Algebraic and metric structures for belief functions

John Klein

Université de Lille - CRIStAL UMR CNRS 9189

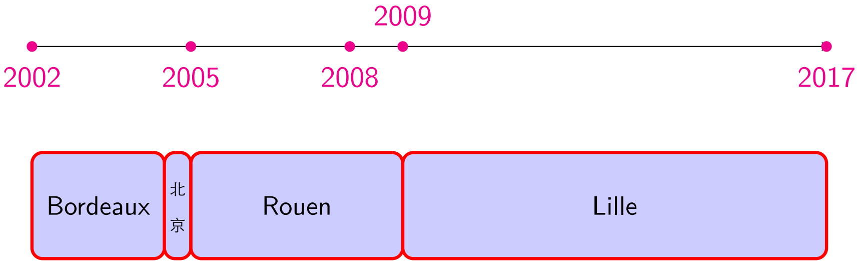
HDR Defense

07/12/2017

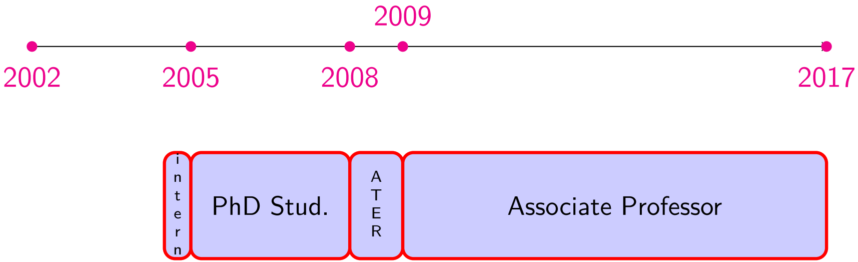
Presentation outline

Presentation outline

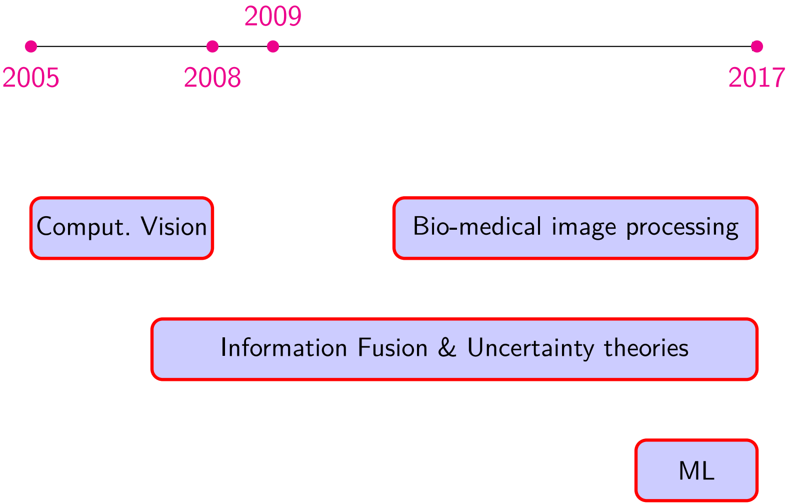
Salient career facts : Places



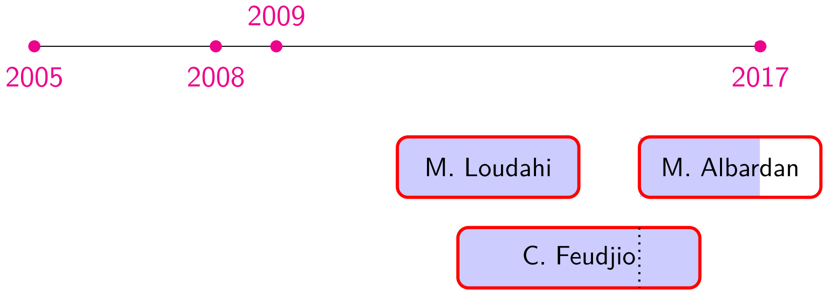
Salient career facts : Positions



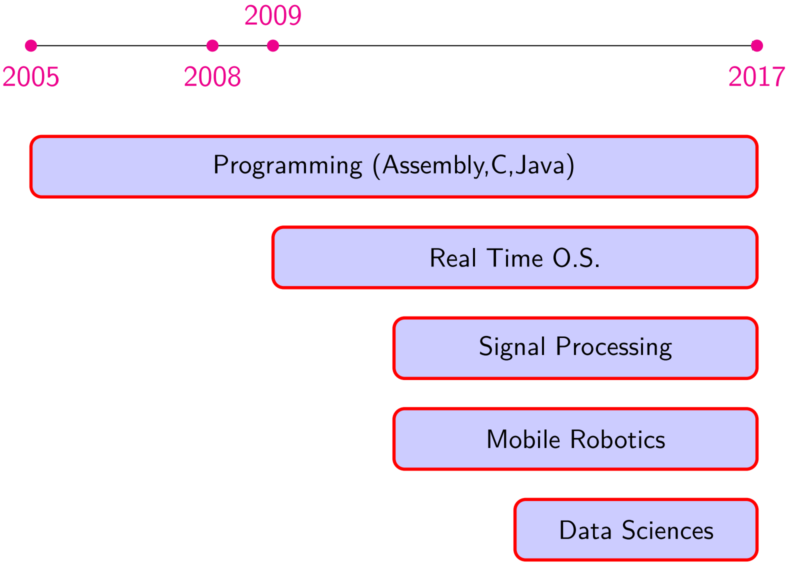
Salient career facts : Research activities



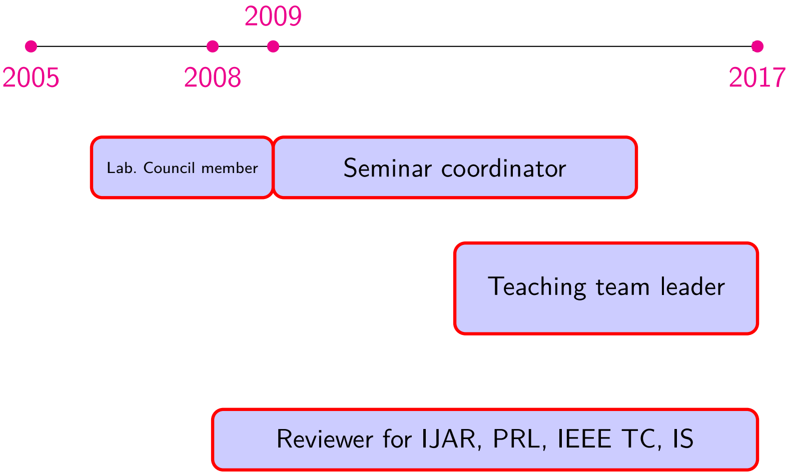
Salient career facts : PhD supervisions



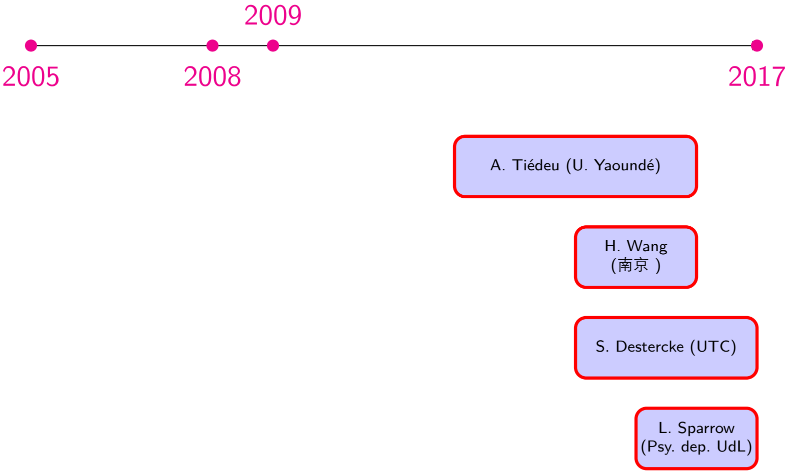
Salient career facts : Teaching activities



Salient career facts : Misc.



Salient career facts : Collaborations



Starting : P. Bas (SIGMA team), C. Versari (Biocomputing team), B. Guedj (INRIA Lille) & A. Bouscayrol (Electrical. Eng. Lab. UdL).

Presentation outline

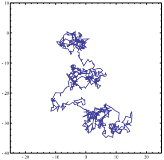
Why do we need uncertainty models ?

- Aleatory nature of the problem :

Light source $\longrightarrow$ sensing device

- The world is too complex :

Trajectory of a particle in a gas ?

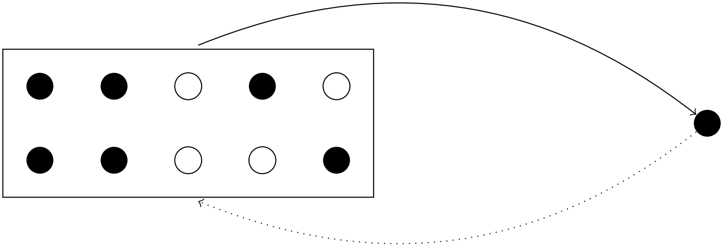


- There is missing information :

Missing entries in a training example $\quad \mathbf{x} = [1.2; ?; 0.4]$

Probabilities

- Limiting frequency of occurrences of some event.



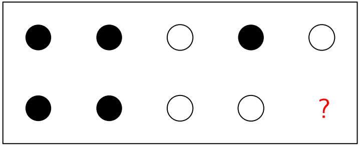
- Subjective degree of belief of an event being true.

  Probability that patient X is sick given observed symptoms

- $\rightarrow$ Unified mathematical framework : probability measures.

Can probabilities cover any situation involving uncertainty ?

- Content of the urn revealed except for one marble !



- Let X denote the color of a randomly picked marble.
- Let Z denote the color of the last marble.
- What is the probability of $X = \bullet$?

$$P(X=\bullet) = \underbrace{P(X=\bullet|Z=\bullet)}_{0.6} P(Z=\bullet) + \underbrace{P(X=\bullet|Z=\circ)}_{0.5} P(Z=\circ)$$

(Total Probability Theorem)

Probabilistic solutions

- Naive idea : choose a uniform distribution for $Z \sim \text{Ber}\left(\frac{1}{2}\right)$
  $\rightarrow$ inconsistent probability value : $P(X = \bullet) = 0.55$.

- Hierarchical idea : investigate candidate distributions for X :

$$X \sim \mu_1 = \text{Ber}(0.5) \quad \text{or} \quad X \sim \mu_2 = \text{Ber}(0.6)$$

  Define a variable H taking values in the set $\{\mu_1; \mu_2\}$ and $H \sim \text{Ber}(\theta)$.
  $\rightarrow$ still need to set a parameter.

- Conservative idea : use lower and upper bounds :

$$\min\{\mu_1(\bullet); \mu_2(\bullet)\} \leq P(X = \bullet) \leq \max\{\mu_1(\bullet); \mu_2(\bullet)\}$$

- The belief function idea : use probabilities with a different hardcore logic.
  $\rightarrow$ 100% neutral while fitting data.

From Probabilities to Belief functions [Dempster 67, Shafer 76, Dempster 08]

- Suppose an assertion $B \in \sigma_\Theta$ is described by the following triplet of epistemic states :

$$\{\text{true; false; don't know}\}.$$

- Let us assign probabilities (u, v, w) to these states and $u + v + w = 1$.

- Obviously, B^c is described by the triplet (v, u, w).

- The set function mapping B to probabilities u is denoted bel and is called a belief function.

- The plausibility function pl is given by

$$pl(B) = 1 - bel(B^c) = 1 - v = u + w.$$

- Our ignorance on B is featured by $w = 1 - u - v = pl(B) - bel(B)$!

Belief functions

- For any $B' \subseteq B$, B' being true logically implies that B is true.

- A convenient representation is given by the mass function m :

$$bel(B) = \sum_{B' \subseteq B} m(B'). \tag{1}$$

- $m(B)$ is the support given to B based on available evidence that cannot be refined to subsets of B and

$$\sum_{B \subseteq \Theta} m(B) = 1 \quad \text{and} \quad m(B) \geq 0, \forall B \subseteq \Theta. \tag{2}$$

Belief functions and the marbles

- Assign the following (u,v,w) triplet to $H = \mu_1$:

$$\{H = \mu_1\} \rightarrow (0; 0; 1) \tag{3}$$

$\rightarrow$ In the continuous case, the wine/water example also features the ability of belief functions to fit imprecise data.

Challenge : fit imprecise data to models

- Example 1 : Let $I_{\mathbf{p}}$ denote the value of a grayscale image at pixel $\mathbf{p}$.
- Let $a_{\mathbf{p}}$ denote the sensor value at pixel $\mathbf{p}$.
- For any positive $\tau < \tau'$, $I_{\mathbf{p}} \in [a_{\mathbf{p}} - \tau; a_{\mathbf{p}} + \tau]$ is less likely than $I_{\mathbf{p}} \in [a_{\mathbf{p}} - \tau'; a_{\mathbf{p}} + \tau']$.

[BELIEF 2012]

Choose a non-decreasing function F w.r.t. τ and obtain a mass function with positive masses assigned to nested intervals $[a_{\mathbf{p}} - \tau; a_{\mathbf{p}} + \tau]$,

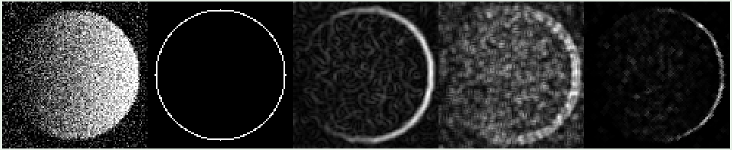
$m_{\mathbf{p}}$: model for the pixel value

Challenge : fit imprecise data to models

[BELIEF 2012]

Take the set of mass functions in the neighborhood of pixel $\mathbf{p}$:

Conflict κ = Edge detector



$$\kappa_{\left\{m_{\mathbf{p}};m_{\mathbf{p}'}\right\}} = \sum_{\substack{A,B\subseteq\Theta \\ A\cap B=\emptyset}} m_{\mathbf{p}}\left(A\right) m_{\mathbf{p}'}\left(B\right)$$

Challenge : fit imprecise data to models

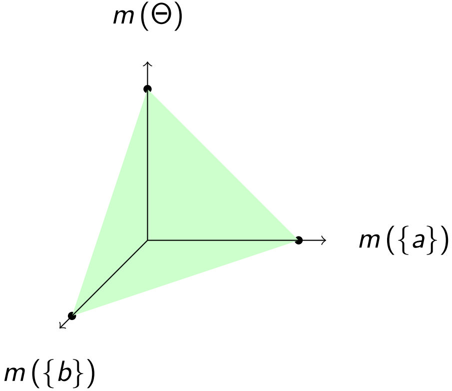
- Example 2 : Let Γ_i denote the output predicted by a 1-vs-all probabilistic classifier.
- Suppose $\Theta = \{y_1; \ldots; y_\ell\}$ is the set of classes.
- Γ_i is a random set, for instance the 1st classifier returns

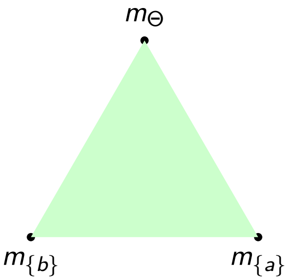
$$\text{either } y_1 \text{ or } \{y_2; \ldots; y_\ell\}$$

- Γ_i is an ill known random variable on Θ and induces a probability distribution on a sub-σ-field, i.e. a belief function on Θ.

Where do belief functions live? Where do mass functions live?

- Let m_B denote the (categorical) mass function encoding the information $X \in B$, i.e. $m_B(B) = 1$.
- The set of mass functions $\mathcal{M}$ (mass space) is the simplex spanned by the categorical mass functions.



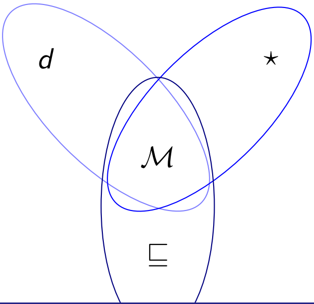


Structures for $\mathcal{M}$

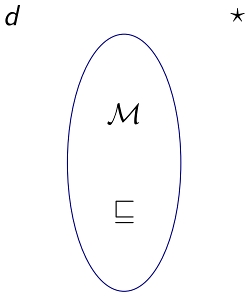
- a partial order $\sqsubseteq$ $\rightarrow$ order theoretic structure,
- a distance d $\rightarrow$ metric space structure,
- a combination rule $\star$ $\rightarrow$ algebraic structure.

Problem

Get consistent results across structure types



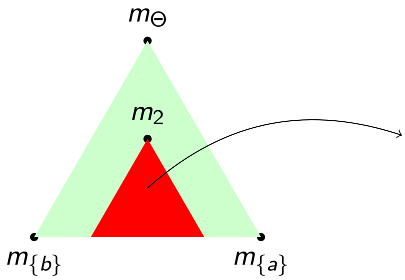
Presentation outline


d
$\star$
$\mathcal{M}$
$\sqsubseteq$

$\mathcal{M}$ as a partially ordered set

- Does $m_1 = \frac{1}{2} m_{\{a\}} + \frac{1}{2} m_{\{b\}}$ convey more information than $m_2 = \frac{1}{3} m_{\{a\}} + \frac{1}{3} m_{\{b\}} + \frac{1}{3} m_\Theta$?
  Example :

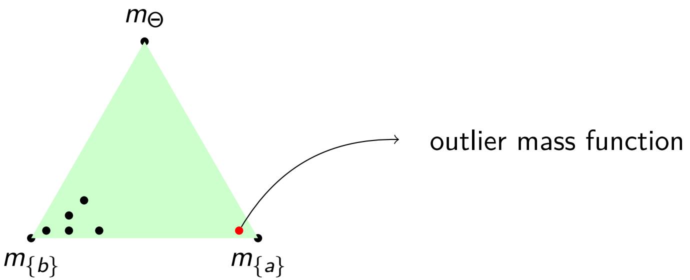
$$m_1 \sqsubseteq_{pl} m_2 \Leftrightarrow pl_1(B) \leq pl_2(B), \forall B \subseteq \Theta. \quad (4)$$



$$\mathcal{S}_{pl}(m_2) = \{m : m \sqsubseteq_{pl} m_2\}$$

Pre-orders for a family $\mathcal{A} = \{m_1, \dots, m_\ell\}$ of mass functions

- How does each mass function contribute to the family inconsistency ?



[IJAR 2011]

$$\xi(m_i) = \frac{1}{\ell - 1} \sum_{m_j \in \mathcal{A} \setminus \{m_i\}} \kappa_{\{m_i;m_j\}}, \tag{5}$$

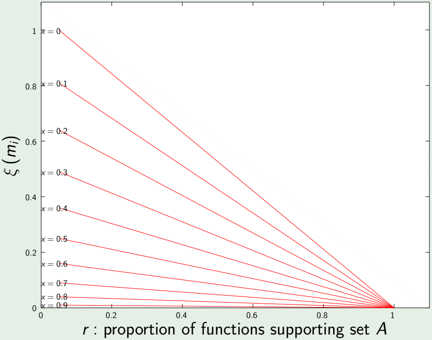
if there is at least one pair with $\kappa_{\{m_i;m_j\}} > 0$.

[IJAR 2011]

- Take two disjoints sets A and B.
- If $m \in \mathcal{A}$, then

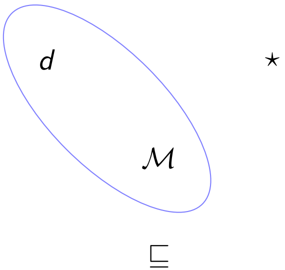
$$m = (1-x)\, m_A + x m_\Theta \quad \text{or} \quad m = (1-x)\, m_B + x m_\Theta .$$

$\rightarrow \xi(m_i)$ is linear w.r.t. r (ratio of those functions supporting A) and does not depend on ℓ.



(not achieved by [Martin 2008] or [Schubert 2010])

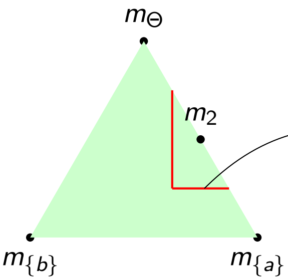
Presentation outline


d
$\star$
$\mathcal{M}$
$\sqsubseteq$

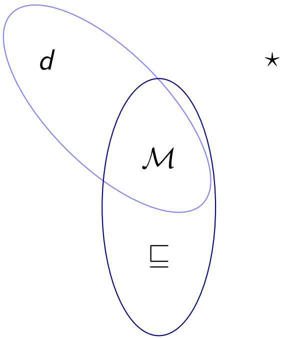
$\mathcal{M}$ as a metric space

- How far away are m_1 and $m_2 = \frac{1}{2} m_{\{a\}} + \frac{1}{2} m_\Theta$?
  Example :

$$d_{pl,k}\left(m_1, m_2\right) = \frac{1}{\rho} \left[\sum_{B \subseteq \Theta} \left| pl_1\left(B\right) - pl_2\left(B\right) \right|^k \right]^{1/k}. \tag{6}$$



$$\left\{ m : d_{pl,1}\left(m, m_2\right) = \frac{1}{8} \right\}$$


d
$\star$
$\mathcal{M}$
$\sqsubseteq$

$\mathcal{M}$ as both a metric space and a partially ordered set

- To what extent $\sqsubseteq$ and d carry compatible attributes of $\mathcal{M}$?

- Notion of consistency

[IJAR 2016] (with S. Destercke)

d is $\sqsubseteq$-**compatible** if for any mass functions m_1, m_2 and m_3 such that $m_1 \sqsubseteq m_2 \sqsubseteq m_3$, we have :

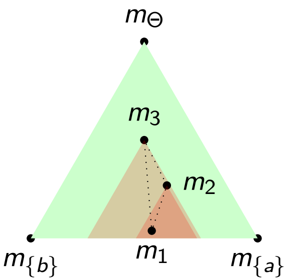
$$\max\left\{d\left(m_1, m_2\right); d\left(m_2, m_3\right)\right\} \leq d\left(m_1, m_3\right), \tag{7}$$

$\rightarrow$ formalization of principles from [Jousselme 2012]

$\mathcal{M}$ as both a metric space and a partially ordered set

[IJAR 2016] (with S. Destercke)

Several consistency results, e.g. $d_{pl,k}$ is $\sqsubset_{pl}$-compatible ($k < \infty$).



$\mathcal{M}$ as both a metric space and a partially ordered set

[IJAR 2016] (with S. Destercke)

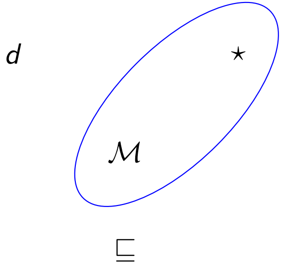
Consistency is useful for belief function approximation

- Problem : find $m_* \in \mathcal{R}$ s.t.
  - (i) m_* approximates m,
  - (ii) m_* is more informative than m.
- Solution : use distance d_{pl} which is consistent with $\sqsubseteq_{pl}$:

$$m_* = \underset{m' \in \mathcal{R} \cap \mathcal{S}_{pl}(m)}{\arg\min} \; d_{pl}\left(m, m'\right)$$

- Guarantee : m_* is a maximal inner approximation :
  For any $m' \in \mathcal{R} \cap \mathcal{S}_{pl}(m)$, we have $m_* \not\sqsubseteq_{pl} m'$.

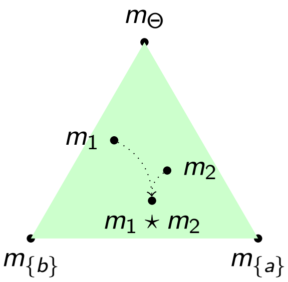
Presentation outline


d
$\star$
$\mathcal{M}$
$\sqsubseteq$

$\mathcal{M}$ as a magma

- One just needs to endow $\mathcal{M}$ with some binary operation (or combination rule)

$$\begin{aligned} \mathcal{M} \times \mathcal{M} &\to \mathcal{M}, \\ (m_1, m_2) &\to m_1 \star m_2. \end{aligned} \tag{8}$$



$\mathcal{M}$ as a magma

- Example : the unnormalized version of Dempster's rule, i.e. the conjunctive rule $\textcircled{\cap}$

$$m_1 \textcircled{\cap} m_2 (E) = \sum_{\substack{A,B \subseteq \Theta, \\ A \cap B = E}} m_1 (A)\, m_2 (B)\,, \tag{9}$$

$\mathcal{M}$ as a monoid

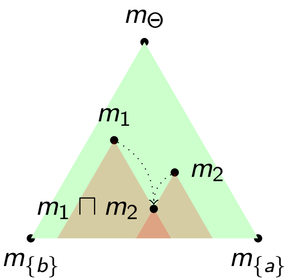
- Monoid = Magma with an associative rule and a neutral element
- Also true for $\textcircled{\cap}$: m_Θ is the neutral element and associativity holds.

$\mathcal{M}$ as a magma + idempotence

- How to cautiously combine mass functions while ensuring increased informative content ?

[BELIEF 2016] (with S. Destercke)

$$m_1 \sqcap m_2 = \underset{m \in \mathcal{S}_{pl}(m_1) \cap \mathcal{S}_{pl}(m_2)}{\arg\min} d(m, m_\Theta) . \quad (10)$$



$\mathcal{M}$ as a magma + idempotence

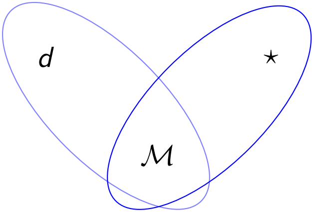
[BELIEF 2016] (with S. Destercke)

Features of rule $\sqcap$

operator	condition for use	commutativity	associativity	idempotence	neutral element
$\textcircled{\cap}$ [Smets 90]	none	yes	yes	no	m_Θ
$\oplus$ [Dempster 67]	$m_{1\textcircled{\cap}2}(\emptyset) < 1$	yes	yes	no	m_Θ
$\textcircled{\wedge}$ [Denoeux 08]	$m_1(\Theta) > 0$ and $m_2(\Theta) > 0$	yes	yes	yes	none
$\sqcap$ [BELIEF 2016]	none	yes	quasi	yes	m_Θ

[IJAR 2018] (with S. Destercke)

$\rightarrow$ an idempotent distance based disjunctive rule.



$\sqsubseteq$

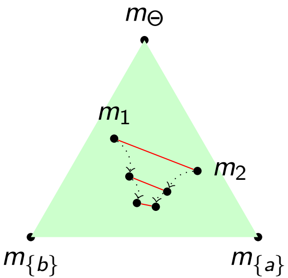
$\mathcal{M}$ as both a monoid and a metric space

- To what extent $\star$ and d carry compatible attributes of $\mathcal{M}$?
- Another notion of consistency is needed

[IJAR 2014 + IEEE TC 2016] (M. Loudahi's PhD)

d is **consistent** w.r.t. $\star$ if for any mass functions m_1, m_2 and m_3 on Θ :

$$d(m_1 \star m_3, m_2 \star m_3) \leq d(m_1, m_2) . \quad (11)$$



$\mathcal{M}$ as both a monoid and a metric space

[IJAR 2014 + IEEE TC 2016] (M. Loudahi's PhD)

Several results proving distance/rule consistencies.

Example :

- map each mass function m to its Dempsterian matrix $\mathbf{D}$
- Each column of $\mathbf{D}$ is $m \circledcap m_B$ for some $B \subseteq \Theta$.
- Define $d(m_1, m_2) = \frac{1}{\rho} \left\| \mathbf{D}_1 - \mathbf{D}_2 \right\|_1$.
- The metric d is consistent with $\circledcap$.

Presentation outline

Conclusions

- Belief functions provide a large spectrum of models for reasoning under uncertainty.
- We use them when there is ignorance (misspecified priors).
- Downside : greater time and memory complexities and more subtle calculus rules.
- 3 selected contributions :

Proved which distance agrees with which combination rule.

Proved which distance agrees with which partial order.

Introduced a new rule that agrees with the least commitment principle.

Research directions

- Structure of the mass space

Open question

Consistency with both a partial order and a combination rule ?

Open question

What about consistency between metrics and specificity pre-orders ?

- Applications of belief functions or imprecise probabilities to signal and image processing.

Under preparation

An upper probability model for pixel values.

$\rightarrow$ What's new ? It takes the sensor technology into account.

Research directions

- Applications of belief functions or imprecise probabilities to machine learning.

Idea

Use behavior based combination [Pichon et al. 2014] to combine classifiers (in the wake of M. Albardan's PhD).

m : 1 vs all classifier output

m_{meta} classifier performances on a validation set

$$m_{\text{rec}}(B) = \sum_{\substack{A \subseteq \Theta, H \subseteq \mathcal{S} \\ \bigcup_{s \in H} \Gamma_A(s) = B}} m(A) \times m_{\text{meta}}(H). \tag{12}$$

Research directions (long term)

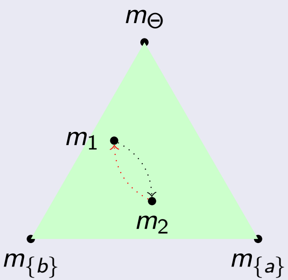
- Belief functions and knowledge representation in AI.

Open question

Can we achieve evidence deletion ?

Idea

Define a pair of rules for insertion/deletion of evidence.



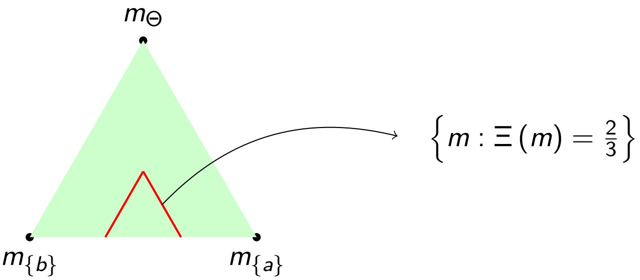
Thank you for your attention.

$\mathcal{M}$ as a pre-ordered set

- Inconsistency pre-order : does $m_1 = \frac{1}{2} m_{\{a\}} + \frac{1}{2} m_{\{b\}}$ encode less consistent information than $m_2 = \frac{1}{2} m_{\{a\}} + \frac{1}{2} m_{\{a,b\}}$?
  Example :

$$\Xi(m) = \max_{a \in \Theta} pl(\{a\}) \quad (13)$$

$\Xi(m_1) = \frac{1}{2}$ while $\Xi(m_2) = 1$.

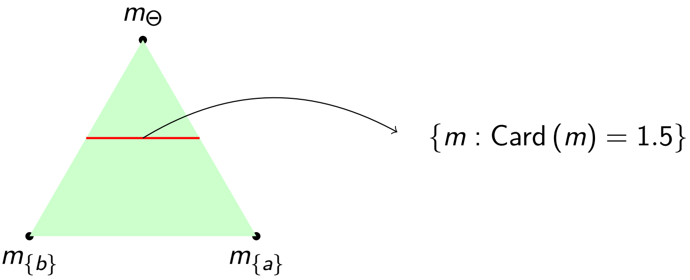


$\mathcal{M}$ as a pre-ordered set

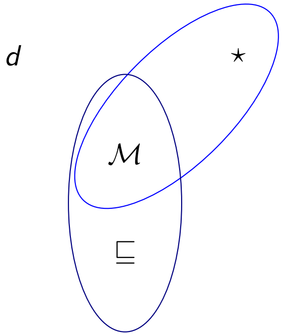
- Specificity : do probability masses in m_1 support large (imprecise) subsets as compared to m_2 ?
  Example :

$$\text{Card}(m) = \sum_{\substack{B \subseteq \Theta \\ B \neq \emptyset}} m(B)\,|B| \tag{14}$$

Card $(m) = 1$ while Card $(m_2) = 1.5$.

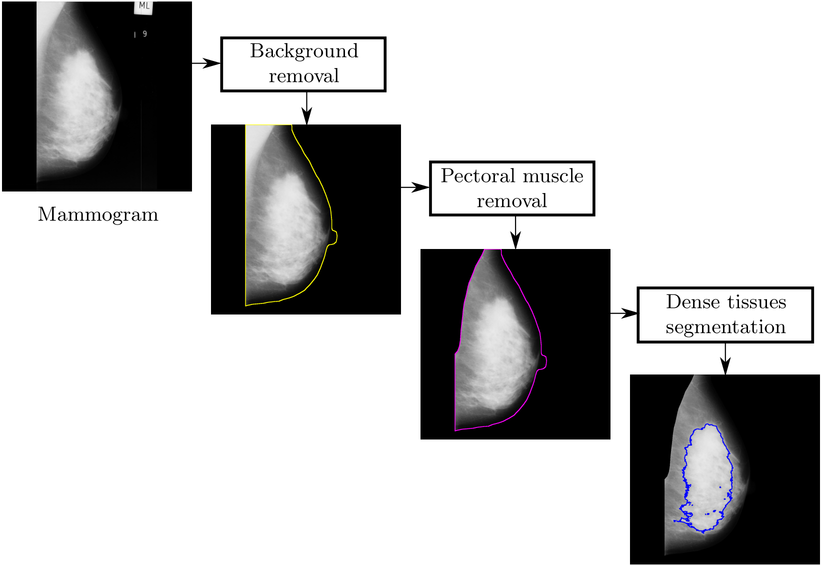


Order theoretic and algebraic structures for $\mathcal{M}$



$$m_1 = m_1 \star m_2 \Leftrightarrow m_1 \sqsubseteq_\star m_2. \tag{15}$$

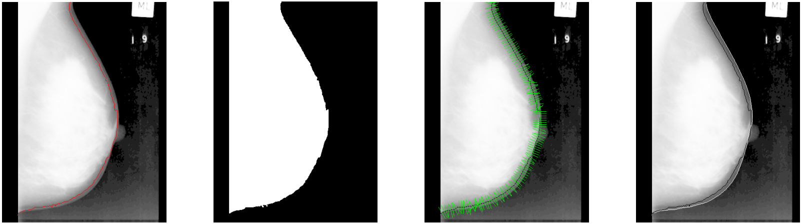
Medical image processing (C. Feudjio's PhD) : Mammogram segmentation



Medical image processing (C. Feudjio's PhD) : Mammogram segmentation

- Goal : prioritize patients (dense tissue ratio is correlated with cancer risk)

- 2 first steps : (spatial) Fuzzy C-means + pre and post-processing.

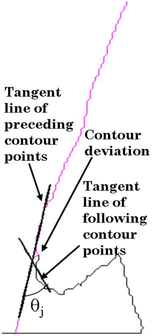


Correct edge points achieve maximal gradient norm among pixels the search segment (in green)

Medical image processing (C. Feudjio's PhD) : Mammogram segmentation

- Goal : prioritize patients (dense tissue ratio is correlated with cancer risk)

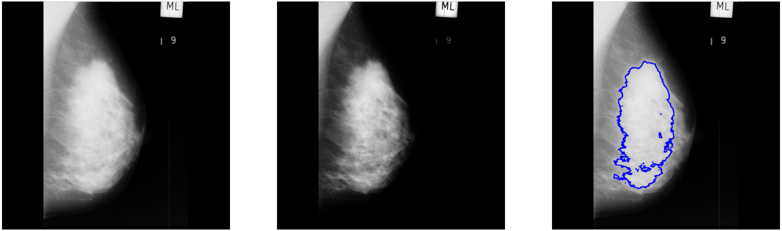
- 2 first steps : (spatial) Fuzzy C-means + pre and post-processing.



Curvature thresholding + similar contour post-processing

Medical image processing (C. Feudjio's PhD) : Mammogram segmentation

- Final step : histogram specification + threshold.

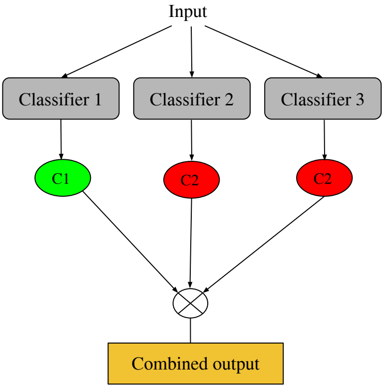


Gamma corrections as candidate transports

Obj. func. : Wasserstein dist. + regularizer based on HoG features

Machine Learning (M. Albardan's PhD) : Classifier combination

- Goal : alleviate the burden of choosing among equally appealing classification algorithms and take the lot !



Machine Learning (M. Albardan's PhD) : Classifier combination

- Idea : Use classification performances to combine classifiers
- Performances are evaluated on a validation set in the form of confusion matrices.
- Using these matrices, we have access to

$$P\left(\text{true class} \middle| k^{\text{th}} \text{ classifier output}\right).$$

- We need

$$P(\text{true class}|\text{all classifier outputs})$$

Machine Learning (M. Albardan's PhD) : Classifier combination

- Solutions :
  - Assume classifier output cond. indep. and apply Bayes theorem.

    $\rightarrow$ Application to physiological signal classification (collab. with L. Sparrow)

  - Propose a parametric model of $P(\text{true class}|\text{all classifier outputs})$ as an aggregation of the distributions derived from confusion matrices :

    $\rightarrow$ t-norm $+$ renormalization.
    $\rightarrow$ copula function (collab. with B. Guedj)

Elec. Engineering (collab. with H. Wang and S. Li) : Wind turbine control

- Goal : maximize energy production of a wind turbine
- For a given wind speed, there is an optimal turbine speed for power production.
- Solution : a controller for rotor currents based on sliding mode control.
- Sliding mode control use a discontinuous control signal to confine state trajectories to a desirable manifold (e.g. one that has an equilibrium state point).