they correspond to another object)
3. ...

Dynamic credibility assessment for source's discounting and discarding

See [Florea etal, 2009]

- Identification problem with fusion
- Distance measures are used to assess the information credibility
- Non-credible pieces of information are discounted or discarded from the fusion process
- Several credibility measures are compared based on the final pignistic probability of the targeted object





Adequate measure of credibility?

1. Reflexivity may not be required for credibility assessment
2. For instance, Dempster's conflict would make sense
3. Separability may not be required (?)

Information loss assessment

Information loss (or gain) refers here to how much the information contents of two *related* belief functions differ

$$m' = f(m)$$

- Either m' is the approximation of m (when not issued from a distance optimization problem) (see [Denoeux, 2001])
- Either $m' = m \oplus m_0$ is the result of the combination of m with m_0
- ...

Adequate measure of information loss?

1. Reflexivity and separability may be both required for information loss assessment
2. Information-based distances may be appropriate

Outline

Some applications
Quality assessment
Decision criterion

Decision criterion

The measure is used as a criterion to optimize:

$$m^* = \arg \min_{m' \in \mathcal{E}_X} d(m, m')$$
$$\text{subjected to } C_1, C_2, \ldots$$

where C_i are some constraints.
Examples are:

- Classification: $m' = m_A$ is a categorical BPA and the optimum hypothesis as the "closest" one to the current BPA m
- Belief function approximation
- Evidential pattern matching
- Evidential information retrieval

Decision criterion - Belief function approximation

See [Cuzzolin, 2007, 2009]
Problem: Find an approximation $\tilde{m}$ of a given belief function m

- The approximation problem can be formalized as an optimization problem under constraints as:

$$\tilde{m} = \arg \min_{m \in \mathcal{E}_X} d(m, \tilde{m})$$
$$\text{subject to } C_1, C_2, \ldots$$

where $\tilde{m}$ is the approximation of m and C_i are some constraints such as:

  - $C : \tilde{m}$ is a Bayesian BPA, can be seen as the projection of m over the subspace of probabilities, $\mathcal{E}_x$ (Bayesian approximation, [Cuzzolin, 2007])
  - $C : \tilde{m}$ is a consistent BPA (consistent approximation, [Cuzzolin, 2009])

Decision criterion - Evidential pattern matching

See [Jousselme & Maupin, 2012]
Problem: Object classification with an evidential training dataset

$$\text{Class}(\mathbf{z}) = \text{Class}\left(\arg\min_{\mathbf{z}_k \in Z} d(m_z^{(k)}, m_z)\right)$$

- Hyp: Similar objects lead to similar belief functions
- Equivalent to blind estimation (when measurements are not (or only partially) accessible)
- May be useful when subjective uncertainty is expressed
- PM result in the feature space (1-NN): 0.30

	Minkowski family								
	L_1			L_2			L_∞		
	min	max	mean	min	max	mean	min	max	mean
I	0.266	0.333	0.291	0.266	0.333	0.302	0.266	0.350	0.308
Jac	0.249	0.333	0.291	0.249	0.315	**0.272**	0.269	0.315	**0.285**
IncInc′	0.252	0.315	**0.279**	0.252	0.333	0.285	0.249	0.333	**0.285**
Int′Int	0.252	0.315	**0.279**	0.252	0.333	0.285	0.249	0.333	**0.285**
Bet′$_x$Bet$_x$	0.271	0.315	0.291	0.271	0.315	0.291	0.271	0.315	0.291
Bet′Bet	0.271	0.315	0.291	0.271	0.315	0.291	0.271	0.315	0.291

	Inner product family						Fidelity family		
	IP			cos			Hellinger		
	min	max	mean	min	max	mean	min	max	mean
I	0.281	0.368	0.335	0.249	0.333	0.285	0.252	0.333	**0.279**
Jac	0.259	0.381	0.325	0.249	0.315	**0.272**	0.253	0.333	0.285
IncInc′	0.241	0.354	0.283	0.252	0.333	0.285	0.252	0.333	**0.279**
Int′Int	0.384	0.512	0.454	0.269	0.315	0.285	0.315	0.464	0.406
Bet′$_x$Bet$_x$	0.252	0.315	**0.274**	0.253	0.315	0.285	0.252	0.333	**0.279**
Bet′Bet	0.252	0.315	**0.274**	0.271	0.315	0.291	-	-	-

AUC for different measures with a real dataset of vehicle identification

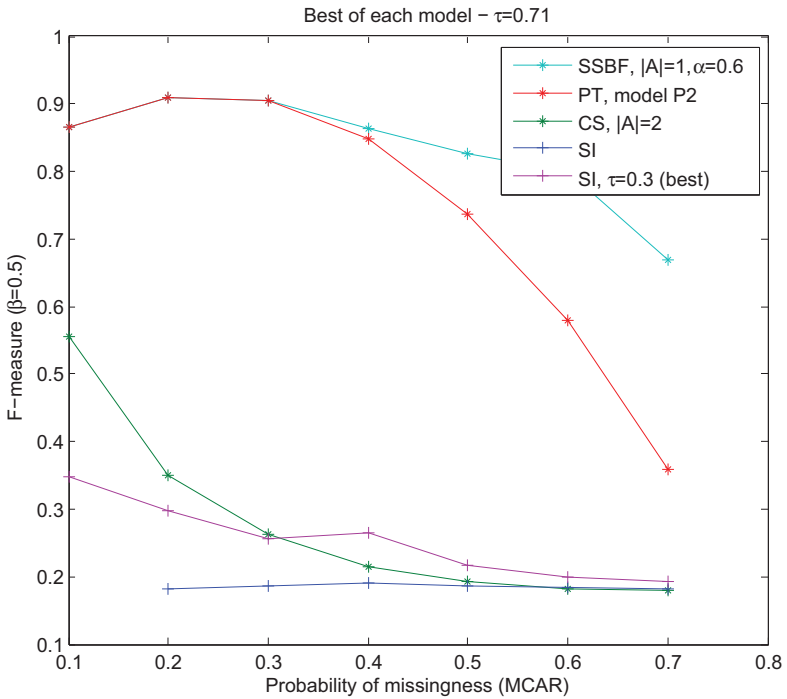
Information retrieval with missing data

See [Jousselme & Maupin, 2013]
Problem: Retrieve the set of relevant items r_q to a given query q from an evidential database

$$r_q = \{\mathbf{z} \in Z | d(\mathbf{z}, q) \leq \tau\}$$

where τ is a threshold over d.

- Missing data are replaced by belief functions with different models (classical sets, probabilities, belief functions)
- $d_J^{(2)}$ is the Jaccard metric
- Consistency properties with classical sets and probabilities were considered



Comparison of several uncertainty models with the F-measure criterion

Summary (1)

1. Basic Probability Assignments (BPAs) are interpreted as vectors in the belief space
2. The belief, plausibility, commonality functions and the pignistic probability are linear transformations of the BPA vector
3. The belief space is an inner product space with the inner product $\mathbf{m}_1' \mathbf{W} \mathbf{m}_2$
4. The belief space is a normed space , the norm being defined by the L_p Minkowski family of norms
5. Inner products (and cosines) measure the orthogonality between belief functions
6. Metrics measure the distance between belief functions
7. Dempster's conflict does not satisfy the basic metric property of reflexivity

Summary (2)

8. The choice of the measure can be guided by
   (1) the metric properties (reflexivity, separability),
   (2) the meaning of the **W** matrix,
   (3) the nature of the norm (p value in L_p)

   and their adequacy to the practical problem at hand
9. Practical uses of distance can be split into
   (1) information/algorithm quality assessment ("accuracy", "credibility", "information loss"): The choice of the measure should be guided by its semantics and properties
   (2) (direct) decision criterion: The choice of the measure can be guided by external performances criteria (AUC, F-measure)

➔ The practical use of distance has the only limit of your imagination ...

References (1)



Distances

- A.-L. Jousselme and P. Maupin, "Distances in evidence theory: Comprehensive survey and generalizations," *International Journal of Approximate Reasoning*, vol. 53, pp. 118–145, February 2012.
- A.-L. Jousselme, D. Grenier, and E. Bossé, "A new distance between two bodies of evidence," *Information Fusion*, vol. 2, pp. 91–101, June 2001.
- M. Bouchard, A.-L. Jousselme, and P.-E. Doré, "A proof for the positive definiteness of the Jaccard index matrix," *International Journal of Approximate Reasoning*, vol. 54, no. 5, pp. 615–626, 2013.



Geometry of belief functions

- F. Cuzzolin, "A Geometric Approach to the Theory of Evidence," IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews, vol. 38, no. 4, pp. 522–534, 2008.
- F. Cuzzolin, "Geometry of Dempster's rule of combination," IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, vol. 34, no. 2, pp. 961–977, 2008.

References (2)

Matrix calculus of belief functions

- P. Smets, "The application of the matrix calculus to belief functions," *International Journal of Approximate Reasoning*, vol. 31, pp. 1–30, 2002.

Other references used to prepare this presentation

- W. Liu, "Analyzing the degree of conflict among belief functions," *Artificial Intelligence*, vol. 170, pp. 909–924, 2006.
- S. Destercke and T. Burger, "Toward an axiomatic definition of conflict between belief functions," *IEEE Trans. of Systems, Man and Cybernetics - Part A: Systems and Humans*, 2013.
- B. Ristic and P. Smets, "The TBM global distance measure for the association of uncertain combat ID declarations," *Information Fusion*, vol. 7, pp. 276–284, 2006.
- Z. Sunberg and J. Rogers, "A belief function distance metric for orderable sets," *Information Fusion*, vol. 14, pp. 361–373, 2013.
- J. Diaz, M. Rifqi, and B. Bouchon-Meunier, "A similarity measure between basic belief assignments," in *Proceedings of the 9th International Conference Information Fusion*, (Firenze, Italy), 2006.
- D. Attiaoui, D. Pierre-Emmanuel, A. Martin, and B. Ben Yaghlane, "A distance between continuous belief functions," in *Proc. of the Scalable Uncertainty Management (SUM) conference*, vol. 7520 of Lecture Notes in Artificial Intelligence, pp. 194–205, Springer-Verlag Berlin Heidelberg, 2012.
- Z. Elouedi, K. Mellouli, and P. Smets, "Belief decision trees: theoretical foundations," International Journal of Approximate Reasoning, vol. 28, no. 2-3, pp. 91–124, 2001.
- M. Louhadi, J. Klein, J.-M. Vannobel, and O. Colot, "Fast computation of lp norm-based specialization distances between bodies of evidence," in Proc. of the int. conf. on Belief Functions (BELIEF 2014), (Oxford, UK), 2014.

References (3)



Applications

- D. Fixsen and R. P. S. Mahler, "The Modified Dempster-Shafer Approach to Classification," *IEEE Trans. on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 27, pp. 96–104, Jan. 1997.
- A.-L. Jousselme, E. Bossé, and A. Jouan, "Analysing an identity information fusion algorithm based on evidence theory," in IST-040/RSY-012, RTO Symposium on Military Data and Information Fusion, (Prague, Czech Republic), 20-22 October 2003.
- T. Denoeux, "Inner and outer approximation of belief structures using a hierarchical clustering approach," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 9, no. 4, pp. 437–460, 2001.
- L. M. Zouhal and T. Denoeux, "An evidence-theoretic k-NN rule with parameter optimization," IEEE Transactions on Systems, Man and Cybernetics - Part C, vol. 28, no. 2, pp. 263–271, 1998.
- M. C. Florea, A.-L. Jousselme, and E. Bossé, "Dynamic estimation of evidence discounting rates based on information credibility," RAIRO Operations Research, vol. 44, pp. 285–306, October 2011.
- F. Cuzzolin, "Consistent approximations of belief functions," in 6th International Symposium on Imprecise Probability: Theories and Applications, (Durham, United Kingdom), 2009.
- F. Cuzzolin, "Two new Bayesian approximations of belief functions based on convex geometry," IEEE Transactions on Systems, Man, and Cybernetics part B, vol. 37, pp. 993–1008, August 2007.
- A.-L. Jousselme and P. Maupin, "Comparison of uncertainty representations for missing data in information retrieval," in Proceedings of the International Conference of Information Fusion, (Istanbul, Turkey), July 2013.
- A.-L. Jousselme and P. Maupin, "An evidential pattern matching approach for vehicle identification," in Advances in Intelligent and Soft Computing 164, Proc. of the 2nd Int. Conf. on Belief Functions (Denoeux and Masson, eds.), (Compiègne, France), pp. 45–52, Springer-Verlag, May 2012.

Anne-Laure Jousselme
Anne-Laure.Jousselme@cmre.nato.int