Information Fusion in the Theory of Evidence

Frédéric Pichon

Laboratory of Computer Engineering and Automation of Artois (LGI2A)
Université d'Artois, Béthune, France

BFTA 2023
Nomi, Japan
October 28, 2023

Contents of this lecture

- Pieces of evidence, represented by belief functions $bel_1, \ldots, bel_n$ à la Shafer, about the actual (unknown) value of a variable X defined on a domain Ω.
- Information fusion

$$bel = f(bel_1, \ldots, bel_n)$$

Contents of this lecture

- Pieces of evidence, represented by belief functions $bel_1, \ldots, bel_n$ à la Shafer, about the actual (unknown) value of a variable X defined on a domain Ω.
- Information fusion

$$bel = f(bel_1, \ldots, bel_n)$$

  - Original proposal for f is Dempster's rule of combination.

Contents of this lecture

- Pieces of evidence, represented by belief functions $bel_1, \ldots, bel_n$ à la Shafer, about the actual (unknown) value of a variable X defined on a domain Ω.
- Information fusion

$$bel = f(bel_1, \ldots, bel_n)$$

  - Original proposal for f is Dempster's rule of combination.
  - Since then, a jungle of combination rules.

Contents of this lecture

- Pieces of evidence, represented by belief functions $bel_1, \ldots, bel_n$ à la Shafer, about the actual (unknown) value of a variable X defined on a domain Ω.
- Information fusion

$$bel = f(bel_1, \ldots, bel_n)$$

  - Original proposal for f is Dempster's rule of combination.
  - Since then, a jungle of combination rules.
- This lecture

Contents of this lecture

- Pieces of evidence, represented by belief functions $bel_1, \ldots, bel_n$ à la Shafer, about the actual (unknown) value of a variable X defined on a domain Ω.
- Information fusion

$$bel = f(bel_1, \ldots, bel_n)$$

  - Original proposal for f is Dempster's rule of combination.
  - Since then, a jungle of combination rules.
- This lecture
  1. Review of Dempster's rule

Contents of this lecture

- Pieces of evidence, represented by belief functions $bel_1, \ldots, bel_n$ à la Shafer, about the actual (unknown) value of a variable X defined on a domain Ω.
- Information fusion

$$bel = f(bel_1, \ldots, bel_n)$$

  - Original proposal for f is Dempster's rule of combination.
  - Since then, a jungle of combination rules.
- This lecture
  1. Review of Dempster's rule
  2. Guided tour of the jungle

Contents of this lecture

- Pieces of evidence, represented by belief functions $bel_1, \ldots, bel_n$ à la Shafer, about the actual (unknown) value of a variable X defined on a domain Ω.
- Information fusion

$$bel = f(bel_1, \ldots, bel_n)$$

  - Original proposal for f is Dempster's rule of combination.
  - Since then, a jungle of combination rules.
- This lecture
  1. Review of Dempster's rule
  2. Guided tour of the jungle
  3. How to choose f?

Outline

Outline

Mass, belief and plausibility functions

- A piece of evidence about a variable X taking values in a finite set Ω (frame of discernement) may be represented by a mass function $m : 2^{\Omega} \rightarrow [0, 1]$ such that

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

  and $m(\emptyset) = 0$.
- Every $A \subseteq \Omega$ such that $m(A) > 0$ is a focal set of m.
- $\mathcal{F}(m)$: set of focal sets of m.
- A piece of evidence may equivalently be represented by the belief function $bel : 2^{\Omega} \rightarrow [0, 1]$ or plausibility function $pl : 2^{\Omega} \rightarrow [0, 1]$

$$\begin{aligned} bel(A) &= \sum_{B \subseteq A} m(B), \\ pl(A) &= \sum_{B \cap A \neq \emptyset} m(B), \end{aligned}$$

Example

- My Apple MacBook has broken down.
- The cause X of the issue is either a power problem, a CPU malfunction, an hard drive failure or a corrupted OS, i.e., $\Omega = \{p, c, h, o\}$.
- A technician conducts an investigation and finds that $X \in \{p, c\}$.
  - If the investigation was conducted properly, we know that $X \in \{p, c\}$.
  - If the investigation was not conducted properly, we just know that $X \in \Omega$ (we know nothing).
- There is a chance 0.1 that the investigation was not conducted properly:
  - The probability of knowing that $X \in \{p, c\}$ is 0.9
  - The probability of knowing nothing is 0.1
- This piece of evidence about X can be represented by

$$m(\{p, c\}) = 0.9, m(\Omega) = 0.1$$

Semantics

- Suppose we receive a coded message containing reliable information about X defined on Ω.
- The actual code used is unknown, but we know that it was one of $c_1, \ldots, c_n$, and that each code had a chance p_i of being selected.
- Furthermore, we know that the meaning of the message is $X \in A_i \subseteq \Omega$ if code c_i was used.
- What do we know about X?
- For all $A \subseteq \Omega$, the probability that the message
  - means $X \in A$ is:
    $$m(A) = \sum_{i:A_i=A} p_i$$
  - implies $X \in A$ is: $bel(A) = \sum_{B \subseteq A} m(B)$
  - is consistent with $X \in A$ is:
    $$pl(A) = \sum_{B \cap A \neq \emptyset} m(B)$$

Special cases

- Logical mass function m_A such that $m_A(A) = 1$ (only one focal set), represents the evidence whose meaning is precisely and surely A ($\sim$ set).
- Total ignorance is represented by the logical mass function m_Ω, called vacuous mass function.
- If all focal sets are singletons, then the mass function is Bayesian ($\sim$ probability distribution).
- A mass function is simple if it has two focal sets: Ω and A for some $A \subset \Omega$. It has the following form

$$\begin{aligned} m(A) &= 1 - w(A), \\ m(\Omega) &= w(A) \end{aligned}$$

It is denoted by $A^{w(A)}$. For instance, $\{p, c\}^{0.1}$ in the faulty MacBook example.

Informational comparison

- Given two pieces of information $X \in A$ and $X \in B$, $X \in A$ is at least as informative as $X \in B$ if $A \subseteq B$.
- An extension of this ordering between sets to mass functions is the specialization ordering.
- Given two mass functions m_1 and m_2, m_1 is at least as informative as m_2, noted $m_1 \sqsubseteq_s m_2$, if m_1 can be obtained from m_2 by distributing each mass $m_2(B)$ to subsets of B, i.e.,

$$m_1(A) = \sum_B S(A, B) m_2(B), \quad \forall A,$$

  where $S(A, B) =$ proportion of $m_2(B)$ transferred to $A \subseteq B$.
- Properties
  - Extension of set inclusion: $m_A \sqsubseteq_s m_B \Leftrightarrow A \subseteq B$
  - Greatest element: vacuous mass function m_Ω
  - $m_1 \sqsubseteq_s m_2 \Rightarrow pl_1 \leq pl_2$

Outline

Outline

Reliable and independent messages

Derivation of Dempster's rule

- Let m_1 and m_2 be two mass functions induced by two randomly coded messages (also called sources), with:
  - $c_1, \ldots, c_n$, $p_1, \ldots, p_n$ and $A_1, \ldots, A_n$, the codes, their chances and their message meanings in the case of the first message,
  - $c'_1, \ldots, c'_m$, $p'_1, \ldots, p'_m$ and $B_1, \ldots, B_m$, the codes, their chances and their message meanings in the case of the second message.
- Assume the messages are independent, i.e., the two random choices of codes are independent: there is a chance $p_i p'_j$ that the pair (c_i, c'_j) of codes was chosen.
- Assume the messages are reliable, i.e., if the actual codes were c_i and c'_j, we can conclude that $X \in A_i \cap B_j$ for sure.

Reliable and independent messages

Derivation of Dempster's rule

- Let m_1 and m_2 be two mass functions induced by two randomly coded messages (also called sources), with:
  - $c_1, \ldots, c_n$, $p_1, \ldots, p_n$ and $A_1, \ldots, A_n$, the codes, their chances and their message meanings in the case of the first message,
  - $c'_1, \ldots, c'_m$, $p'_1, \ldots, p'_m$ and $B_1, \ldots, B_m$, the codes, their chances and their message meanings in the case of the second message.
- Assume the messages are independent, i.e., the two random choices of codes are independent: there is a chance $p_i p'_j$ that the pair (c_i, c'_j) of codes was chosen.
- Assume the messages are reliable, i.e., if the actual codes were c_i and c'_j, we can conclude that $X \in A_i \cap B_j$ for sure.
  - If $A_i \cap B_j = \emptyset$, then we know that (c_i, c'_j) could not be the pair of codes actually used.

$\rightarrow$ We must condition the chance distribution on the event $\{(c_i, c'_j) | 1 \leq i \leq n, 1 \leq j \leq m, A_i \cap B_j \neq \emptyset\}$

Reliable and independent messages

Derivation of Dempster's rule (continued)

- The probability of the overall message being $X \in C$, $C \neq \emptyset$ is

$$\begin{aligned} m(C) &= K \sum_{i,j: A_i \cap B_j = C} p_i p'_j \\ &= K \sum_{A \cap B = C} m_1(A) m_2(B) \end{aligned} \quad (1)$$

with $K = (1 - \kappa)^{-1}$ where κ is the degree of conflict defined as

$$\kappa = \sum_{i,j: A_i \cap B_j = \emptyset} p_i p'_j = \sum_{A \cap B = \emptyset} m_1(A) m_2(B)$$

Reliable and independent messages

Derivation of Dempster's rule (continued)

- The probability of the overall message being $X \in C$, $C \neq \emptyset$ is

$$\begin{aligned} m(C) &= K \sum_{i,j:A_i \cap B_j = C} p_i p'_j \\ &= K \sum_{A \cap B = C} m_1(A) m_2(B) \qquad (1) \end{aligned}$$

  with $K = (1-\kappa)^{-1}$ where κ is the degree of conflict defined as

$$\kappa = \sum_{i,j:A_i \cap B_j = \emptyset} p_i p'_j = \sum_{A \cap B = \emptyset} m_1(A) m_2(B)$$

- If $\kappa < 1$, the result of the combination of m_1 and m_2 by Dempster's rule $\oplus$ is the mass function $m_{1\oplus 2} = m_1 \oplus m_2$ called orthogonal sum and defined as $m_{1\oplus 2}(C) = m(C)$ given by (1).

Faulty MacBook example continued

- The technician's analysis is represented by the (simple) mass function $m_1 = \{p, c\}^{0.1}$
- Now, assume that a friend (Apple enthusiast) returns the mass function $m_2 = \{c, h\}^{0.2}$
- Assuming the pieces of evidence to be independent and reliable, we obtain:

$m_2 \setminus m_1$	$\{p, c\}$ 0.9	Ω 0.1
$\{c, h\}$ 0.8	$\{c, h\} \cap \{p, c\} = \{c\}$ 0.8*0.9=0.72	$\{c, h\}$ 0.08
Ω 0.2	$\{p, c\}$ 0.18	Ω 0.02

$$
\begin{aligned}
m_{1\oplus 2}(\{c\}) &= 0.72\\
m_{1\oplus 2}(\{c, h\}) &= 0.08\\
m_{1\oplus 2}(\{p, c\}) &= 0.18\\
m_{1\oplus 2}(\Omega) &= 0.02
\end{aligned}
$$

Justification in the Transferable Belief Model (TBM)

- The TBM allows $m(\emptyset) > 0$ (open-world assumption).
- Let $\odot$ be a combination rule for two mass functions.
- Assume $\odot$ must satisfy the following requirements:
  1. $m_1 \odot m_2$ is more informative than (a specialization of) m_1 and m_2
  2. $m_1 \odot m_2 = m_2 \odot m_1$ (commutativity)
  3. $(m_1 \odot m_2) \odot m_3 = m_1 \odot (m_2 \odot m_3)$ (associativity)
  4. $m \odot m_A$ is the least informative among the more informative mass functions m' than m such that $pl'(\overline{A}) = 0$.
- Then $m_1 \odot m_2 = m_1 \oplus^* m_2$, with $\oplus^*$ the unnormalized Dempster's rule (conjunctive rule)

$$
\begin{aligned}
m_{1\oplus^*2}(C) &= \sum_{A\cap B=C} m_1(A)m_2(B), \quad \forall C \subseteq \Omega, \\
m_{1\oplus 2}(C) &= \frac{m_{1\oplus^*2}(C)}{1-\kappa} = \frac{m_{1\oplus^*2}(C)}{1-m_{1\oplus^*2}(\emptyset)}, \quad \forall C \neq \emptyset.
\end{aligned}
$$

Outline

Properties of Dempster's rule

- Commutativity: $m_1 \oplus m_2 = m_2 \oplus m_1$
- Associativity: $(m_1 \oplus m_2) \oplus m_3 = m_1 \oplus (m_2 \oplus m_3)$
- Insensitivity to vacuous information (vacuous mass function as neutral element): $m \oplus m_\Omega = m$
- Generalization of set intersection: if $A \cap B \neq \emptyset$ then

$$m_A \oplus m_B = m_{A \cap B}$$

- Generalization of probabilistic conditioning: if m is a Bayesian mass function and m_A is a logical mass function, then

$$m \oplus m_A$$

  is a Bayesian mass function corresponding to the conditioning of m by A.

Expression using commonalities

- Commonality function $q : 2^\Omega \to [0, 1]$

$$q(A) = \sum_{B \supseteq A} m(B)$$

- Conversely,

$$m(A) = \sum_{B \supseteq A} (-1)^{|B \setminus A|} q(B)$$

- We have

$$q_{1 \oplus^* 2}(A) = q_1(A) \cdot q_2(A), \quad \forall A,$$

and

$$\begin{aligned} q_{1\oplus 2}(A) &= K \cdot q_{1 \oplus^* 2}(A), \quad \forall A \neq \emptyset, \\ q_{1\oplus 2}(\emptyset) &= 1 \end{aligned}$$

with $K = \left(\sum_{\emptyset \neq B \subseteq \Omega} (-1)^{|B|+1} q_{1\oplus^* 2}(B)\right)^{-1}$.

Simple and separable mass functions

- Let $A^{w_1(A)}$ and $A^{w_2(A)}$ be two simple mass functions.
- We have

$$A^{w_1(A)} \oplus A^{w_2(A)} = A^{w_1(A) \cdot w_2(A)}$$

Simple and separable mass functions

- Let $A^{w_1(A)}$ and $A^{w_2(A)}$ be two simple mass functions.
- We have

$$A^{w_1(A)} \oplus A^{w_2(A)} = A^{w_1(A) \cdot w_2(A)}$$

- A mass function is separable if it can be written as the $\oplus$ combination of simple mass functions

$$m = \bigoplus_{\emptyset \neq A \subset \Omega} A^{w(A)}$$

with $0 \leq w(A) \leq 1$ for all $A \subset \Omega, A \neq \emptyset$.

Canonical decomposition

- Let m be a non dogmatic mass function ($m(\Omega) > 0$).
- Weight function $w : 2^{\Omega} \setminus \{\emptyset, \Omega\} \to (0, +\infty)$

$$w(A) = \prod_{B \supseteq A} q(B)^{(-1)^{|B|-|A|+1}}$$

- m can be recovered from w by

$$m = \bigoplus_{\emptyset \neq A \subset \Omega} A^{w(A)}$$

(Some $A^{w(A)}$ in this decomposition may not be proper mass functions, since $w(A) > 1$.)

Expression using weights

- Let m_1 and m_2 be two non dogmatic mass functions.
- We have

$$m_1 \oplus m_2 = \bigoplus_{\emptyset \neq A \subset \Omega} A^{w_1(A) \cdot w_2(A)}$$

- The $\oplus$-decomposition is at play in various approaches (GBT, E-KNN, DS analysis of GLR classifiers, contextual corrections,...) and is the foundation to solutions to important problems (fusion of non independent sources, distributed fusion).

Outline

Exact computation

- Orthogonal sum $m_{1\oplus 2}$:
  - Mass-based approach;
  - Commonality-based approach (program).

Exact computation

- Orthogonal sum $m_{1\oplus 2}$:
  - Mass-based approach;
  - Commonality-based approach (▸▸ program).
- Computing times proportional to, respectively:
  - $|\Omega||\mathcal{F}(m_1)||\mathcal{F}(m_2)|$
  - $|\Omega|^2 2^{|\Omega|}$ (using the Fast Möbius Tranform to perform $m \leftrightarrow q$)
- Which approach to use ?
  - if $\forall m_i$, $|\mathcal{F}(m_i)| << 2^{|\Omega|}$, use the mass-based approach;
  - if $\exists m_i$, $|\mathcal{F}(m_i)| \sim 2^{|\Omega|}$, use the commonality-based approach.

Exact computation

- Orthogonal sum $m_{1\oplus 2}$:
  - Mass-based approach;
  - Commonality-based approach (program).
- Computing times proportional to, respectively:
  - $|\Omega||\mathcal{F}(m_1)||\mathcal{F}(m_2)|$
  - $|\Omega|^2 2^{|\Omega|}$ (using the Fast Möbius Tranform to perform $m \leftrightarrow q$)
- Which approach to use ?
  - if $\forall m_i$, $|\mathcal{F}(m_i)| << 2^{|\Omega|}$, use the mass-based approach;
  - if $\exists m_i$, $|\mathcal{F}(m_i)| \sim 2^{|\Omega|}$, use the commonality-based approach.
- In the worst case, exponential complexity with respect to $|\Omega|$.
- However, for practical applications (typically involving several mass functions), this is rarely an issue...

Exact computation

Particular forms of mass functions

- If each mass function is of the form ${\{\omega\}}^{w(\{\omega\})}$ or $\overline{\{\omega\}}^{w(\overline{\{\omega\}})}$, for some $\omega \in \Omega$, the complexity becomes linear.

Exact computation

Particular forms of mass functions

- If each mass function is of the form $\{\omega\}^{w(\{\omega\})}$ or $\overline{\{\omega\}}^{w(\overline{\{\omega\}})}$, for some $\omega \in \Omega$, the complexity becomes linear.
- If Ω is linearly ordered, and the focal sets of the mass functions are constrained to be intervals, the complexity becomes polynomial.
  - Example: duration (in days) of the repair of the faulty MacBook
  - $\Omega = \{1, \ldots, 30\}$
  - $A \subseteq \Omega$ is an interval if there exist elements a and b of Ω such that $A = \{\omega \in \Omega | a \leq \omega \leq b\}$.
  - Such A is denoted by $[a, b]$.
  - For instance, $A = \{12, 13, 14, 15, 16\} = [12, 16]$.

Exact computation

Particular forms of mass functions (continued)

- More generally, the complexity is polynomial if there is a partial ordering $\leq$ of Ω such that $(\Omega, \leq)$ is a lattice and the focal sets of the mass functions are constrained to be intervals of that lattice.
- Refresher on lattices:
  - Partial ordering $\leq$ on finite set L: a reflexive, antisymmetric and transitive relation on L.
  - $(L, \leq)$ is a partially ordered set (poset).
  - The poset $(L, \leq)$ is a lattice if for every $x, y \in L$, there is a unique greatest lower bound (denoted $x \wedge y$ and called meet) and a unique least upper bound (denoted $x \vee y$ and called join).
- Remark: The intersection of two intervals $[a, b]$ and $[c, d]$ of $(\Omega, \leq)$, required by Dempster's rule, is an interval

$$[a, b] \cap [c, d] = \begin{cases} [a \vee c, b \wedge d] & \text{if } a \vee c \leq b \wedge d, \\ \emptyset & \text{otherwise.} \end{cases}$$

Exact computation

Particular forms of mass functions (continued)

- This result makes it possible to tackle applications, such as multi-label classification, ensemble clustering, and preference aggregation, involving the manipulation of mass functions defined on very large Ω and which are thus intractable in the usual case.
- Indeed, in such applications, mass functions having only (lattice) interval focal sets are naturally encountered...

Exact computation

Examples of lattice intervals: Multi-label classification

- Instances belong to several classes at the same time.
- E.g., a song (instance) can generate several emotions (classes).
- Let $\Theta = \{\theta_1, \ldots, \theta_c\}$ be the set of classes.
- Class label X of an instance takes values in $\Omega = 2^\Theta$.
- Let ω_A be the element of Ω corresponding to $A \subseteq \Theta$

Exact computation

Examples of lattice intervals: Multi-label classification

- Instances belong to several classes at the same time.
- E.g., a song (instance) can generate several emotions (classes).
- Let $\Theta = \{\theta_1, \ldots, \theta_c\}$ be the set of classes.
- Class label X of an instance takes values in $\Omega = 2^\Theta$.
- Let ω_A be the element of Ω corresponding to $A \subseteq \Theta$
- Partial ordering on Ω: $\omega_A \leq \omega_B \Leftrightarrow A \subseteq B$, for all $A, B \subseteq \Theta$.
- Interval $[\omega_A, \omega_B]$ of lattice $(\Omega, \leq)$, for $A \subseteq B$, is an imprecise specification of X: it surely contains all elements of A and surely contains no element of $\overline{B}$.
- Natural way to express expert imprecise knowledge about the class label of a training instance.
- Predicting the class label of a test instance from such training data amounts, using the E-KNN classifier, to combining mass functions with interval focal sets.

Exact computation

Examples of lattice intervals: Ensemble clustering

- Clustering a set Θ of n objects = finding a partition of Θ.
- Let Ω be set of all partitions of Θ.
- The "true" partition X of the objects takes values in Ω.
- Partial ordering on Ω: $\omega \leq \omega'$ (ω is finer than ω'), for all $\omega, \omega' \in \Omega$, if the clusters of ω can be obtained by splitting those of ω'.

Exact computation

Examples of lattice intervals: Ensemble clustering

- Clustering a set Θ of n objects = finding a partition of Θ.
- Let Ω be set of all partitions of Θ.
- The "true" partition X of the objects takes values in Ω.
- Partial ordering on Ω: $\omega \leq \omega'$ (ω is finer than ω'), for all $\omega, \omega' \in \Omega$, if the clusters of ω can be obtained by splitting those of ω'.
- Interval $[\underline{\omega}, \overline{\omega}]$ of lattice $(\Omega, \leq)$, for $\underline{\omega} \leq \overline{\omega}$, is an imprecise specification of X: it is coarser than $\underline{\omega}$ and finer than $\overline{\omega}$.
- For instance, "the objets of a set $A \subseteq \Theta$ belong to the same cluster" can be represented by the interval $[\omega_A, \omega_\Theta]$, where ω_B is the partition where only the objects in B are clustered together.
- Natural way to interpret the output of a clustering algorithm.
- Predicting the true partition from an ensemble of such clustering algorithms, while accounting for their validity, amounts to combining mass functions with interval focal sets.

Exact computation

Decision making

- The goal is often to make decisions.
- A usual decision rule is to select the singleton $\{\omega\}$ of Ω with the largest plausibility or, equivalently (since $pl(\{\omega\}) = q(\{\omega\})$), with the largest commonality.
- The complexity is linear, thanks to the property

$$q_{1\oplus 2}(\{\omega\}) = K \cdot q_1(\{\omega\}) \cdot q_2(\{\omega\}), \quad \forall \omega \in \Omega.$$

Approximate computation

- Approximate computation when the exact computation is not possible.
- Stochastic approximation procedures:
  - (Approximate) Combined belief for some $A \subset \Omega$ can be computed by Monte Carlo algorithms in time linear in $|\Omega|$;
  - Not feasible when one is interested in the whole combined belief function.
- Deterministic approximation procedures: provide upper and lower bounds on combined belief
  - Mass-based approach;
  - Commonality-based approach.

Approximation for the mass-based approach

- Complexity depends on the number of focal sets $\rightarrow$ approximate mass functions by simpler ones with fewer focal sets.
- Simplest method: Summarization algorithm.
- Let $F_1, \ldots, F_r$ be the focal sets of a mass function m ranked by decreasing mass, i.e., $m(F_1) \geq m(F_2) \geq \ldots \geq m(F_r)$.
- Let k be the maximum allowed number of focal sets.
- If $r > k$, the $r - k + 1$ focal sets $F_k, \ldots, F_r$ are replaced by their union, and m is approximated by the mass function $\varphi^+(m)$ defined as

$$\begin{aligned}
\varphi^+(m)\,(F_i) &= m(F_i), \quad i = 1, \ldots, k-1, \\
\varphi^+(m)\left(\bigcup_{i=k}^{r} F_i\right) &= \sum_{i=k}^{r} m(F_i).
\end{aligned}$$

- For short, we say that $F_k, \ldots, F_r$ are "aggregated".

Approximation for the mass-based approach

- We have $m \sqsubseteq_s \varphi^+(m)$: it is called an outer approximation of m.

Proposition (Monotonicity of $\oplus^*$ with respect to $\sqsubseteq_s$)

$$m \sqsubseteq_s m' \Rightarrow m \oplus^* m_0 \sqsubseteq_s m' \oplus^* m_0, \quad \forall m_0$$

- From these properties, we have

$$m_{\oplus^*} \sqsubseteq_s m^+$$

  with

$$\begin{aligned} m_{\oplus^*} &:= m_1 \oplus^* \ldots \oplus^* m_n, \\ m^+ &:= \varphi^+(\varphi^+(\ldots \varphi^+(\varphi^+(m_1 \oplus^* m_2) \oplus^* m_3) \oplus^* \ldots m_{n-1}) \oplus^* m_n). \end{aligned}$$

- m^+ is an outer approximation of the conjunctive combination of mass functions $m_1, \ldots, m_n$.
- The combinatorial explosion of the combination is avoided.

Approximation for the mass-based approach

- In the summarization procedure of a mass function m, if we replace the focal sets $F_k, \ldots, F_r$ by their intersection rather than their union, we get another approximation $\varphi^-(m)$ of m.
- We have $\varphi^-(m) \sqsubseteq_s m$: it is called an inner approximation of m.
- Furthermore,

$$m^- \sqsubseteq_s m_{\oplus^*} \sqsubseteq_s m^+$$

with

$$m^- := \varphi^-(\varphi^-(\ldots\varphi^-(\varphi^-(m_1 \oplus^* m_2) \oplus^* m_3) \oplus^* \ldots m_{n-1}) \oplus^* m_n).$$

- We have

$$pl^- \leq pl_{\oplus^*} \leq pl^+$$

- Bounds on $bel_{\oplus^*}$ can also be obtained.

Approximation for the mass-based approach

- Let $m_{\oplus}$ denote the orthogonal sum of mass functions $m_1, \ldots, m_n$.
- We have

$$pl_{\oplus}(A) = \frac{pl_{\oplus^*}(A)}{pl_{\oplus^*}(\Omega)}, \quad \forall A \subseteq \Omega$$

- Inner and outer approximations m^- and m^+ of $m_{\oplus^*}$ allow thus to obtain lower and upper bounds on $pl_{\oplus}$ (and also on $bel_{\oplus}$):

$$\frac{pl^-(A)}{pl^+(\Omega)} \leq pl_{\oplus}(A) \leq \frac{pl^+(A)}{pl^-(\Omega)}$$

Approximation for the mass-based approach

- Let m be a mass function.
- The summarization algorithm produces a less informative approximation $\varphi^+(m)$ of m (we have $m \sqsubseteq_s \varphi^+(m)$).
- It does so by aggregating "unimportant" focal sets (those with lowest masses).
- They are unimportant in the sense that they will not incur too much information loss.
- When approximating m, we indeed want to lose as less as possible of its informative content.

Approximation for the mass-based approach

- How much information is lost by $\varphi^+(m)$?
- (Generalized) Cardinality of a mass function m:

$$|m| := \sum_A m(A)|A|,$$

  the greater the cardinality of m, the less informative m is.
- We have

$$m_1 \sqsubseteq_s m_2 \Rightarrow |m_1| \leq |m_2|.$$

- Hence, a measure of the information lost if we replace m by $\varphi^+(m)$ can be

$$\Delta(\varphi^+(m), m) := |\varphi^+(m)| - |m|.$$

Approximation for the mass-based approach

- Furthermore, we can remark that the summarization algorithm involves a specific partition $\mathcal{P} = \{I_1, \ldots, I_k\}$ of $\mathcal{F}(m) = \{F_1, \ldots, F_r\}$ with

$$\begin{aligned} I_i &= \{F_i\}, \quad i = 1, \ldots, k-1, \\ I_k &= \{F_k, \ldots, F_r\}. \end{aligned}$$

- The mass function $\varphi^+(m)$ can then be rewritten simply as

$$\varphi^+(m)\left(\bigcup_{F \in I} F\right) = \sum_{F \in I} m(F), \quad \forall I \in \mathcal{P}.$$

Approximation for the mass-based approach

- Other partitions of $\mathcal{F}(m)$ of size k exist!
- Let $\varphi^+_{\mathcal{P}}(m)$ be the outer approximation of m obtained for some partition $\mathcal{P}$ of $\mathcal{F}(m)$ using the equation on the previous slide.

→ Find the best outer approximation $\varphi^+_{\mathcal{P}^*}(m)$ of m by searching a partition $\mathcal{P}^*$ minimizing the information loss:

$$\Delta(\varphi^+_{\mathcal{P}^*}(m), m) := \min_{\mathcal{P} \in \mathcal{P}_k} \Delta(\varphi^+_{\mathcal{P}}(m), m),$$

with $\mathcal{P}_k$ the set of all partitions of $\mathcal{F}(m)$ of size k.

Approximation for the mass-based approach

- An exhaustive search in $\mathcal{P}_k$ is in general not possible, as $|\mathcal{P}_k|$ rapidly explodes, even for small values of r.
- We need to resort to heuristic search techniques.
- A hierarchical clustering algorithm has been proposed for that purpose: pairs of focal sets are grouped sequentially (at each step, the two "closest" focal sets are aggregated), until the desired number k of focal sets has been reached.
- Time proportional to r^3.

Approximation for the mass-based approach

- The algorithm relies on the "distance" $\delta^+(F_i, F_j)$ between any pair (F_i, F_j) of focal sets of a mass function m defined as

$$\delta^+(F_i, F_j) := \Delta(\varphi^+_{\mathcal{P}_{i,j}}(m), m),$$

with $\mathcal{P}_{i,j}$ the partition of $\mathcal{F}(m)$ of size $|\mathcal{F}(m)| - 1$ such that

$$\begin{aligned} \exists I \in \mathcal{P}_{i,j}, I &= \{F_i, F_j\}, \\ \forall I' \in \mathcal{P}, I' \neq I, I' &= \{F\}, F \in \mathcal{F}(m), F \neq F_i, F_j. \end{aligned}$$

- $\delta^+(F_i, F_j)$ evaluates how much information is lost, with respect to a given mass function m, if its focal sets F_i and F_j are aggregated.

Approximation for the mass-based approach

- Let $\varphi^{+}_{\hat{\mathcal{P}}}(m)$ denote the outer approximation of a mass function m obtained using this hierarchical clustering-based approach (outer clustering approximation for short)
  - There is no guarantee that it yields the same (lowest) information loss as $\varphi^{+}_{\mathcal{P}^*}(m)$.
  - It has been shown empirically to yield better results than $\varphi^{+}(m)$.
- Much as the summarization procedure can be adapted to obtain an inner approximation $\varphi^{-}(m)$ of m, this more complex approximation procedure can be adapted to find an inner (clustering) approximation $\varphi^{-}_{\hat{\mathcal{P}}}(m)$ of m.
- Remark: contrarily to the summarization procedure, approximations $\varphi^{+}_{\hat{\mathcal{P}}}(m)$ and $\varphi^{-}_{\hat{\mathcal{P}}}(m)$ of m rely in general on different partitions $\hat{\mathcal{P}}$ of $\mathcal{F}(m)$.

Approximation for the mass-based approach

- Similarly as for the summarization procedure, we can obtain inner and outer approximations of the conjunctive combination of n mass functions:

$$m^-_{\hat{\mathcal{P}}} \sqsubseteq_s m_{\oplus^*} \sqsubseteq_s m^+_{\hat{\mathcal{P}}}$$

  with

$$\begin{aligned} m^-_{\hat{\mathcal{P}}} &:= \varphi^-_{\hat{\mathcal{P}}}(\varphi^-_{\hat{\mathcal{P}}}(\ldots \varphi^-_{\hat{\mathcal{P}}}(\varphi^-_{\hat{\mathcal{P}}}(m_1 \oplus^* m_2) \oplus^* m_3) \oplus^* \ldots m_{n-1}) \oplus^* m_n), \\ m^+_{\hat{\mathcal{P}}} &:= \varphi^+_{\hat{\mathcal{P}}}(\varphi^+_{\hat{\mathcal{P}}}(\ldots \varphi^+_{\hat{\mathcal{P}}}(\varphi^+_{\hat{\mathcal{P}}}(m_1 \oplus^* m_2) \oplus^* m_3) \oplus^* \ldots m_{n-1}) \oplus^* m_n). \end{aligned}$$

- They induce bounds on $pl_\oplus$ (and on $bel_\oplus$):

$$\frac{pl^-_{\hat{\mathcal{P}}}(A)}{pl^+_{\hat{\mathcal{P}}}(\Omega)} \leq pl_\oplus(A) \leq \frac{pl^+_{\hat{\mathcal{P}}}(A)}{pl^-_{\hat{\mathcal{P}}}(\Omega)}.$$

Approximation for the commonality-based approach

- Complexity depends on $|\Omega|$ → approximate Ω by a simpler (coarser) frame Θ with fewer elements.
- Algorithm for the combination of n mass functions $m_1, \ldots, m_n$:
  1. Search, using a hierarchical clustering procedure, for a partition (coarsening) Θ of Ω of size c, minimizing information loss defined as
  $$\sum_{i=1}^{n} \Delta(\rho(\overline{m}_i^{\Theta}), m_i)$$
  with $\rho(\overline{m}_i^{\Theta})$ the outer approximation of m_i obtained by carrying m_i to Θ (restriction $\overline{m}_i^{\Theta}$) and carrying it back to Ω (vacuous extension $\rho(\cdot)$)
  2. Using the commonality-based approach, combine the mass functions in the coarsened frame, i.e., compute $\overline{m}^{\Theta} := \oplus_{i=1}^{*n} \overline{m}_i^{\Theta}$
  3. Carry the result to Ω, i.e., compute $\overline{m} := \rho(\overline{m}^{\Theta})$
- $\overline{m}$ is an outer approximation of $m_{\oplus^*}$.

Approximation for the commonality-based approach

- Computing time proportional to $\max(|\Omega|^3, nc^2 2^c)$.
- Algorithm can be adapted to obtain an inner approximation $\underline{m}$ of $m_{\oplus^*}$.
- We have thus

$$\underline{m} \sqsubseteq m_{\oplus^*} \sqsubseteq \overline{m}$$

- This algorithm thus also yields lower and upper bounds for $bel_\oplus$ and for $pl_\oplus$.

Outline

Zadeh's example

- Let $X \in \Omega = \{a, b, c\}$ and two experts providing mass functions m_1 and m_2 about X:

$$m_1(\{a\}) = 0.99, m_1(\{b\}) = 0.01, m_1(\{c\}) = 0$$

$$m_2(\{a\}) = 0, m_2(\{b\}) = 0.01, m_2(\{c\}) = 0.99$$

- We have $m_{1 \oplus 2}(\{b\}) = 1$.
- As both experts considered b to be very unlikely, some authors claim this result to be counterintuitive, and use it to question Dempster's rule.
- However, if you accept the assumptions underlying Dempster's rule, then this is the only reasonable conclusion: expert 1 tells that c is impossible, and expert 2 tells that a is impossible, hence b is the only remaining possibility.
- The question is not whether Dempster's rule produces sound results or not, but rather whether its underlying assumptions hold.

$\rightarrow$ We need a way to assess their validity.

Degree of conflict

- $\kappa = \sum_{A \cap B = \emptyset} m_1(A) m_2(B)$ has been shown to satisfy a set of desirable properties for a conflict measure, i.e., a measure of the inconsistency resulting from making the assumptions that the received messages are reliable and independent.
- In Zadeh's example, we have $\kappa = 0.9999$, which suggests that these assumptions may not be valid.

Degree of conflict

- $\kappa = \sum_{A \cap B = \emptyset} m_1(A) m_2(B)$ has been shown to satisfy a set of desirable properties for a conflict measure, i.e., a measure of the inconsistency resulting from making the assumptions that the received messages are reliable and independent.
- In Zadeh's example, we have $\kappa = 0.9999$, which suggests that these assumptions may not be valid.

$\rightarrow$ We need alternative rules, corresponding to other assumptions.

Outline

Beyond Dempster's rule

- Let m_1 and m_2 be two mass functions induced by two randomly coded messages.
- The assumptions leading to Dempster's rule are that the
  - Messages are independent: the probability that the messages mean A and B, respectively, is $m_1(A)m_2(B)$ (at least before making an assumption about their reliability)
  - Messages are reliable: if the messages mean A and B, respectively, then the overall message is $C = A \cap B$.
- Rules corresponding to other assumptions about
  1. the reliability (still assuming independence)
  2. the dependence

Outline

Disjunctive rule

- Let us now suppose only that at least one of the two messages inducing m_1 and m_2 is reliable, i.e., if the actual codes were c_i and c'_j, we can only conclude that $X \in A_i \cup B_j$ for sure.
- We obtain the disjunctive rule:

$$m_{1 \bigcirc\!\!\!\!\cup 2}(C) = \sum_{A \cup B = C} m_1(A) m_2(B), \quad \forall C \subseteq \Omega,$$

- It satisfies similar properties as Dempster's rule, in particular commutativity, associativity and expression based on pointwise product of belief functions.
- $\cup$ can be replaced by any other binary Boolean connective $\otimes$ (may require normalization using $\kappa_{\otimes} = \sum_{A \otimes B = \emptyset} m_1(A) m_2(B)$) and an interpretation provided (see lecture at BFTA'2019).

Discount and combine (DC)

Independent reliabilities

- Suppose a given message inducing a mass function m is reliable with probability α (and unreliable with probability $1-\alpha$).
- Then the probability that the message means
  - $X \in A$ is ${}^{\alpha}m(A) = \alpha m(A)$, for $A \subset \Omega$,
  - nothing is ${}^{\alpha}m(\Omega) = \alpha m(\Omega) + 1 - \alpha$.
- ${}^{\alpha}m$ is the mass function resulting from the discounting of m with discount rate $1-\alpha$.

Discount and combine (DC)

Independent reliabilities

- Suppose a given message inducing a mass function m is reliable with probability α (and unreliable with probability $1-\alpha$).
- Then the probability that the message means
  - $X \in A$ is ${}^{\alpha}m(A) = \alpha m(A)$, for $A \subset \Omega$,
  - nothing is ${}^{\alpha}m(\Omega) = \alpha m(\Omega) + 1 - \alpha$.
- ${}^{\alpha}m$ is the mass function resulting from the discounting of m with discount rate $1-\alpha$.
- Suppose the two messages inducing m_1 and m_2 have independent probabilities α_1 and α_2 of being reliable, then what we know about X is represented by the mass function m_{DC} such that

$$m_{DC} = {}^{\alpha_1}m_1 \oplus {}^{\alpha_2}m_2$$

- Combination method often used in practice and that can be extended to refined form of knowledge about the behaviour of each source (using, e.g., contextual corrections).

Weighted average (WA)

Dependent reliabilities

- A close, yet different, assumption that can be made about the reliability of the two messages is:
  - the first message is reliable and the second is not reliable with probability α_1,
  - the first message is not reliable and the second is reliable with probability $\alpha_2 = 1 - \alpha_1$.
- In this case, our knowledge about X is represented by the mass function m_{WA} such that

$$m_{WA} = \alpha_1 m_1 + \alpha_2 m_2$$

⚠ In general, $\alpha_1 m_1 + \alpha_2 m_2 \neq {}^{\alpha_1} m_1 \oplus {}^{\alpha_2} m_2$

- Neither commutative nor associative combination methods.

Dubois and Prade's rule

- Dubois and Prade's rule $\circledstar$:

$$m_{1\circledstar 2}(C) = \sum_{A\cap B=C} m_1(A)m_2(B) + \sum_{A\cap B=\emptyset, A\cup B=C} m_1(A)m_2(B), \forall C \neq \emptyset,$$

  and $m_{1\circledstar 2}(\emptyset) = 0$.
- Properties:
  - If $\kappa = 0$, then $m_{1\circledstar 2} = m_{1\oplus 2}$
  - If $\kappa = 1$, then $m_{1\circledstar 2} = m_{1\bigcirc\!\!\!\!\cup\, 2}$
  - Commutativity, insensitivity to vacuous information, as well as 6 other basic fusion properties
  - Not associative

Outline

Handling dependence

- Let m_1 and m_2 be two mass functions induced by two randomly coded messages.
- Dependence between the messages induces a joint probability $m_{12}(A, B)$ that they mean A and B, respectively.
- $m_{12}(\cdot, \cdot)$ is a joint mass function s.t. its marginals are m_1 and m_2:

$$m_1(A) = \sum_B m_{12}(A, B), \quad \forall A,$$

  and likewise for m_2.
- Independence: $m_{12}(A, B) = m_1(A)m_2(B)$, for all A and B.
- All preceding combination rules can be extended to a known dependence between the messages by replacing $m_1(A)m_2(B)$ in their definitions by $m_{12}(A, B)$.

Handling dependence

- Let m_1 and m_2 be two mass functions induced by two randomly coded messages.
- Dependence between the messages induces a joint probability $m_{12}(A, B)$ that they mean A and B, respectively.
- $m_{12}(\cdot, \cdot)$ is a joint mass function s.t. its marginals are m_1 and m_2:

$$m_1(A) \quad = \quad \sum_B m_{12}(A, B), \quad \forall A,$$

  and likewise for m_2.
- Independence: $m_{12}(A, B) = m_1(A)m_2(B)$, for all A and B.
- All preceding combination rules can be extended to a known dependence between the messages by replacing $m_1(A)m_2(B)$ in their definitions by $m_{12}(A, B)$.
- However, in practice, describing the nature of the dependence is generally difficult.

$\rightarrow$ We thus need a different approach to handle non independent pieces of evidence.

Least commitment principle

Least commitment principle (LCP)

When several belief functions are compatible with a set of constraints, the least informative according to some informational ordering should be selected (if it exists).

- To become operational, the LCP needs a means (informational ordering) to compare the informative content of belief functions.

Informational orderings

- We have already seen an informational ordering: $\sqsubseteq_s$.
- Other proposals exist to establish whether a mass function m_1 is more informative than another mass function m_2 (noted $m_1 \sqsubseteq_x m_2$ for some informational ordering x):
- *w* (weight) ordering: $m_1 \sqsubseteq_w m_2$ if $m_1 = m_2 \oplus m$ for some separable mass function m, which is equivalent to

$$w_1(A) \leq w_2(A), \quad \forall A.$$

Cautious merging

- Two reliable but non independent sources providing mass functions m_1 and m_2.
- The result m of their combination should be more informative than each one of them

$$m \in \mathcal{S}_x(m_1) \cap \mathcal{S}_x(m_2),$$

with $\mathcal{S}_x(m)$ the set of mass functions m' such that $m' \sqsubseteq_x m$.

- The LCP dictates that the x-least informative element in $\mathcal{S}_x(m_1) \cap \mathcal{S}_x(m_2)$ should be selected, if it exists
  - Neither existence nor uniqueness can be guaranteed with the s-ordering.
  - Existence and uniqueness can be guaranteed with the w-ordering (if m_1 and m_2 are nondogmatic): it is defined by

$$m_{1 \textcircled{\wedge} 2} = m_1 \textcircled{\wedge} m_2 = \bigoplus_{\emptyset \neq A \subset \Omega} A^{\min(w_1(A), w_2(A))}.$$

Rule $\textcircled{\wedge}$ is called the cautious rule.

Properties of the cautious rule

- Commutativity
- Associativity
- Monotonicity with respect to $\sqsubseteq_w$:

$$m_1 \sqsubseteq_w m_2 \Rightarrow m_1 \owedge m_3 \sqsubseteq_w m_2 \owedge m_3$$

- Idempotence: $m \owedge m = m$
- Distributivity of $\oplus$ with respect to $\owedge$:

$$(m_1 \oplus m_2) \owedge (m_1 \oplus m_3) = m_1 \oplus (m_2 \owedge m_3)$$

- Sensitivity to vacuous information

Bold rule

- A disjunctive counterpart of $\owedge$, called the bold rule and noted $\ovee$, can also be obtained.
- It corresponds to non independent pieces of evidence, at least one of which is reliable.
- It amounts to taking the minimum of *disjunctive* weights, which are a disjunctive counterpart to the weights we have seen so far.

Infinite families of conjunctive and disjunctive rules

- Infinite families of conjunctive and disjunctive combination rules, of which $\oplus$, $\owedge$, $\ocup$ and $\ovee$ belong, can be defined by replacing the product/minimum in their definitions by uninorms/triangular norms defined on $(0, +\infty]$.
- Example : t-norm based conjunctive rules (t-rules)

$$m_1 \owedge_{\mathcal{T}} m_2 = \bigoplus_{\emptyset \neq A \subset \Omega} A^{\mathcal{T}(w_1(A), w_2(A))}$$

  with $\mathcal{T}$ a triangular norm on $(0, +\infty]$ ($\owedge$ recovered for $\mathcal{T} = \min$)
- These conjunctive rules are commutative, associative, monotonic with respect to $\sqsubseteq_w$. The uninorm-based ones are in addition insensitive to vacuous information.
- Parameterized versions of these rules can be obtained, which can be useful to adapt the combination to the degree of dependence between the sources.

Outline

Fusion of n mass functions

- We receive mass functions $m_1, \ldots, m_n$

$$m = f(m_1, \ldots, m_n)$$

- All the preceding combination rules can be extended to n mass functions.

Associative rules

- The n-ary extension $\odot(m_1, \ldots, m_n)$ of the binary combination rule $\odot \in \{\oplus, \textcircled{\cup}, \textcircled{\wedge}, \textcircled{\vee}\}$ satisfies

$$\odot(m_1, \ldots, m_n) = ((...((m_1 \odot m_2) \odot m_3) \odot ...m_{n-1}) \odot m_n)$$

- Associated assumptions for the n-ary extension of
  - $\oplus$: the pieces of evidence are independent and reliable.
  - $\textcircled{\cup}$: the pieces of evidence are independent and at least one of them is reliable (it is not known which one)
  - $\textcircled{\wedge}$: the pieces of evidence are non independent (their actual dependence is unknown) and reliable
  - $\textcircled{\vee}$: the pieces of evidence are non independent and at least one of them is reliable

Non-associative rules

- Dubois and Prade's rule:

$$\circledast(m_1,\ldots,m_n)(B) = \sum_{B=h(A_1,\ldots,A_n)} m_1(A_1)\cdot\ldots\cdot m_n(A_n)$$

  with

$$h(A_1,\ldots,A_n) = \cup_{\mathcal{M}\in MCS(A_1,\ldots,A_n)} \cap_{A_i\in\mathcal{M}} A_i$$

  where $MCS(A_1,\ldots,A_n)$ is the set of maximal consistent subsets of $\{A_1,\ldots,A_n\}$.
- Discount and combine: $m_{DC} = \bigoplus_{i=1}^{n} {}^{\alpha_i}m_i$, with $\alpha_i \in [0,1]$
- Weighted average: $m_{WA} = \sum_{i=1}^{n} \alpha_i m_i$, with $\alpha_i \in [0,1]$ and $\sum_{i=1}^{n} \alpha_i = 1$.

q-relaxation rule

- The q-relaxation rule $\otimes_r$ has the same definition as rule $\circledast$ except that $MCS(A_1, \ldots, A_n)$ is replaced by $Relax_r(A_1, \ldots, A_n)$, which is the set of subsets of $\{A_1, \ldots, A_n\}$ of sizes r.
- Associated assumption: r out of the n pieces of evidence are reliable.
- Normalization (using $\kappa_r = \sum_{f(A_1,\ldots,A_n)=\emptyset} m_1(A)m_2(B)$) may be required.
- Properties:
  - Generalization of the (unnormalized) Dempster ($r = n$) and disjunctive ($r = 1$) rules.
  - Generalization of the q-relaxation technique from interval analysis ($r = n - q$), which is designed to implement some form of robustness to outliers.
  - Commutativity.

Outline

Which f to use?

$$m = f(m_1, \ldots, m_n)$$

Which f to use?

$$m = f(m_1, \ldots, m_n)$$

- Ideal situation: reliability and independence of the m_i are clear $\rightarrow$ use $\oplus$.

Which f to use?

$$m = f(m_1, \ldots, m_n)$$

- Ideal situation: reliability and independence of the m_i are clear $\rightarrow$ use $\oplus$.
- What if the reliability is unknown or the independence cannot be assumed?
  - No labelled data to assess the effectiveness (performance) of a particular f:
    - Use a robust (and principled) rule;
    - Select a rule according to a consistency-informativeness trade-off.
  - Labelled data available: learn the “best” rule.

Outline

Robust rules

- Independent but unknown reliability: $\circledast$ or, if n large, RANSAC-based rule $\odot$:
  1. Estimate which m_i should be considered reliable ($i \in \mathcal{R}$):
     1. $\mathcal{I}^\nu$: set of N random subsets $\mathcal{I} \subset \{1, \ldots, n\}$ of size ν
     2. $\forall \mathcal{I} \in \mathcal{I}^\nu, \mathcal{I}_\tau := \{i | 1 \leq i \leq n,\ \kappa(m_i, m_\mathcal{I}) \leq \tau\}$ with $m_\mathcal{I} = \oplus_{i \in \mathcal{I}} m_i$
     3. $\mathcal{R} = \text{argmax}_{\{\mathcal{I}_\tau |\ \mathcal{I} \in \mathcal{I}^\nu\}} |\mathcal{I}_\tau|$
  2. Return $\oplus$-combination of the reliable ones:
     $\odot(m_1, \ldots, m_n) = \oplus_{i \in \mathcal{R}} m_i$
- Not independent but reliable: $\owedge$
- Not independent and unknown reliability: $\ovee$

Robust rules

- Independent but unknown reliability: $\circledast$ or, if n large, RANSAC-based rule $\circledcirc$:
  1. Estimate which m_i should be considered reliable ($i \in \mathcal{R}$):
     1. $\mathcal{I}^\nu$: set of N random subsets $\mathcal{I} \subset \{1, \ldots, n\}$ of size ν
     2. $\forall \mathcal{I} \in \mathcal{I}^\nu, \mathcal{I}_\tau := \{i | 1 \leq i \leq n,\ \kappa(m_i, m_\mathcal{I}) \leq \tau\}$ with $m_\mathcal{I} = \oplus_{i \in \mathcal{I}} m_i$
     3. $\mathcal{R} = \operatorname{argmax}_{\{\mathcal{I}_\tau |\ \mathcal{I} \in \mathcal{I}^\nu\}} |\mathcal{I}_\tau|$
  2. Return $\oplus$-combination of the reliable ones: $\circledcirc(m_1, \ldots, m_n) = \oplus_{i \in \mathcal{R}} m_i$
- Not independent but reliable: $\owedge$
- Not independent and unknown reliability: $\ovee$

	rel	$\neg$*rel*
ind	$\oplus$	$\circledast$, $\circledcirc$
$\neg$*ind*	$\owedge$	$\ovee$

Consistency-informativeness trade-off (CIT)

Basic idea

- 3 sources about $X \in \Omega = \{a, b, c\}$ supplying A_1, A_2 and A_3 s.t.

$$A_1 = \{a\}, A_2 = \{a, b\}, A_3 = \{b, c\}$$

- Assumption r_1 = "all sources are reliable" yields

$$X \in C_1 = A_1 \cap A_2 \cap A_3 = \emptyset$$

  i.e. an inconsistent result, and thus cannot hold.

- In contrast, the assumption r_3 = "at least one of the sources is reliable" yields

$$X \in C_3 = A_1 \cup A_2 \cup A_3 = \mathcal{X}$$

  and is thus plausible (it does not yield a contradiction). However, it is useless as it is not informative at all.

Consistency-informativeness trade-off (CIT)

Basic idea (continued)

- The intermediate assumption $r_2 =$ "at least two of the sources are reliable" yields

$$X \in C_2 = (A_1 \cap A_2) \cup (A_1 \cap A_3) \cup (A_2 \cap A_3) = \{a, b\}$$

  r_2 is plausible (the result is consistent) and informative (or, at least, more informative than r_3).

Consistency-informativeness trade-off (CIT)

Basic idea (continued)

- The intermediate assumption $r_2 =$ "at least two of the sources are reliable" yields

$$X \in C_2 = (A_1 \cap A_2) \cup (A_1 \cap A_3) \cup (A_2 \cap A_3) = \{a, b\}$$

  r_2 is plausible (the result is consistent) and informative (or, at least, more informative than r_3).
- r_2 is preferable, but for other (A_1, A_2, A_3), it could be r_1 or r_3 due to

$$C_1 \subseteq C_2 \subseteq C_3, \quad \forall A_1, A_2, A_3.$$

  r_{j+1} will always yield a result that is on the one hand at least as consistent as that of r_j, but also on the other hand at most as specific as that of r_j.

$\rightarrow$ Consistency and informativeness are antagonist goals

- Sensible strategy for a given (A_1, A_2, A_3): test iteratively each r_j and select the first one which yields a consistent result (it will then be the most informative and consistent possible result).

Extension to rule selection

Strategy

1. Consider a set of rules $\mathcal{F} = \{f^1, ..., f^J\}$ such that:
   - $m^j \sqsubseteq m^{j+1}$
   - $\phi(m^j) \leq \phi(m^{j+1})$

   with $m^j = f^j(m_1, \ldots, m_n)$, $\phi(m)$ a measure of the consistency of m and $\sqsubseteq$ an informational ordering between mass functions;
2. Test iteratively each f^j until $\phi(m^j) \geq \delta$.

Extension to rule selection

Strategy

1. Consider a set of rules $\mathcal{F} = \{f^1, ..., f^J\}$ such that:
   - $m^j \sqsubseteq m^{j+1}$
   - $\phi(m^j) \leq \phi(m^{j+1})$

   with $m^j = f^j(m_1, \ldots, m_n)$, $\phi(m)$ a measure of the consistency of m and $\sqsubseteq$ an informational ordering between mass functions;
2. Test iteratively each f^j until $\phi(m^j) \geq \delta$.

- Approach originally studied in the context of the TBM ($m(\emptyset) \geq 0$ allowed), with $\sqsubseteq_s$ and $\phi_\pi(m) = \max_{\omega\in\Omega} pl(\{\omega\})$.

Proposition

$$m \sqsubseteq_s m' \Rightarrow \phi_\pi(m) \leq \phi_\pi(m'), \forall m, m'$$

→ Consistency and informativeness are at odds !

Choosing $\mathcal{F}$

- A similar proposition holds for:
  - $\sqsubseteq_{w^*}$: informational ordering based on the "TBM" weight function w^*, which assigns a weight $\frac{\prod_{|B|\notin 2\mathbb{N}} q(B)}{\prod_{|B|\in 2\mathbb{N}} q(B)}$ to $\emptyset$ and such that $w^*(A) = w(A)$ for all $A \subset \Omega$;
  - $\sqsubseteq_v$: informational ordering based on the disjunctive weight function v.
- Thanks to these propositions, it suffices to choose $\mathcal{F}$ such that $m^j \sqsubseteq_x m^{j+1}$, with $x \in \{s, w^*, v\}$, in order to have $\phi_\pi(m^j) \leq \phi_\pi(m^{j+1})$ (required in Step 1 of the strategy).
- There exist in particular such $\mathcal{F}$ suited to the different unknown reliability/dependence situations.

CIT for selecting reliability or dependence

- Independent but unknown reliability: use CIT approach with
  - $f^j = \otimes^*_{r_j}$ with $r_j = n - j + 1$ and $\otimes^*_r$ the TBM (unnormalized) q-relaxation rule
  - $f^j(m_1, \ldots, m_n) = \oplus^{*n}_{i=1} {}^{\alpha^j_i} m_i$, with $\alpha^j_i \geq \alpha^{j+1}_i$ (independent reliabilities)
- Not independent but reliable: use $f^j = \owedge^*_{\mathcal{T}_j}$ with $\mathcal{T}_j \leq \mathcal{T}_{j+1}$ and $\owedge^*_{\mathcal{T}}$ the TBM conjunctive t-rule (relying on w^* rather than w) for some t-norm $\mathcal{T}$.
- Not independent and unknown reliability: use $f^j = \ovee_{\mathcal{T}_j}$ with $\mathcal{T}_j \geq \mathcal{T}_{j+1}$ and $\ovee_{\mathcal{T}}$ the disjunctive t-rule for some t-norm $\mathcal{T}$.

CIT application, independent but unknown reliability

Nuclear reactor safety

- Project BEMUSE of the Nuclear Energy Agency.
- $n = 10$ sources (CEA, IRSN,...) providing uncertain estimates of parameter values of a nuclear power plant.
- Costly data and complex phenomena involved $\rightarrow$ no reliable means to know the source reliabilities.
- Chose $\mathcal{F}$ with $f_j = \otimes^*_{r_j}$ ($n - j + 1$ out of n reliable).
- PCT2 parameter with domain $\Omega = \{x_1, \ldots, x_6\}$
  - $\phi_\pi(m^1) = 0.19$ (all sources reliable)
  - $\phi_\pi(m^2) = 0.81$ (9 out of 10 reliable)
  - $\phi_\pi(m^3) = 1$ (8 out of 10 reliable)
  - Values x_4 and x_5 are definitely more plausible.

$\rightarrow$ Results that are consistent, informative and readable by the end-user.

Summary: principled rule selection

	rel	$\neg rel$
ind	$\oplus$	$\circledast$, $\odot$ $CIT(\otimes_r)$, $CIT(DC)$
$\neg ind$	$\owedge$ $CIT(\owedge_{\mathcal{T}})$	$\ovee$ $CIT(\ovee_{\mathcal{T}})$

Outline

1. Preliminaries
2. Dempster's rule
   - Justifications
   - Properties
   - Computation
   - Conflict
3. Alternative rules
   - Reliability
   - Dependence
   - *n*-ary extensions
4. Rule selection
   - Principled
   - Performance-oriented

Learning data: typical setting

- ℓ objects $o_1 \ldots, o_\ell$ for which we have observed the true values $\hat{x}_1, \ldots, \hat{x}_\ell$ of some X.
- For each object o_i, mass functions $m_{1,i}, \ldots, m_{n,i}$ about the true value of X.
- A loss function $\mathcal{L}(m, \hat{x})$ evaluating the error of knowing m about X for a given object whose true value for X is $\hat{x}$.
- From a set $\mathcal{F}$ of possible rules, choose

$$f^* = \underset{f \in \mathcal{F}}{\operatorname{argmin}} \frac{1}{\ell} \sum_{i=1}^{\ell} \mathcal{L}(f(m_{1,i}, \ldots, m_{n,i}), \hat{x}_i)$$

- Remark: more or less complex optimisation problem to solve depending on chosen $\mathcal{F}$ and $\mathcal{L}$.

$\mathcal{F}$ and $\mathcal{L}$

- Typically, $\mathcal{L}(m, \hat{x})$ corresponds to transforming m into a probability measure P, and using the squared error (SE) or cross-entropy (CE) loss:

$$
\begin{aligned}
\mathcal{L}_{SE}(m, \hat{x}) &= \sum_{x \in \Omega} (1_{\hat{x}}(x) - p(x))^2 \\
\mathcal{L}_{CE}(m, \hat{x}) &= -\sum_{x \in \Omega} 1_{\hat{x}}(x) \log p(x)
\end{aligned}
$$

- $\mathcal{F}$ corresponds to a parameterized family of rules:
  - Independent but unknown reliability: $\otimes_r$ with parameter r, or DC with parameters α_i.
  - Not independent but reliable: $\bigcirc\!\!\!\!\wedge_{\mathcal{T}_s}$ for some family of t-norms $\mathcal{T}_s$ determined by parameter s.
  - Not independent and unknown reliability: $\bigcirc\!\!\!\!\vee_{\mathcal{T}_s}$.

Application, not independent but reliable

Classifier fusion [Quost et al. 2011]

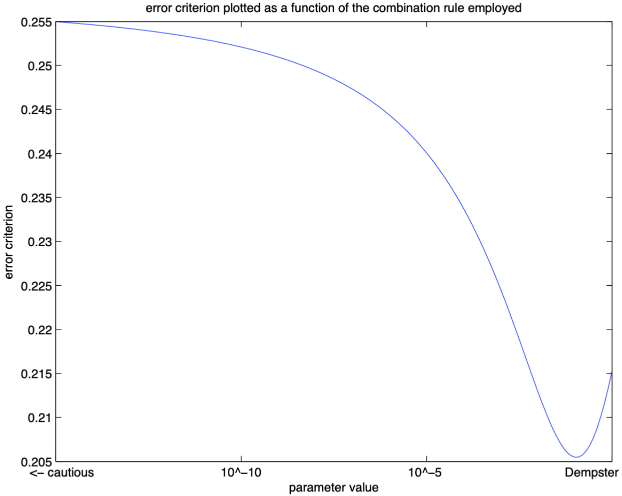
- Binary classification problem, with 10 features.
- One classifier learnt per feature, hence 10 classifiers.
- Conditionally on each class, correlation σ between any two of the first 9 features, last feature independent from all the others.

→ Experimental framework intended to ressemble a situation where there are 9 dependent classifiers, and a tenth classifier independent from the others.

- For each object, the *i*-th classifier produces a mass function m_i.
- The 10 obtained mass functions are combined using a parameterized t-rule $\owedge_{\mathcal{T}_s}$, with $s \in (0, 1]$ and such that $\owedge_{\mathcal{T}_s} = \owedge$ for $s \to 0$ and $\owedge_{\mathcal{T}_s} = \oplus$ for $s = 1$.
- Error criterion (loss function): pignistic probability transformation with SE.

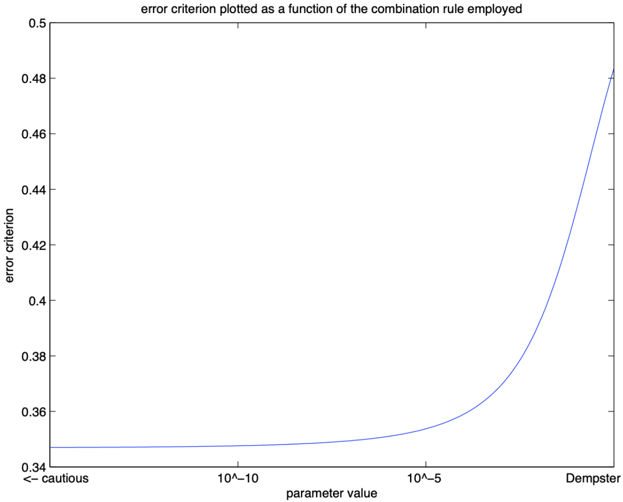
Results for $\sigma = 0.1$

[Quost et al. 2011]



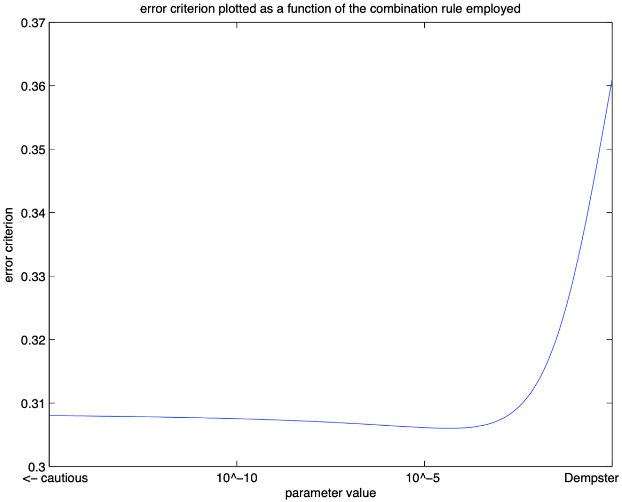
Results for $\sigma = 0.9$

[Quost et al. 2011]



Results for $\sigma = 0.5$

[Quost et al. 2011]



Summary: rule selection

	rel	$\neg$*rel*
ind	$\oplus$	$\circledast$, $\odot$ $CIT/\mathcal{L}(\otimes_r)$, $CIT/\mathcal{L}(DC)$
$\neg$*ind*	$\owedge$ $CIT/\mathcal{L}(\owedge_{\mathcal{T}})$	$\ovee$ $CIT/\mathcal{L}(\ovee_{\mathcal{T}})$

Summary

- Dempster's rule is a well-justified combination rule, satisfying important properties, appearing in numerous approaches to various problems and whose complexity can be managed.
- There exist alternative, well-justified, combination rules, corresponding to other assumptions/requirements.
- In practice, Dempster's rule is often effective and its underlying assumptions met (at least approximatively).
- However, if there is some uncertainty about the validity of these assumptions, there exist several (principled/performance-oriented) means to select an alternative rule addressing this uncertainty.

References I

Justifications

A. Dempster.
Upper and lower probabilities induced by a multivalued mapping.
The Annals of Statistics, 28:325–339, 1967.

G. Shafer.
A Mathematical Theory of Evidence.
Princeton University Press, Princeton, NJ, 1976.

G. Shafer.
Constructive probability.
Synthese, 48(1):1–60, 1981.

P. Smets.
Analyzing the combination of conflicting belief functions.
Information Fusion, 8:397–412, 2007.

F. Klawonn and Ph.Smets
The dynamic of belief in the transferable belief model and specialization–generalization matrices.
In Proc. of UAI 1992, p. 130–137. Morgan Kaufman, 1992.

D. Dubois and H. Prade.
On the Unicity of Dempster Rule of Combination.
Int. J. Intell. Syst., 1:133–142, 1986.

References II

P.Hájek
Deriving Dempster's rule.
In Uncertainty in Intelligent Systems, p. 75–83, 1993.

F. Pichon and T. Denœux.
The unnormalized Dempster's rule of combination: a new justification from the least commitment principle and some extensions.
Journal of Automated Reasoning, 45:61–87, 2010.

Canonical Decomposition

P. Smets.
The canonical decomposition of a weighted belief.
In Proc. of IJCAI 1995, p.1896–1901. Morgan Kaufmann, 1995.

D. Dubois, F. Faux, H. Prade.
Prejudice in uncertain information merging: Pushing the fusion paradigm of evidence theory further.
Int. Journal of Approximate Reasoning, 121:1–22, 2020.

F. Pichon.
Canonical decomposition of belief functions based on Teugels' representation of the multivariate Bernoulli distribution.
Information Sciences, 428:76–104, 2018.

References III

Ph. Smets.
Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem.
Int. Journal of Approximate Reasoning, 9:1–35, 1993.

T. Denœux.
A *k*-nearest neighbor classification rule based on Dempster-Shafer theory.
IEEE Trans on Syst, Man and Cybern, 25(5):804–813, 1995.

T. Denœux.
Logistic regression, neural networks and Dempster-Shafer theory: a new perspective.
Knowledge-Based Systems, 176:54–67, 2019.

F. Pichon, D. Mercier, É. Lefèvre, and F. Delmotte.
Proposition and learning of some belief function contextual correction mechanisms.
Int. Journal of Approximate Reasoning, 72:4–42, 2016.

Computation

N. Wilson.
Algorithms for Dempster-Shafer theory.
In D. M. Gabbay and P. Smets, editors, Hanbook of defeasible reasoning and uncertainty management, vol. 5, p.421–475. Kluwer Academic Publishers, 2000.

References IV

R. Kennes and Ph. Smets.
Computational aspects of the Möbius transformation.
In P. Bonissone, and M. Henrion, editors, Proc. 6th Conference on Uncertainty in Artificial Intelligence, p.344–351. MIT, 1990.

J.A. Barnett.
Computational methods for a mathematical theory of evidence
In Proc. of IJCAI-81, p.868–875. 1981.

J.D. Lowrance, T.D. Garvey, T.M. Strat
A framework for evidential-reasoning systems.
In Kehler et al, editors, Proc. AAAI'86, vol.2, p.896–903. AAAI Press,1986.

T. Denœux.
Inner and outer approximation of belief structures using a hierarchical clustering approach.
Int. Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 9(4):437–460, 2001.

T. Denœux and A. Ben Yaghlane.
Approximating the Combination of Belief Functions using the Fast Moebius Transform in a coarsened frame.
Int. Journal of Approximate Reasoning, 31(1-2):77–101, 2002.

T. Denœux and M.-H. Masson.
Evidential reasoning in large partially ordered sets. Application to multi-label classification, ensemble clustering and preference aggregation.
Annals of Operations Research, 195(1): 135–161, 2012.

References V

Conflict

L.A. Zadeh.
On the validity of Dempster's rule of combination of evidence.
Technical Report UCB/ERL M79/24, University of California, Berkely, 1979.

S. Destercke and T. Burger.
Toward an axiomatic definition of conflict between belief functions.
IEEE Trans. of Systems, Man and Cybernetics - Part B: Cybernetics, 42(2):585-596, 2013.

F. Pichon, A.-L. Jousselme, N. Ben Abdallah.
Several shades of conflict.
Fuzzy Sets and Systems, 366:63–84, 2019.

W. Liu.
Analyzing the degree of conflict among belief functions.
Artificial Intelligence, 170, p. 909-924, 2006.

A. Bronevich and I. Rozenberg.
The contradiction between belief functions: Its description, measurement, and correction based on generalized credal sets.
Int. Journal of Approximate Reasoning, 112:119–139, 2019.

J. Schubert.
Conflict management in Dempster-Shafer theory using the degree of falsity.
Int. Journal of Approximate Reasoning, 52(3):449–460, 2011.

References VI

Reliability

D. Dubois and H. Prade.
A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets.
Int. Journal of General Systems, 12(3):193–226, 1986.

F. Pichon, D. Dubois, and T. Denœux.
Relevance and truthfulness in information correction and fusion.
Int. Journal of Approximate Reasoning, 53(2):159–175, 2012.

F. Pichon, D. Dubois, and T. Denœux.
Quality of information sources in information fusion.
In Information Quality in Information Fusion and Decision Making, p. 31–49, Springer, 2019.

R. Haenni, and S. Hartmann.
Modeling partially reliable information sources: a general approach based on Dempster-Shafer theory.
Information Fusion, 7(4):361–379, 2006.

E. Lefèvre, O. Colot and P. Vannoorenberghe.
Belief function combination and conflict management.
Information Fusion, 3(2):149–162, 2002.

References VII

D. Dubois, W. Liu, J. Ma and H. Prade.
The basic principles of uncertain information fusion. An organised review of merging rules in different representation frameworks.
Information Fusion, 32:12–39, 2016.

Dependence

T. Denœux.
Conjunctive and disjunctive combination of belief functions induced by non-distinct bodies of evidence.
Artificial Intelligence 172:234–264, 2008.

S. Destercke and D. Dubois.
Idempotent conjunctive combination of belief functions: Extending the minimum rule of possibility theory.
Information Sciences, 181(18):3925–3945, 2011.

M.E.G.V. Cattaneo
Belief functions combination without the assumption of independence of the information sources.
Int. Journal of Approximate Reasoning, 52(3):299–315, 2011.

A. Bronevich and I. Rozenberg.
The choice of generalized dempster–shafer rules for aggregating belief functions.
Int. Journal of Approximate Reasoning, 56:122–136, 2015.

J. Klein, S. Destercke and O. Colot,
Idempotent conjunctive and disjunctive combination of belief functions by distance minimization,
Int. Journal of Approximate Reasoning, 92:32–48, 2018.

References VIII

Rule selection

T. Denœux.
Distributed combination of belief functions.
Information Fusion 65:179–191, 2021.

S. Zair and. S. Le Hégarat-Mascle
Evidential framework for robust localization using raw GNSS data.
Engineering Applications of Artificial Intelligence, 61:126–135, 2017.

F. Pichon, S. Destercke, and T. Burger.
A consistency-specificity trade-off to select source behavior in information fusion.
IEEE Transactions on Cybernetics, 45(4):598–609, 2015.

Z. Elouedi, K. Mellouli and Ph. Smets.
Assessing sensor reliability for multisensor data fusion within the transferable belief model.
IEEE Trans. Syst. Man Cybern. Part B, 34(1):782–787, 2004.

Z.-G. Liu, Q. Pan, J. Dezert and A. Martin.
Combination of classifiers with optimal weight based on evidential reasoning.
IEEE Transactions on Fuzzy Systems, 26:(3):1217–1230 2018.

B. Quost, M.-H. Masson and T. Denœux.
Classifier fusion in the Dempster-Shafer framework using optimized t-norm based combination rules.
Int. Journal of Approximate Reasoning, 52(3):353–374, 2011.

Software libraries for belief function combination

- Matlab:
  - FMT (Smets, extended versions from Denoeux, Martin (DST))
  - Approximation (Denoeux)
- R:
  - iBelief (Zhou & Martin)
  - Belief Package (Destercke)
  - dst (Boivin)
  - EvCombR (Karlsson)
- Python:
  - pyds (Reineking)
- C++:
  - 2CoBel (Pellicanò & Le Hégarat-Mascle & Aldea)
  - bft (Kurdej)
  - eVidenZ (Burrus & Lesage)
- Java:
  - Java Dempster Shafer Library (Reineking)
  - evidence4j (based on eVidenZ)

Demo Matlab FMT

Dempster's rule followed by outer clustering approximation

- Goal: computation of $\varphi^{+}_{\hat{\mathcal{P}}}(m_1 \oplus m_2)$ for $k = 2$ with m_1 and m_2 from the faulty MacBook example, i.e.,

$$\begin{aligned} m_1 &= \{p, c\}^{0.1}, \\ m_2 &= \{c, h\}^{0.2}. \end{aligned}$$

- We proceed in four steps:
  1. input m_1 and m_2 using the “focal set format”;
  2. compute $m_1 \oplus^* m_2$ using the commonality-based approach;
  3. compute $m_1 \oplus m_2$ by normalizing $m_1 \oplus^* m_2$;
  4. compute $\varphi^{+}_{\hat{\mathcal{P}}}(m_1 \oplus m_2)$ for $k = 2$.

Focal set format

- Let m be a mass function defined on $\Omega = \{\omega_1, \ldots, \omega_p\}$, with r focal sets: $\mathcal{F}(m) = \{F_1, \ldots, F_r\}$.
- m can be represented by a pair $(\mathbf{mass}, \mathbf{F})$, where **mass** is the r-dimensional column vector of masses

$$\begin{bmatrix} m(F_1) \\ \vdots \\ m(F_r) \end{bmatrix}$$

  and **F** is a $r \times p$ binary matrix such that

$$\mathbf{F}_{ij} = \begin{cases} 1 & \text{if } \omega_j \in F_i, \\ 0 & \text{otherwise.} \end{cases}$$

- This format is convenient to input a mass function.

Step 1: input m_1 and m_2

- Let $\Omega = \{p, c, h, o\}$.
- $m_1 = \{p, c\}^{0.1}$, i.e., $m_1(\{p, c\}) = 0.9, m_1(\Omega) = 0.1$.

```
mass1=[0.9;0.1];
F1=[1 1 0 0 % {p,c}
1 1 1 1]; % {p,c,h,o}
```

- $m_2 = \{c, h\}^{0.2}$, i.e., $m_2(\{c, h\}) = 0.8, m_2(\Omega) = 0.2$.

```
mass2=[0.8;0.2];
F2=[0 1 1 0 % {c,h}
1 1 1 1]; % {p,c,h,o}
```

Step 2: compute $m_1 \oplus^* m_2$ using the commonality-based approach

- A commonality function q is specified with the $2^{|\Omega|}$ numbers $q(A), A \subseteq \Omega$.
- It can be represented by a $2^{|\Omega|}$-dimensional column vector **q** whose element j stores $q(A_j)$ with A_j the subset of Ω such that $\omega_i \in A_j$ if the i-th bit in the binary representation of $j-1$ equals 1.
- Example for $\Omega = \{\omega_1, \omega_2, \omega_3\}$

Position	$\omega_3\, \omega_2\, \omega_1$	**q**
1	0 0 0	$q(\emptyset)$
2	0 0 1	$q(\{\omega_1\})$
3	0 1 0	$q(\{\omega_2\})$
4	0 1 1	$q(\{\omega_1, \omega_2\})$
5	1 0 0	$q(\omega_3)$
6	1 0 1	$q(\{\omega_1, \omega_3\})$
7	1 1 0	$q(\{\omega_2, \omega_3\})$
8	1 1 1	$q(\{\omega_1, \omega_2, \omega_3\})$

Step 2: compute $m_1 \oplus^* m_2$ using the commonality-based approach

- This "vector format" can also be used to represent mass, belief and plausibility functions.
- For instance, $\mathbf{m}_1$ is the $2^{|\Omega|}$-dimensional column vector whose element j stores $m_1(A_j)$.
- It is the format expected by the Matlab functions of the FMT toolbox that perform the transformations from one function (e.g. the mass function) to another (e.g. the commonality function).
- To be able to compute q_1 and q_2, we thus need first to convert m_1 and m_2 from the focal set format to the vector format. This is done with the function 'mtobbm':

```
m1 = mtobbm(mass1,F1);
m2 = mtobbm(mass2,F2);
```

Step 2: compute $m_1 \oplus^* m_2$ using the commonality-based approach

- The transformation from the mass function to the commonality function is done with the function 'mtoq'

```
q1=mtoq(m1);
q2=mtoq(m2);
```

- Computation of $q_{1\oplus^*2}$

```
q12=q1.*q2;
```

- Computation of $m_{1\oplus^*2}$, using the function 'qtom' which tranforms any commonality function into its associated mass function

```
m12=qtom(q12);
```

Step 3: compute $m_1 \oplus m_2$ by normalizing $m_1 \oplus^* m_2$

- Computation of $m_{1\oplus 2}$, using the function 'mtonm' which, given a mass function m, returns the mass function m' such that (normalization):

$$m'(A) = \begin{cases} \frac{m(A)}{1-m(\emptyset)} & \text{if } A \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

```
M12 = mtonm(m12);
```

Step 4: compute $\varphi^{+}_{\hat{\mathcal{P}}}(m_1 \oplus m_2)$ for $k = 2$

- The function 'apphier' performs the outer clustering approximation $\varphi^{+}_{\hat{\mathcal{P}}}(m)$ of a mass function m.
- It expects m to be provided in the focal set format.
- The conversion from the vector format of m to its focal set format is done with the function 'bbmtom'.

```
[Mass12, F12]= bbmtom(M12);
[Mass12out,F12out,C,N]=apphier(Mass12,F12,2,'out
   '); % C(i) is the cluster id of focal set i (
   in the original mass function) in the
   partition, N is the cardinality of the
   approximation
```

Step 4: compute $\varphi^{+}_{\hat{\mathcal{P}}}(m_1 \oplus m_2)$ for $k = 2$

- We obtain

```
Mass12out =

    0.7200
    0.2800

F12out =

    0    1    0    0
    1    1    1    1
```

Full program

```
mass1=[0.9;0.1];
F1=[1 1 0 0 % {p,c}
1 1 1 1]; % {p,c,h,o}
mass2=[0.8;0.2];
F2=[0 1 1 0 % {c,h}
1 1 1 1]; % {p,c,h,o}
m1 = mtobbm(mass1,F1); % Focal set format to vector format
m2 = mtobbm(mass2,F2);

q1=mtoq(m1);
q2=mtoq(m2);
q12=q1.*q2;
m12=qtom(q12); % conjunctive combination of m1 and m2
M12 = mtonm(m12); % Dempster's combination of m1 and m2

[Mass12, F12]= bbmtom(M12);
[Mass12out,F12out,C,N]=apphier(Mass12,F12,2,'out'); % Outer
    clustering approximation
```