Combination of belief functions

Frédéric Pichon

Laboratory of Computer Engineering and Automation of Artois (LGI2A)
Université d'Artois, Béthune, France
http://www.lgi2a.univ-artois.fr/~pichon

7th School on Belief Functions and their Applications
Granada, Spain, October 20, 2025

Contents of this lecture

- Dempster's rule of combination is the cornerstone of Shafer's theory of evidence.
- It allows the combination of independent and reliable pieces of evidence.
- However, two issues:
  1. Its computational complexity;
  2. Pieces of evidence are not always independent and reliable.
- This lecture:
  1. In practice, its complexity is manageable.
  2. An extension of Dempster's rule allowing us to account for various assumptions with respect to the reliability and dependence of the pieces of evidence.
  3. Methods to determine which assumptions to make about the reliability and dependence of the pieces of evidence.

Outline

1. Computation of Dempster's rule
   - Expressions
   - Exact
   - Approximate
2. Extension of Dempster's rule
   - Forms of unreliability for a piece of evidence
   - Partially reliable pieces of evidence
   - Dependent pieces of evidence
3. Rule selection
   - Uncertainty principle-based
   - Performance-based

Outline

1. Computation of Dempster's rule
    - Expressions
    - Exact
    - Approximate
2. Extension of Dempster's rule
    - Forms of unreliability for a piece of evidence
    - Partially reliable pieces of evidence
    - Dependent pieces of evidence
3. Rule selection
    - Uncertainty principle-based
    - Performance-based

Outline

1. Computation of Dempster's rule
   - Expressions
   - Exact
   - Approximate
2. Extension of Dempster's rule
   - Forms of unreliability for a piece of evidence
   - Partially reliable pieces of evidence
   - Dependent pieces of evidence
3. Rule selection
   - Uncertainty principle-based
   - Performance-based

Refresher

Unnormalized Dempster's rule (conjunctive rule)

- Let $(\Omega_1, P_1, \Gamma_1)$ and $(\Omega_2, P_2, \Gamma_2)$, with $\Gamma_i : \Omega_i \rightarrow 2^{\Theta} \setminus \{\emptyset\}$, $i = 1, 2$, be two messages representing two pieces of evidence about $X \in \Theta$ and inducing mass functions m_1 and m_2, respectively.
- If these messages are assumed to be independent and reliable, then our body of evidence is represented by the random set $(\Omega_1 \times \Omega_2, P_{12}, \Gamma_{\cap})$, with

$$
\begin{aligned}
P_{12}(\omega_1, \omega_2) &= P_1(\omega_1) P_2(\omega_2), \\
\Gamma_{\cap}(\omega_1, \omega_2) &= \Gamma_1(\omega_1) \cap \Gamma_2(\omega_2)
\end{aligned}
$$

- This random set induces the state of knowledge about X modeled by the conjunctive sum $m_{1 \textcircled{\cap} 2}$ such that

$$m_{1 \textcircled{\cap} 2}(A) = \sum_{B \cap C = A} m_1(B) m_2(C), \quad \forall A \subseteq \Theta.$$

Refresher

Dempster's rule, mass-based expression

- If one handles the inconsistency ($m_{1\textcircled{\cap}2}(\emptyset)$) that may be present in such a random set *à la Shafer*, i.e., by conditioning P_{12} on $\Theta_\cap = \{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \Gamma_\cap(\omega_1, \omega_2) \neq \emptyset\}$, then the probability of knowing that $X \in A$ from these messages satisfies, for all $A \neq \emptyset$,

$$\begin{aligned} m_{1\oplus 2}(A) &= m^*_{1\textcircled{\cap}2}(A) \quad \text{(normalized conjunctive sum)} \\ &= \frac{m_{1\textcircled{\cap}2}(A)}{1-(m_{1\textcircled{\cap}2}(\emptyset)}, \\ &= \frac{\sum_{B\cap C=A} m_1(B)m_2(C)}{1-\sum_{B\cap C=\emptyset} m_1(B)m_2(C)} \end{aligned}$$

- The orthogonal sum $m_{1\oplus 2}$ is the mass function generated by the random set $(\Omega_1 \times \Omega_2, P_\cap, \Gamma_\cap)$ with $P_\cap$ the probability measure resulting from the conditioning of P_{12} on the event $\Theta_\cap$.
- It is well defined if $1 - m_{1\textcircled{\cap}2}(\emptyset) = P_{12}(\Theta_\cap) > 0$.
- The binary operation $\oplus$ is called Dempster's rule.

Expression using commonalities

- Commonality function $q : 2^{\Theta} \to [0,1]$

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

- Conversely,

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

- We have

$$q_{1 \bigcirc\!\!\!\!\cap\, 2}(A) = q_1(A) \cdot q_2(A), \quad \forall A,$$

and

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

with $K = \left(\sum_{\emptyset \neq B \subseteq \Theta} (-1)^{|B|+1} q_{1 \bigcirc\!\!\!\!\cap\, 2}(B) \right)^{-1} = \left(1 - m_{1 \bigcirc\!\!\!\!\cap\, 2}(\emptyset)\right)^{-1}$.

Complexity

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

Complexity

- Orthogonal sum $m_{1\oplus 2}$:
  - Mass-based approach;
  - Commonality-based approach (program).
- Computing times proportional to, respectively:
  - $|\Theta||\mathcal{F}(m_1)||\mathcal{F}(m_2)|$
  - $|\Theta|^2 2^{|\Theta|}$ (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.

Complexity

- Orthogonal sum $m_{1\oplus 2}$:
  - Mass-based approach;
  - Commonality-based approach (program).
- Computing times proportional to, respectively:
  - $|\Theta||\mathcal{F}(m_1)||\mathcal{F}(m_2)|$
  - $|\Theta|^2 2^{|\Theta|}$ (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 $|\Theta|$.
- However, for practical applications (typically involving several mass functions), this is rarely an issue...

Outline

1. Computation of Dempster's rule
   - Expressions
   - Exact
   - Approximate
2. Extension of Dempster's rule
   - Forms of unreliability for a piece of evidence
   - Partially reliable pieces of evidence
   - Dependent pieces of evidence
3. Rule selection
   - Uncertainty principle-based
   - Performance-based

Particular forms of mass functions

Simple mass functions

- A mass function is simple if it has two focal sets: Θ and A for some $A \subset \Theta$, which means it is of the form

$$m(A) = 1 - d, \quad m(\Theta) = d,$$

for some $d \in [0, 1]$. It is denoted by A^d.

  - It represents a message that means $X \in A$ with probability $1 - d$, and that is useless, i.e., means $X \in \Theta$, with probability d.
  - Prototypical example: a sensor reporting $X \in A$ and faulty with probability d.

- If each mass function is of the form $\{\theta\}^d$ or $\overline{\{\theta\}}^d$, for some $\theta \in \Theta$, the complexity becomes linear.

Particular forms of mass functions

Interval focal sets

- If Θ is linearly ordered, and the focal sets of the mass functions are constrained to be intervals, the complexity becomes polynomial.
- Example: X is the number of days before the attack
  - $\Theta = \{1, \ldots, 30\}$
  - $A \subseteq \Theta$ is an interval if there exist elements a and b of Θ such that $A = \{\theta \in \Theta | a \leq \theta \leq b\}$.
  - Such A is denoted by $[a, b]$.
  - For instance, $A = \{12, 13, 14, 15, 16\} = [12, 16]$.

Particular forms of mass functions

Lattice interval focal sets

- More generally, the complexity is polynomial if there is a partial ordering $\leq$ of Θ such that $(\Theta, \leq)$ 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).
- Polynomial complexity because the intersection of two intervals $[a, b]$ and $[c, d]$ of $(\Theta, \leq)$, required by Dempster's rule, is an interval

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

Particular forms of mass functions

Lattice interval focal sets (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 Θ and which are thus intractable in the usual case.
- Indeed, in such applications, mass functions having only (lattice) interval focal sets are naturally encountered...

Particular forms of mass functions

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 $\Xi = \{\xi_1, \ldots, \xi_c\}$ be the set of classes.
- Class label X of an instance takes values in $\Theta = 2^{\Xi}$.
- Let θ_A be the element of Θ corresponding to $A \subseteq \Xi$

Particular forms of mass functions

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 $\Xi = \{\xi_1, \ldots, \xi_c\}$ be the set of classes.
- Class label X of an instance takes values in $\Theta = 2^{\Xi}$.
- Let θ_A be the element of Θ corresponding to $A \subseteq \Xi$
- Partial ordering on Θ: $\theta_A \leq \theta_B \Leftrightarrow A \subseteq B$, for all $A, B \subseteq \Xi$.
- Interval $[\theta_A, \theta_B]$ of lattice $(\Theta, \leq)$, 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.

Particular forms of mass functions

Examples of lattice intervals: Ensemble clustering

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

Particular forms of mass functions

Examples of lattice intervals: Ensemble clustering

- Clustering a set Ξ of n objects = finding a partition of Ξ.
- Let Θ be set of all partitions of Ξ.
- The "true" partition X of the objects takes values in Θ.
- Partial ordering on Θ: $\theta \leq \theta'$ (θ is finer than θ'), for all $\theta, \theta' \in \Theta$, if the clusters of θ can be obtained by splitting those of θ'.
- Interval $[\underline{\theta}, \overline{\theta}]$ of lattice $(\Theta, \leq)$, for $\underline{\theta} \leq \overline{\theta}$, is an imprecise specification of X: it is coarser than $\underline{\theta}$ and finer than $\overline{\theta}$.
- For instance, "the objets of a set $A \subseteq \Xi$ belong to the same cluster" can be represented by the interval $[\theta_A, \theta_\Xi]$, where θ_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.

Decision making

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

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

Outline

1. Computation of Dempster's rule
   - Expressions
   - Exact
   - Approximate
2. Extension of Dempster's rule
   - Forms of unreliability for a piece of evidence
   - Partially reliable pieces of evidence
   - Dependent pieces of evidence
3. Rule selection
   - Uncertainty principle-based
   - Performance-based

Approximate computation

- Approximate computation when the exact computation is not possible.
- Stochastic approximation procedures:
  - (Approximate) Combined belief for some $A \subset \Theta$ can be computed by Monte Carlo algorithms in time linear in $|\Theta|$;
  - 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 $m \sqsubseteq \varphi^+(m)$: it is called an outer approximation of m.

Proposition (Monotonicity of $\bigcirc\!\!\!\!\cap$ with respect to $\sqsubseteq$)

$$m \sqsubseteq m' \Rightarrow m \bigcirc\!\!\!\!\!\cap\, m_0 \sqsubseteq m' \bigcirc\!\!\!\!\!\cap\, m_0, \quad \forall m_0$$

- From these properties, we have

$$m_{\bigcirc\!\!\!\!\cap} \sqsubseteq m^+$$

with

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

- 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 $\varphi^-(m) \sqsubseteq m$: it is called an inner approximation of m.
- Furthermore,

$$m^- \sqsubseteq m_{\cap} \sqsubseteq m^+$$

with

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

- We have

$$pl^- \leq pl_{\cap} \leq pl^+$$

- Bounds on $bel_{\cap}$ 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

$$pl_\oplus(A) = \frac{pl_{\bigcirc\!\!\!\!\cap}(A)}{pl_{\bigcirc\!\!\!\!\cap}(\Theta)}, \quad \forall A \subseteq \Theta$$

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

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

Approximation for the mass-based approach

- Let m be a mass function.
- The summarization algorithm produces a less informative (in the sense of $\sqsubseteq$) approximation $\varphi^+(m)$ of m (we have $m \sqsubseteq \varphi^+(m)$).
- 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 content change.
- When approximating m, we indeed want to preserve as much as possible of its information content.

Approximation for the mass-based approach

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

$$|m| := \sum_{\emptyset \neq A \subseteq \Theta} m(A)|A|,$$

  the greater the cardinality of m, the less informative m is.
- Recall that we have

$$m_1 \sqsubseteq m_2 \Rightarrow |m_1| \leq |m_2|,$$

  hence, if information content is measured using cardinality, $\varphi^+(m)$ constitutes a loss of information.
- 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.

$\rightarrow$ 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.

- Remark: this is a case where m is substituted by a mass function chosen in a set $\mathcal{M}$ of less specialized mass functions than m, and by choosing the cardinality as the uncertainty measure, we have seen that we should apply the minimum uncertainty principle, which is what is happening here.

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.
- This algorithm takes 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:

$$m^-_{\hat{\mathcal{P}}} \sqsubseteq m_{\bigcirc\!\!\!\!\cap} \sqsubseteq m^+_{\hat{\mathcal{P}}}$$

with

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

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

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

Approximation for the commonality-based approach

- Complexity depends on $|\Theta| \rightarrow$ approximate Θ by a simpler (coarser) frame Ξ 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) Ξ of Θ of size c, minimizing information loss defined as

     $$\sum_{i=1}^{n} \Delta(m_i^{\downarrow\Xi\uparrow\Theta}, m_i)$$

     with $m_i^{\downarrow\Xi\uparrow\Theta}$ the outer approximation of m_i obtained by carrying m_i to Ξ (restriction $m_i^{\downarrow\Xi}$) and carrying it back to Θ (vacuous extension $\uparrow\Theta$)
  2. Using the commonality-based approach, combine the mass functions in the coarsened frame, i.e., compute $m^{\Xi} := \text{\textcircled{$\cap$}}_{i=1}^{n} m_i^{\downarrow\Xi}$
  3. Carry the result to Θ, i.e., compute $\overline{m} := m^{\Xi\uparrow\Theta}$
- $\overline{m}$ is an outer approximation of $m_{\text{\textcircled{$\cap$}}}$.

Approximation for the commonality-based approach

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

$$\underline{m} \sqsubseteq m_{\textcircled{\cap}} \sqsubseteq \overline{m}$$

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

Outline

1. Computation of Dempster's rule
   - Expressions
   - Exact
   - Approximate
2. Extension of Dempster's rule
   - Forms of unreliability for a piece of evidence
   - Partially reliable pieces of evidence
   - Dependent pieces of evidence
3. Rule selection
   - Uncertainty principle-based
   - Performance-based

Motivation

- The assumptions leading to Dempster's rule are that the pieces of evidence to be combined are independent and reliable.
- These assumptions clearly do not always hold.

$\rightarrow$ An extension of Dempster's rule allowing us to account for various assumptions with respect to the reliability and dependence of the pieces of evidence.

- A prism to understand most of the main alternative combination schemes to Dempster's rule.

Outline

1. Computation of Dempster's rule
   - Expressions
   - Exact
   - Approximate
2. Extension of Dempster's rule
   - Forms of unreliability for a piece of evidence
   - Partially reliable pieces of evidence
   - Dependent pieces of evidence
3. Rule selection
   - Uncertainty principle-based
   - Performance-based

Reliability

- The reliability of a piece of evidence is classically understood in terms of relevance, i.e., it is reliable if it provides useful information regarding the variable of interest.
- Examples:
  - A broken watch is useless to try and find the time it is since there is no way to know whether the supplied information is correct or not: it is not reliable for the time;
  - My twelve-year-old son is ignorant about the name of the latest Nobel Peace Prize laureate: he is not reliable for this question (in contrast to nobelprize.org).
- Basic idea : a piece of evidence is valid if it is reliable, whereas it is useless if it is unreliable.

Formalization

- Assume a piece of evidence corresponding to a message whose meaning is $X \in A \subseteq \Theta$.
  - If it is unreliable, we replace $X \in A$ by $X \in \Theta$
  - If it is reliable, we keep $X \in A$
- Let R be the variable denoting its reliability, defined on $\mathcal{R} = \{rel, unrel\}$.
- The interpretation of the message according to the reliability may be modeled by $\Pi_A : \mathcal{R} \to 2^\Theta$ such that

$$\begin{aligned} \Pi_A(rel) &= A, \\ \Pi_A(unrel) &= \Theta. \end{aligned}$$

Uncertain reliability

- Let (Ω, P, Γ) be a message representing a piece of evidence about X and inducing mass function m.
- Assume this message to be unreliable with probability $P^{\mathcal{R}}(\mathit{unrel}) = \alpha$.
- What can then be inferred about X?
- If the actual code was $\omega \in \Omega$ and
  - the message is reliable, we know that $X \in \Pi_{\Gamma(\omega)}(\mathit{rel})$
  - the message is unreliable, we know that $X \in \Pi_{\Gamma(\omega)}(\mathit{unrel}) = \Theta$
- Hence, the probability to know $X \in A \subset \Theta$ is

$$
\begin{aligned}
{}^{\alpha}m(A) &:= P^{\mathcal{R}}(\mathit{rel}) \cdot \sum_{\omega:\Pi_{\Gamma(\omega)}(\mathit{rel})=A} P(\omega) \\
&= (1-\alpha) \cdot m(A).
\end{aligned}
$$

Uncertain reliability (continued)

- The random set

$$(\Omega \times \mathcal{R}, P \times P^{\mathcal{R}}, \Gamma^{\mathcal{R}})$$

with

$$\Gamma^{\mathcal{R}}(\omega, r) := \Pi_{\Gamma(\omega)}(r)$$

for all $(\omega, r) \in \Omega \times \mathcal{R}$, represents all the available information and the knowledge it induces about X is represented by ${}^{\alpha}m$.

- ${}^{\alpha}m$ is known as the discounting with discount rate α of mass function m.
- It is the most basic so-called correction operation for a mass function.

Contextual reliability

- Unreliability can be refined into contextual unreliability, leading to a more general correction model.
- State *unrel* and a given message $X \in A$: one must discard what the message tells with respect to each $\theta \in \Theta$ (each $\theta \in \Theta$ is then still a possible value for X).
- Unreliable only for some $\theta \in \Theta$: one must discard what the message tells only for these values.
- Assume unreliability for all the values in some $B \subseteq \Theta$ ($unrel_B$ for short): we must then replace $X \in A$ by $X \in A \cup B$.
- Let $\mathcal{R}^{\mathcal{C}} = \{unrel_B\}_{B \subseteq \Theta}$.
- The interpretation of the message according to this refined, "contextual", reliability may be modeled by $\Gamma_A^{\mathcal{C}} : \mathcal{R}^{\mathcal{C}} \rightarrow 2^{\Theta}$ s.t.

$$\Gamma_A^{\mathcal{C}}(unrel_B) = A \cup B.$$

- The crude reliability model corresponds to states $unrel_{\emptyset}$ (=*rel*) and $unrel_{\Theta}$ (=*unrel*).

Uncertain contextual reliability

- Assume uncertainty in the form of a probability distribution $P^{\mathcal{R}^{\mathcal{C}}}$, about the contextual reliability of a message (Ω, P, Γ) representing a piece of evidence and inducing mass function m.
- Our knowledge about X is then represented by the mass function ${}^{\mathcal{C}}m$ generated by the random set

$$(\Omega \times \mathcal{R}^{\mathcal{C}}, P \times P^{\mathcal{R}^{\mathcal{C}}}, \Gamma^{\mathcal{C}})$$

where $\Gamma^{\mathcal{C}}(\omega, r) := \Pi^{\mathcal{C}}_{\Gamma(\omega)}(r)$ for all $(\omega, r) \in \Omega \times \mathcal{R}^{\mathcal{C}}$.

Uncertain contextual reliability

Particular case

- If we have independent probabilities α_k of being unreliable for value θ_k, for all $1 \leq k \leq K$, then

$$P^{\mathcal{R}^{\mathcal{C}}}(unrel_B) = \prod_{\theta_k \in B} \alpha_k \prod_{\theta_\ell \in \overline{B}} (1 - \alpha_\ell)$$

- ${}^{\mathcal{C}}m$ is then known as the contextual discounting of m with discount rate vector $(\alpha_1, \ldots, \alpha_K)$.

Truthfulness

- Another refined form of reliability is when reliability includes another dimension besides the relevance: the truthfulness.
- Being truthful means actually supplying the information possessed.
- Lack of truthfulness can take several forms, and can be intentional or not.
- For instance, a sensor that has a systematic bias is a kind of unintentional lack of truthfulness.
- We consider here the crudest form, where non truthful means telling the contrary of what is known.

Formalization

- Assume a piece of evidence corresponding to a message whose meaning is $X \in A \subseteq \Theta$.
  - If it is not relevant, we replace $X \in A$ by $X \in \Theta$.
  - If it is relevant,
    - either it is truthful, in which case we keep $X \in A$.
    - or it lies, in which case we replace $X \in A$ by $X \in \overline{A}$.
- Relevance R defined on $\mathcal{R} = \{rel, \neg rel\}$.
- Truthfulness T defined on $\mathcal{T} = \{tru, \neg tru\}$.
- Let $\mathcal{R}^{\mathcal{T}} := \mathcal{R} \times \mathcal{T}$.
- The interpretation of the message according to the relevance and truthfulness may be modeled by $\Pi_A^{\mathcal{T}} : \mathcal{R}^{\mathcal{T}} \to 2^{\Theta}$ such that

$$\Pi_A^{\mathcal{T}}(rel, tru) = A, \quad \Pi_A^{\mathcal{T}}(rel, \neg tru) = \overline{A},$$
$$\Pi_A^{\mathcal{T}}(\neg rel, tru) = \Pi_A^{\mathcal{T}}(\neg rel, \neg tru) = \Theta.$$

- Uncertainty can be considered, leading to a generalization of discounting.
- Contextual non truthfulness can also be considered.

Outline

Uncertain reliability

- Let $(\Omega_1, P_1, \Gamma_1)$ and $(\Omega_2, P_2, \Gamma_2)$ be two messages representing two pieces of evidence about X and inducing mass functions m_1 and m_2, respectively.
- Assume that these messages are independent, i.e.,

$$P_{12}(\omega_1, \omega_2) = P_1(\omega_1)P_2(\omega_2), \quad \forall(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2.$$

- Let R_i defined on $\mathcal{R}_i = \{rel_i, unrel_i\}$ denote the reliability of message i, $i = 1, 2$, and let $\boldsymbol{\mathcal{R}} := \mathcal{R}_1 \times \mathcal{R}_2$.
- Assume uncertainty $P^{\boldsymbol{\mathcal{R}}}$ on their reliabilities.
- Our knowledge about X may then be defined as the mass function ${}^{\boldsymbol{\mathcal{R}}}m$ induced by the random set

$$(\Omega_1 \times \Omega_2 \times \boldsymbol{\mathcal{R}}, P_{12} \times P^{\boldsymbol{\mathcal{R}}}, \Gamma^{\boldsymbol{\mathcal{R}}})$$

where, for all $\mathbf{r} = (r_1, r_2) \in \boldsymbol{\mathcal{R}}$,

$$\Gamma^{\boldsymbol{\mathcal{R}}}(\omega_1, \omega_2, \mathbf{r}) := \Gamma_1^{\mathcal{R}}(\omega_1, r_1) \cap \Gamma_2^{\mathcal{R}}(\omega_2, r_2).$$

Uncertain reliability

Normalization

- It is possible that ${}^{\mathcal{R}}m(\emptyset) > 0$, reflecting some inconsistency in the body of evidence.
- It can be resolved by extending Shafer's reasoning leading to Dempster's rule: if the decoding, according to some pair of codes (ω_1, ω_2), of the messages, leads to a contradiction under some reliability state $\mathbf{r}$, then we know that this pair of codes together with this reliability state could not be the actual ones.

$\rightarrow$ $P_{12} \times P^{\mathcal{R}}$ conditioned on
$\Theta_{\mathcal{R}} = \{(\omega_1, \omega_2, \mathbf{r}) \in \Omega_1 \times \Omega_2 \times \mathcal{R} : \Gamma^{\mathcal{R}}(\omega_1, \omega_2, \mathbf{r}) \neq \emptyset\}$

- Let $P_{\mathcal{R}}$ denote the probability measure resulting from this conditioning. Our knowledge about X may then be defined as the mass function generated by the random set

$$(\Omega_1 \times \Omega_2 \times \mathcal{R}, P_{\mathcal{R}}, \Gamma^{\mathcal{R}}).$$

- This mass function is nothing but ${}^{\mathcal{R}}m^*$, i.e., ${}^{\mathcal{R}}m$ normalized.

Uncertain reliability

Particular cases

${}^{\mathcal{R}}m^*$ reduces to

- $m_1 \oplus m_2$ if $P^{\mathcal{R}}(rel_1, rel_2) = 1$, i.e., the messages are reliable

$\rightarrow$ Dempster's rule

- ${}^{\alpha_1}m_1 \oplus {}^{\alpha_2}m_2$ if $P^{\mathcal{R}} = P^{\mathcal{R}_1} \times P^{\mathcal{R}_2}$, with $P^{\mathcal{R}_i}(unrel_i) = \alpha_i$, i.e., the messages have independent probabilities α_1 and α_2 of being unreliable

$\rightarrow$ "Discount and combine" scheme

- $\alpha m_1 + (1-\alpha)m_2$ if $P^{\mathcal{R}}(rel_1, unrel_2) = \alpha$, $P^{\mathcal{R}}(unrel_1, rel_2) = 1-\alpha$, i.e., the messages have dependent reliabilities such that $R_2 = \neg R_1$

$\rightarrow$ Weighted average

⚠ In general, $\alpha m_1 + (1-\alpha)m_2 \neq {}^{\alpha}m_1 \oplus {}^{1-\alpha}m_2$

Imprecise reliability

- Assume the reliability is known in the form of $\mathbf{R} \subseteq \mathcal{R}$.
- Then we obtain (following Shafer's resolution of inconsistency) the mass function ${}^{\mathbf{R}}m^*$ about X induced by the random set

$$(\Omega_1 \times \Omega_2, P_{\mathbf{R}}, \Gamma_{\mathbf{R}})$$

  where
  - $\Gamma_{\mathbf{R}}(\omega_1, \omega_2) := \cup_{\mathbf{r} \in \mathbf{R}} \Gamma^{\mathcal{R}}(\omega_1, \omega_2, \mathbf{r})$
  - $P_{\mathbf{R}}$: P_{12} conditioned on $\Theta_{\mathbf{R}} = \{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \Gamma_{\mathbf{R}}(\omega_1, \omega_2) \neq \emptyset\}$
- Remark: Both imprecision and uncertainty about the reliability can be taken into account by considering a mass function on $\mathcal{R}$, leading to a general model subsuming the previous ones.

Imprecise reliability

Particular cases

${}^{\mathbf{R}}m^*$ reduces to

- $m_1 \circledcup m_2$ (same definition as $\circledcap$, except that $\cap$ is replaced by $\cup$) for $\mathbf{R} = \{(rel_1, rel_2), (rel_1, unrel_2), (unrel_1, rel_2)\}$, i.e., at least one of the two messages is reliable.

$\rightarrow$ Disjunctive rule, which satisfies similar properties as Dempster's rule, in particular commutativity, associativity and expression based on pointwise product of belief functions (if m_1 and m_2 are normalized).

- in the case where we receive $N \geq 2$ messages, $\mathbf{R} =$ "$N - Q$ out of the N messages are reliable".

$\rightarrow$ Q-relaxation rule, which is commutative, extends Dempster's rule ($Q = 0$) and the disjunctive rule ($Q = N - 1$), and generalizes the q-relaxation technique from interval analysis, which is designed to implement some form of robustness to outliers.

Relevance and truthfulness

- Assume two pieces of evidence corresponding to two messages $X \in A_1$ and $X \in A_2$, respectively.
- Let R_i defined on $\mathcal{R}_i^{\mathcal{T}} := \mathcal{R}_i \times \mathcal{T}_i$ denote the relevance and truthfulness of message i and let $\boldsymbol{\mathcal{R}}^{\mathcal{T}} := \mathcal{R}_1^{\mathcal{T}} \times \mathcal{R}_2^{\mathcal{T}}$.
- For any assumption $\mathbf{r} = (r_1, r_2) \in \boldsymbol{\mathcal{R}}^{\mathcal{T}}$, we deduce

$$X \in \Pi(\mathbf{r}) := \Pi_{A_1}^{\mathcal{T}}(r_1) \cap \Pi_{A_2}^{\mathcal{T}}(r_2)$$

  and, for an imprecise assumption $\mathbf{R} \subseteq \boldsymbol{\mathcal{R}}^{\mathcal{T}}$, we know

$$X \in \Pi(\mathbf{R}) = \cup_{\mathbf{r} \in \mathbf{R}} \Pi(\mathbf{r})$$

- Example: $\mathbf{R} = \{(rel_1, tru_1, rel_2, \neg tru_2), (rel_1, \neg tru_1, rel_2, tru_2)\}$

$$\begin{aligned}\Pi(\mathbf{R}) &= \Pi(rel_1, tru_1, rel_2, \neg tru_2) \cup \Pi(rel_1, \neg tru_1, rel_2, tru_2)\\ &= (A_1 \cap \overline{A_2}) \cup (\overline{A_1} \cap A_2)\\ &= A_1 \Delta A_2 \text{ (exclusive or)}\end{aligned}$$

$\rightarrow$ All connectives of Boolean logic can be reinterpreted in terms of assumptions wrt the relevance and truthfulness

General case

- Consider a mass function $m^{\mathcal{R}^{\mathcal{T}}}$ representing uncertain and imprecise knowledge about the relevance and truthfulness of two independent messages $(\Omega_1, P_1, \Gamma_1)$ and $(\Omega_2, P_2, \Gamma_2)$.
- Let $\mathcal{B}$ be the set of binary Boolean connectives.
- Any focal set **R** of $m^{\mathcal{R}^{\mathcal{T}}}$ yields a connective $b \in \mathcal{B}$.
- A connective $b \in \mathcal{B}$ may be retrieved for different $\mathbf{R} \subseteq \mathcal{R}^{\mathcal{T}}$.

$\rightarrow$ $m^{\mathcal{R}^{\mathcal{T}}}$ actually induces a probability distribution $P^{\mathcal{B}}$ over the connectives to be used to combine the messages.

General case

- Consider a mass function $m^{\mathcal{R}^{\mathcal{T}}}$ representing uncertain and imprecise knowledge about the relevance and truthfulness of two independent messages $(\Omega_1, P_1, \Gamma_1)$ and $(\Omega_2, P_2, \Gamma_2)$.
- Let $\mathcal{B}$ be the set of binary Boolean connectives.
- Any focal set **R** of $m^{\mathcal{R}^{\mathcal{T}}}$ yields a connective $b \in \mathcal{B}$.
- A connective $b \in \mathcal{B}$ may be retrieved for different $\mathbf{R} \subseteq \mathcal{R}^{\mathcal{T}}$.

→ $m^{\mathcal{R}^{\mathcal{T}}}$ actually induces a probability distribution $P^{\mathcal{B}}$ over the connectives to be used to combine the messages.

- Our knowledge about X given $m^{\mathcal{R}^{\mathcal{T}}}$ may then be defined as the mass function ${}^{\mathcal{B}}m^*$ induced by the random set

$$(\Omega_1 \times \Omega_2 \times \mathcal{B}, P_{\mathcal{B}}, \Gamma^{\mathcal{B}})$$

where $\Gamma^{\mathcal{B}}(\omega_1, \omega_2, b) := \Gamma_1(\omega_1) \otimes_b \Gamma_2(\omega_2)$ with $\otimes_b$ the set-theoretic connective associated to b, and $P_{\mathcal{B}}$ is $P_{12} \times P^{\mathcal{B}}$ conditioned on $\Theta_{\mathcal{B}} = \{(\omega_1, \omega_2, b) \in \Omega_1 \times \Omega_2 \times \mathcal{B} : \Gamma^{\mathcal{B}}(\omega_1, \omega_2, b) \neq \emptyset\}$.

General case (continued)

Theorem

$$^{\mathcal{B}}m^*(A) = \frac{\sum_b P^{\mathcal{B}}(b) \sum_{B \otimes_b C = A} m_1(B) m_2(C)}{1 - \sum_b P^{\mathcal{B}}(b) \sum_{B \otimes_b C = \emptyset} m_1(B) m_2(C)}, \textit{ for all } A \in 2^{\Theta} \setminus \{\emptyset\}.$$

$\rightarrow$ Generalization of Dempster's rule to all Boolean connectives, interpretable in terms of reliability assumptions

Theorem

$^{\mathcal{B}}m^*$ is the marginal on Θ of a graphical belief function model representing the messages and the assumptions on their reliabilities.

Dubois and Prade's rule

- Dempster's rule is obtained by considering $\Gamma(\omega_1, \omega_2, \cap) = \emptyset$ as an observation that (ω_1, ω_2) cannot be the pair of codes actually used (and keeping the assumption "$\cap$", i.e., the messages are reliable).
- An alternative is to view such contradiction as an indication that $\cap$ cannot be the actual reliability of the messages.
- In particular, if such a contradiction happens, we can instead safely assume that at least one of the messages is reliable, i.e., assumption "$\cup$".
- This is represented by, for all $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$, $P^{\mathcal{B}}(\cup | (\omega_1, \omega_2)) = 1$ if $\Gamma(\omega_1, \omega_2, \cap) = \emptyset$ and $P^{\mathcal{B}}(\cap | (\omega_1, \omega_2)) = 1$ otherwise.
- This yields the following probability distribution P_{DP} on $\Omega_1 \times \Omega_2 \times \mathcal{B}$:

$$
\begin{aligned}
P_{DP}(\omega_1, \omega_2, \cap) &= P_{12}(\omega_1, \omega_2), \quad \forall (\omega_1, \omega_2), \Gamma(\omega_1, \omega_2, \cap) \neq \emptyset, \\
P_{DP}(\omega_1, \omega_2, \cup) &= P_{12}(\omega_1, \omega_2), \quad \forall (\omega_1, \omega_2), \Gamma(\omega_1, \omega_2, \cap) = \emptyset.
\end{aligned}
$$

Dubois and Prade's rule (continued)

- The evidence is then represented by the random set

$$(\Omega_1 \times \Omega_2 \times \mathcal{B}, P_{DP}, \Gamma^{\mathcal{B}})$$

- The mass function ${}^{DP}m$ that it generates about X admits the following expression

$${}^{DP}m(A) = \sum_{B \cap C = A} m_1(B) m_2(C) + \sum_{B \cap C = \emptyset, B \cup C = A} m_1(B) m_2(C), \forall A \neq \emptyset$$

  and ${}^{DP}m(\emptyset) = 0$.
- This combination is known as Dubois and Prade's rule.
- Properties:
  - If $m_{1 \textcircled{\cap} 2}(\emptyset) = 0$, then ${}^{DP}m = m_{1 \oplus 2}$
  - If $m_{1 \textcircled{\cap} 2}(\emptyset) = 1$, then ${}^{DP}m = m_{1 \textcircled{\cup} 2}$
  - Commutativity, insensitivity to vacuous information, not associative.
- Remark: the codes and the reliability are not independent, but neither are they in the extension of Dempster's rule (due to the normalization step).

Applications

- Alternatives to Dempster's rule offer some flexibility for combining pieces of evidence that can be useful in practice.
- Examples from the literature:
  - Discount and combine: evidential k-nearest neighbor (EkNN) [Denœux, 1995] and evidential neural network (ENN) classifiers [Denœux, 2000]
  - Weighted average: tree ensembles [Zhang et al., 2023]
  - Contextual discounting: fusion of deep neural networks [Huang et al., 2025]
  - Q-relaxation rule: robustness to outliers [Pellicanò et al., 2018]

Outline

Beyond independence

- The preceding combinations rely on m_1 and m_2 being induced by independent messages $(\Omega_1, P_1, \Gamma_1)$ and $(\Omega_2, P_2, \Gamma_2)$, i.e., $P_{12} = P_1 \times P_2$.
- In principle, any dependence structure, and thus any P_{12} having P_1 and P_2 as marginals, can be selected.
- Example [Shafer, 1986]: $\Omega_i = \{\omega_i, \neg\omega_i\}$
  - $\Gamma_1(\neg\omega_1) = A$, $\Gamma_1(\omega_1) = \Theta$ and $P_1(\omega_1) = 0.2$
  - $\Gamma_2(\neg\omega_2) = \overline{A}$, $\Gamma_2(\omega_2) = \Theta$ and $P_2(\omega_2) = 0.01$
  - Dependence specified by $P_{12}(\omega_1|\omega_2) = 0.9$.
  - We have $P_{12} \neq P_1 \times P_2$.
  - Remark: such messages are called elementary, they induce simple mass functions ($A^{0.2}$ and $\overline{A}^{0.01}$, respectively).
- All preceding combinations can be extended to a known dependence between the messages by replacing $P_1 \times P_2$ by some P_{12} representing this dependence.

Example: reliable and dependent messages

- Let $(\Omega_1, P_1, \Gamma_1)$ and $(\Omega_2, P_2, \Gamma_2)$ be two messages representing two pieces of evidence about $X \in \Theta$ and inducing mass functions m_1 and m_2, respectively.
- Assume the messages are reliable and have some dependence structure described by the joint probability distribution P_{12} on $\Omega_1 \times \Omega_2$.
- This body of evidence is represented by the random set $(\Omega_1 \times \Omega_2, P_{12}, \Gamma_\cap)$, which induces knowledge about X modeled by the following mass function

$$m_{1\cap 2}(A) = P_{12}(\{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \Gamma_\cap(\omega_1, \omega_2) = A\}), \quad \forall A.$$

- An eventual inconsistency manifested in $m_{1\cap 2}$ may be resolved, as in the previous combinations, by normalization.

Example: reliable and dependent messages (cont'd)

- Let $B_1, \ldots, B_r$ denote the focal sets of m_1, $C_1, \ldots, C_s$ the focal sets of m_2, $p_i = m_1(B_i)$, $q_j = m_2(C_j)$ and

$$p_{ij} = P_{12}(\{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \Gamma_1(\omega_1) = B_i, \Gamma_2(\omega_2) = C_j\}).$$

- The following expression can be obtained for $m_{1 \cap 2}$:

$$m_{1 \cap 2}(A) = \sum_{B_i \cap C_j = A} p_{ij}, \quad \forall A.$$

$\rightarrow$ The induced knowledge $m_{1 \cap 2}$ about X, given some dependence between the messages encoded in P_{12}, is determined by p_{ij}, $\forall (i,j)$, which is the probability that 1st message means $X \in B_i$ and the 2nd message means $X \in C_j$.

- When $P_{12} = P_1 \times P_2$ (independence assumption, unnormalized Dempster's rule), we have $\forall (i,j), p_{ij} = p_i q_j$.
- When the dependence P_{12} is unknown, the p_{ij}'s are unknown. How to find them? $\rightarrow$ solutions in next section!

Outline

Typology of approaches

- This general combination scheme allows us to account for a wide range of assumptions about the reliability and dependence of pieces of evidence about a variable X.
- It does not however indicate which assumptions to make.

$\rightarrow$ Means to determine them.

- Two possible situations:
  - The only available information are the pieces of evidence received.

  $\rightarrow$ Uncertainty principle-based selection.
  - Similar pieces of evidence were received previously and the actual value of X was observed, allowing us to assess the effectiveness of assumptions.

  $\rightarrow$ Performance-based selection.
- Both approaches will be illustrated on two cases:

  $P1$ Independence assumed, (partially[1]) unknown reliability;

  $P2$ Reliability assumed, (totally) unknown dependence.

[1] Typically, a set of candidate assumptions about the reliability is considered.

Outline

Approach

- Refresher: if a mass function m has to be replaced by another chosen in a set $\mathcal{M}$ of mass functions that are more specialized than m, using some uncertainty measure U, then if $U = I$ (imprecision), choose $\operatorname{argmax}_{m \in \mathcal{M}} I(m)$, and if $U = C$ (conflict), choose $\operatorname{argmin}_{m \in \mathcal{M}} C(m)$. Moreover, in the particular case where $\mathcal{M}$ is a chain, its $\sqsubseteq$-greatest element is a solution to these two optimization problems.
- The solutions to problems $P1$ and $P2$ presented in the following are essentially instantiations of this general principled procedure.
- They were originally proposed in [Pichon et al., 2015] and [Destercke et al., 2007], respectively.

Unknown reliability

(independence assumed)

- When several messages representing independent pieces of evidence and inducing by mass functions $m_1, \ldots, m_r$ are received, they are classically first combined using the assumption that they are all reliable.
- This assumption induces no transformation/altering of each of the pieces of evidence, i.e., they are accepted as they are. It is a natural default assumption.
- Let $m_0^{\mathcal{R}}$ be the mass function on the messages' reliabilities representing this assumption and let ${}^{\mathcal{R}}m_0 = \bigcirc\!\!\!\!\cap_{i=1}^{r} m_i$ denote the initial state of knowledge it induces about X.
- The validity of this assumption may be assessed through $C({}^{\mathcal{R}}m_0)$ for some conflict measure C.

Tolerable inconsistency

- Specifically, assume some threshold δ of tolerable inconsistency in a piece of evidence modeled by a mass function m
  - below this threshold, the assumptions leading to m may be deemed valid (and Dempster's normalization may safely be used to resolve the inconsistency)
  - above this threshold, such assumptions cannot be considered valid (and normalization becomes too hazardous a solution to be used to resolve the inconsistency)
- Hence, if $C({}^{\mathcal{R}}m_0) \leq \delta$, then the assumption of reliability may be deemed valid and our state of knowledge about X may be represented by ${}^{\mathcal{R}}m_0^* = \oplus_{i=1}^{r} m_i$.
- However, if $C({}^{\mathcal{R}}m_0) > \delta$, then the assumption of reliability is not tenable. In such a case, weaker assumptions, inducing less inconsistent knowledge states, are typically considered..

Weaker valid assumptions

- More precisely, we typically consider a set of assumptions $\mathcal{M}^{\mathcal{R}} = \{m_1^{\mathcal{R}}, \ldots, m_s^{\mathcal{R}}\}$ on the messages' reliabilities, such that

$$^{\mathcal{R}}m_i \sqsubseteq {}^{\mathcal{R}}m_{i+1}$$

with $^{\mathcal{R}}m_i$ the result of combining $m_1, \ldots, m_r$ according to assumption $m_i^{\mathcal{R}}$, $i = 0, \ldots, s$.

- Given the property, $m \sqsubseteq m' \Rightarrow C(m) \geq C(m')$, we know that such a set of assumptions allows to decrease inconsistency.
- Given the threshold δ, we may reduce this set to that of assumptions deemed valid, i.e.,

$$\mathcal{M}^{\mathcal{R}}_{\leq\delta} = \{m_i^{\mathcal{R}}, i = k, \ldots, s\},$$

with $k = \min\{i \in \{1, \ldots, s\} | C(^{\mathcal{R}}m_i) \leq \delta\}$.

Assumption selection

- $\mathcal{M}^{\mathcal{R}}_{\leq\delta}$ is our set of candidate reliability assumptions, in which to seek a better assumption than $m^{\mathcal{R}}_0$.
- To choose among the assumptions in $\mathcal{M}^{\mathcal{R}}_{\leq\delta}$, we can remark that choosing one assumption means that the initial (default) knowledge ${}^{\mathcal{R}}m_0$ about X will be replaced by a mass function in the set
$$\mathcal{M}_{\leq\delta} = \{{}^{\mathcal{R}}m_i, i = k, \ldots, s\},$$
such that ${}^{\mathcal{R}}m_0 \sqsubseteq {}^{\mathcal{R}}m_i$ for all $i = k, \ldots, s$, and ${}^{\mathcal{R}}m_i \sqsubseteq {}^{\mathcal{R}}m_{i+1}$ for all $i = k, \ldots, s-1$.
- We recognize a situation discussed previously: $\mathcal{M}_{\leq\delta}$ is a chain of less specialized mass functions than ${}^{\mathcal{R}}m_0$, hence we should select the $\sqsubseteq$-least informative element in $\mathcal{M}_{\leq\delta}$, which is ${}^{\mathcal{R}}m_k$, and thus the assumption $m^{\mathcal{R}}_k$.
- The final state of knowledge about X, given $m_1, \ldots, m_r$, and using this reasoning, is then ${}^{\mathcal{R}}m^*_k$.

Examples

- This general approach subsumes sequential discounting, which is a classical approach (used in several fusion schemes) for the combination of potentially conflicting mass functions.
- Sequential discounting amounts to considering a set $\mathcal{M}^{\mathcal{R}}$ such that $m_i^{\mathcal{R}}$ is the assumption that message j, $j = 1, \ldots, r$, has independent probability α_j^i of being unreliable, hence ${}^{\mathcal{R}}m_i = \bigcirc\!\!\!\!\cap_{j=1}^{r} {}^{\alpha_j^i}m_j$, and such that $\alpha_j^i \leq \alpha_j^{i+1}$, $i = 1, \ldots, s$.
- It can be instantiated with other sets of sensible reliability assumptions, such as the set of assumptions corresponding to the *Q*-relaxation rule, for $1 \leq Q \leq r - 1$ (${}^{\mathcal{R}}m_i$ is then the assumption that $r - i$ of the messages are reliable).

Application to nuclear reactor safety

- Project BEMUSE of the Nuclear Energy Agency.
- $r = 10$ 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 $\mathcal{M}^{\mathcal{R}}$ with ${}^{\mathcal{R}}m_i$ the assumption that $r - i$ of the sources are reliable.
- Used strong conflict for measure C ($C(m) = 1 - \max_{\theta \in \Theta} c(\theta)$).
- PCT2 parameter with domain $\Theta = \{\theta_1, \ldots, \theta_6\}$
  - $C({}^{\mathcal{R}}m_0) = 0.81$ (all sources reliable)
  - $C({}^{\mathcal{R}}m_1) = 0.19$ (9 out of 10 reliable)
  - $C({}^{\mathcal{R}}m_2) = 0$ (8 out of 10 reliable)
  - Values θ_4 and θ_5 are definitely more plausible.

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

Unknown independence

(reliability assumed)

- Let $(\Omega_1, P_1, \Gamma_1)$ and $(\Omega_2, P_2, \Gamma_2)$ be two messages representing two pieces of evidence about $X \in \Theta$ and inducing mass functions m_1 and m_2, respectively.
- Refresher: if the messages are assumed reliable and to have some dependence structure described by the joint probability distribution P_{12} on $\Omega_1 \times \Omega_2$, then the induced knowledge $m_{1\cap 2}$ about X satisfies

$$m_{1\cap 2}(A) = \sum_{B_i \cap C_j = A} p_{ij}, \quad \forall A,$$

with

$$p_{ij} = P_{12}(\{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \Gamma_1(\omega_1) = B_i, \Gamma_2(\omega_2) = C_j\}).$$

- When the dependence P_{12} is unknown, the p_{ij}'s are unknown. How to find them?

First solution

Entropy maximization

- Refresher: given two marginal probability distributions $P_1^{\Omega_1}$ and $P_2^{\Omega_2}$, applying the maximum uncertainty principle to find the joint distribution $P_{12}^{\Omega_{12}}$ yields, using Shannon entropy as an uncertainty measure, $P_{12}^{\Omega_{12}} = P_1^{\Omega_1} \times P_2^{\Omega_2}$ and thus

$$p_{ij} = p_i q_j, \quad \forall (i, j)$$

  with $p_i = m_1(B_i)$ and $q_j = m_2(C_j)$ for mass functions m_1 and m_2 with focal sets $B_1, \ldots, B_r$ and $C_1, \ldots, C_s$, respectively.
- In other words, we recover the independence assumption and thus Dempster's rule (when normalization is used subsequently to specifying the dependence).
- It can be seen as a reassuring argument in favor of using Dempster's rule when the dependence is unknown. Yet it is not totally satisfying as it does not take into account the structure of the focal sets, in particular their precision and/or their consistency.

Second solution

Conjunctive revision

- A more satisfying solution is obtained by remarking that whatever the dependence, $m_{1\cap 2}$ is a specialization of both m_1 and m_2.
- Formally, let

$$\begin{aligned}\mathcal{M}_{1\cap 2} &= \{m_{1\cap 2}|\ m_{1\cap 2}(A) = \sum_{B_i \cap C_j = A} p_{ij}, \forall A; \sum_{i,j} p_{ij} = 1; \\ &\quad \sum_j p_{ij} = p_i, i = 1,\ldots,r; \sum_i p_{ij} = q_j, j = 1,\ldots,s\}.\end{aligned}$$

We have $m_{1\cap 2} \sqsubseteq m_i$, $i = 1, 2$, for all $m_{1\cap 2} \in \mathcal{M}_{1\cap 2}$.

- This means, in particular, that for any $m_{1\cap 2} \in \mathcal{M}_{1\cap 2}$, there exists a specialization matrix S such that $m_{1\cap 2} = S \cdot m_1$.

$\rightarrow$ The combination setting considered produces a conjunctive revision [Smets, 2002] of some (initial) knowledge state m_1 into a new knowledge state $m_{1\cap 2} \in \mathcal{M}_{1\cap 2}$, given some received evidence m_2.

Second solution

Imprecision maximization

- We recognize a situation discussed previously: $\mathcal{M}_{1\cap 2}$ is a set of more specialized mass functions than m_1.
- Using an imprecision measure I such as the cardinality to select a mass function in this set leads to the following linear optimization problem

$$\max_{p_{ij}} \sum_{(i,j)|B_i\cap C_j\neq\emptyset} p_{ij}|B_i \cap C_j|$$

under the constraints $\sum_{i,j} p_{ij} = 1$ and

$$\begin{aligned} \sum_j p_{ij} &= p_i, \quad i = 1,\ldots,r, \\ \sum_i p_{ij} &= q_j, \quad j = 1,\ldots,s. \end{aligned}$$

- The mass function found by solving this problem can be normalized.

Second solution

Remarks

- This solution was first studied in [Destercke et al., 2007].
- Alternatively, using a conflict measure C to select a mass function in $\mathcal{M}_{1\cap 2}$, it would be justified to look for the least conflicting mass function $m_{1\cap 2}$ in $\mathcal{M}_{1\cap 2}$.
- Taking into account both the inconsistency (minimization) and imprecision (maximization) objectives may be done in several ways.
- For instance, [Cattaneo, 2003] proposes to combine them in a single function $f(m) = (1 - C(m))I(m) - C(m)|\Theta|$ to be maximized, with C the conflict and I the nonspecificity measures.

Outline

1. Computation of Dempster's rule
   - Expressions
   - Exact
   - Approximate
2. Extension of Dempster's rule
   - Forms of unreliability for a piece of evidence
   - Partially reliable pieces of evidence
   - Dependent pieces of evidence
3. Rule selection
   - Uncertainty principle-based
   - Performance-based

Approach

- Consider a system which outputs for a given object o, an estimate of the actual value x^* of some feature $X \in \Theta$ of o.
- To produce this output, the system uses internally a combination of some pieces of evidence. This combination involves some reliability and dependence assumptions. Assume such assumptions can be characterized by some parameters $\boldsymbol{\lambda}$.
- The output for object o may thus be noted $f(o; \boldsymbol{\lambda})$.
- Assume a set of ℓ objects for the which the true value of X is known, i.e., $\{x_i^*\}_{i=1}^{\ell}$ is available.
- Assume outputs $\{f(o_i; \boldsymbol{\lambda})\}_{i=1}^{\ell}$ may be obtained for any $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$.
- The $\hat{\boldsymbol{\lambda}}$ to be used to produce the output for a new object may then be chosen as the one in $\boldsymbol{\Lambda}$ minimizing the average loss

$$J(\boldsymbol{\lambda}) = \frac{1}{n} \sum_{i=1}^{\ell} \mathcal{L}(f(o_i; \boldsymbol{\lambda}), x_i^*)$$

for some loss function $\mathcal{L}(f(o; \boldsymbol{\lambda}), x^*)$.

Evidential classification

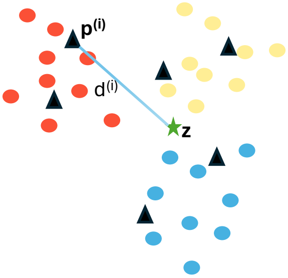
- An evidential classification system fits this general approach.
- It is a system that, given a feature vector $\mathbf{z}$ of an instance, produces a mass function $m_{\mathbf{z}}$ wrt its unknown class $X \in \Theta$.
- Two main kinds of evidential classification systems can be distinguished:
  1. classifier fusion-based;
  2. evidence-theoretic classifiers;
- In the first kind, the combination is at play to combine the outputs of some classifiers, see, e.g., [Huang et al., 2025] and [Quost et al., 2011] for works addressing problems $P1$ (unknown reliability) and $P2$ (unknown dependence), respectively, by following this general approach.
- Among evidence-theoretic classifiers, some, such as the Evidential Neural Network (ENN) classifier, are based on an explicit combination of pieces of evidence.

$\rightarrow$ Two recent examples of evidential classifiers where addressing $P1$ and $P2$ are involved (Serigne Diène ongoing PhD work)

Unknown reliability

Contextual Discounting ENN classifier [Diène et al., 2025]

- The learning set is summarized by ℓ prototypes.
- Each prototype $\mathbf{p}^{(i)}$ has probability $u_q^{(i)}$ of belonging to class θ_q.
- Each prototype $\mathbf{p}^{(i)}$ is a piece of evidence about the class of $\mathbf{z}$; its reliability $s_q^{(i)}$ for class θ_q decreases with the distance $d^{(i)}$ between $\mathbf{z}$ and $\mathbf{p}^{(i)}$ (the decrease is more or less rapid depending on the class).



Computation of the output of the CD-ENN classifier

- Prototype $\mathbf{p}^{(i)}$ induces mass function ${}^{\mathcal{C}}m^{(i)}$ about the class of $\mathbf{z}$, which is the contextual discounting with discount rate vector $(1 - s_1^{(i)}, \ldots, 1 - s_c^{(i)})$ of the (Bayesian) mass function $m^{(i)}(\{\theta_q\}) = u_q^{(i)}$ where

$$s_q^{(i)} = \alpha^{(i)} \exp\left(-\gamma_q^{(i)} d_i^2\right).$$

- The output mass function $m_{\mathbf{z}}$ of CD-ENN is then:

$$m_{\mathbf{z}} = \oplus_{i=1}^{\ell} {}^{\mathcal{C}}m^{(i)}$$

Learning of the parameters

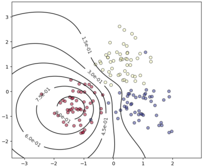
- The reliability parameters are
  - The prototypes $\mathbf{p}^{(i)}$, $i = 1, \ldots, \ell$
  - The $\alpha^{(i)}$ and decay rates $\gamma_q^{(i)}$, $i = 1, \ldots, \ell$, $q = 1, \ldots, c$
- Additional parameters: probabilities $u_q^{(i)}$, $i = 1, \ldots, \ell$, $q = 1, \ldots, c$.
- Estimating these parameters using the cross-entropy of the normalized contour function $p_{m_{\mathbf{z}^{[i]}}}$ of the output mass function $m_{\mathbf{z}^{[i]}}$, i.e., minimizing

$$-\sum_{i=1}^{n} \ln(p_{m_{\mathbf{z}^{[i]}}}(y^{[i]})),$$

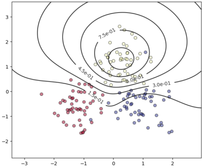
  where $y^{[i]}$ is the true class of $\mathbf{z}^{[i]}$, is linear in the number of classes.

Example and comparison with ENN (singletons and Θ)

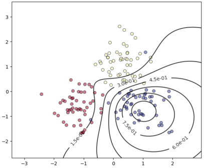
ENN



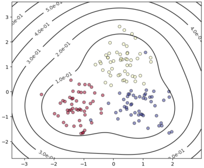
$m(\{\theta_1\})$



$m(\{\theta_2\})$

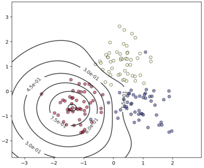


$m(\{\theta_3\})$

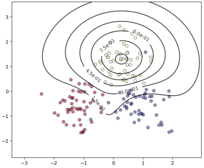


$m(\Theta)$

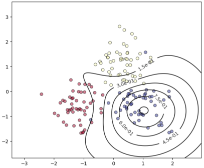
CD-ENN



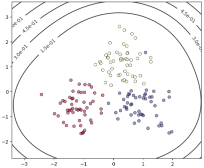
$m(\{\theta_1\})$



$m(\{\theta_2\})$

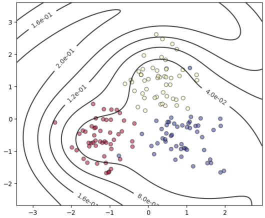


$m(\{\theta_3\})$

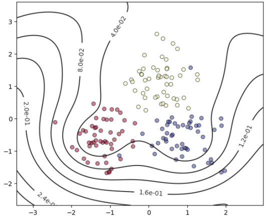


$m(\Theta)$

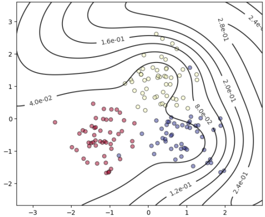
Additional masses generated by CD-ENN



$m(\{\theta_1, \theta_2\})$



$m(\{\theta_1, \theta_3\})$



$m(\{\theta_2, \theta_3\})$

Experiments

GNLL[2] for ENN and CD-ENN

Datasets	ENN(↓)	CD-ENN(↓)
Pima	3.06 ± 0.47	$\mathbf{0.593} \pm 0.13$
Wine red	1.04 ± 0.03	$\mathbf{1.005} \pm 0.02$
Wine white	1.14 ± 0.01	$\mathbf{1.12} \pm 0.01$
Dry Bean	0.64 ± 0.03	$\mathbf{0.63} \pm 0.02$
Ecoli	$\mathbf{0.57} \pm 0.08$	0.66 ± 0.06
Glass	1.41 ± 0.17	$\mathbf{1.27} \pm 0.06$
Heart	1.01 ± 0.32	$\mathbf{0.66} \pm 0.04$
Ionosphere	3.1 ± 0.52	$\mathbf{2.97} \pm 0.99$
Vertebral	$\mathbf{0.68} \pm 0.06$	0.73 ± 0.07
Sonar	1.06 ± 0.38	$\mathbf{0.85} \pm 0.07$

[2] Instance **z** with true class y, GNLL is: $-(1/2)\ln bel_{\mathbf{z}}(\{y\}) - (1/2)\ln pl_{\mathbf{z}}(\{y\})$.

Unknown dependence

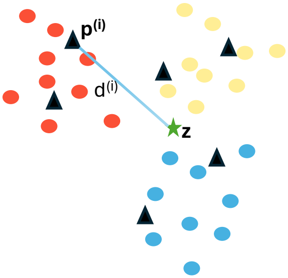
r-sum of two simple mass functions

- Let $(\Omega_1, P_1, \Gamma_1)$ and $(\Omega_2, P_2, \Gamma_2)$ be two messages representing reliable and elementary pieces of evidence about X, inducing simple mass functions $m_1 = A_1^{d_1}$ and $m_2 = A_2^{d_2}$, respectively.
- Given d_1 and d_2, the joint distribution P_{12} on $\Omega_1 \times \Omega_2$, and thus their dependence, can be completely characterized by a correlation $r \in [-1, 1]$ ($r = 0$ corresponds to independence).
- Assume some $r \in [-1, 1]$ specifying the dependence between the messages.
- Then, knowledge about X given such messages is represented by the mass function induced by the random set $(\Omega_1 \times \Omega_2, P_\cap^r, \Gamma_\cap)$, with $P_\cap^r$ the result of conditioning P_{12} on $\Theta_\cap$, denoted $A_1^{d_1} \oplus_r A_2^{d_2}$ and called the r-sum of $A_1^{d_1}$ and $A_2^{d_2}$.
- Binary operation $\oplus_r$ is a generalization of Dempster's rule for the combination of two simple mass functions ($\oplus$ recovered for $r = 0$).

Dependence-aware evidential radial basis function network (r-ERBFN) classifier [Pichon et al., 2024]

Binary classification case $\Theta = \{\theta_1, \theta_2\}$

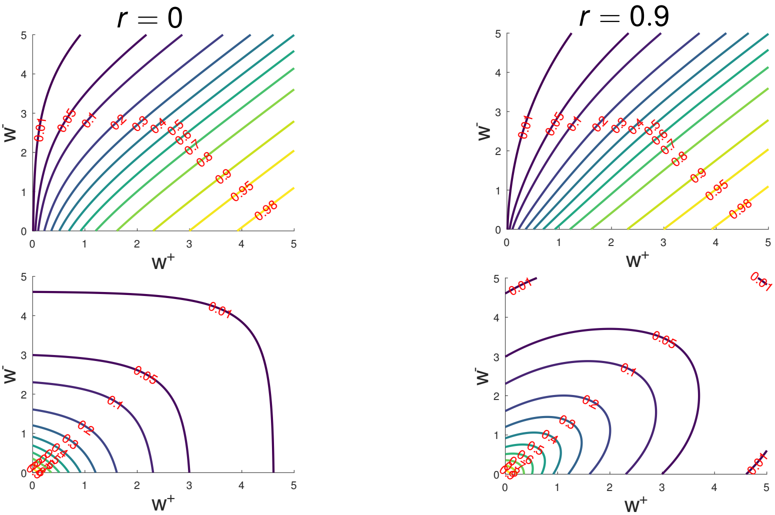
- Similarly as for CD-ENN, ℓ prototypes.
- Each prototype $\mathbf{p}^{(i)}$ has a parameter $v^{(i)} \in \mathbb{R}$.
- Each prototype $\mathbf{p}^{(i)}$ is a piece of evidence about the class of $\mathbf{z}$: it supports class θ_1 if $v^{(i)} \geq 0$, θ_2 otherwise, such support decreasing with distance $d^{(i)}$ and increasing with $|v^{(i)}|$



Computation of the output of the r-ERBFN classifier

- Prototype $\mathbf{p}^{(i)}$ induces mass function $m^{(i)}$ about the class of $\mathbf{z}$, such that
$$m^{(i)} = \{\theta^{(i)}\}^{\exp(-w^{(i)})}$$
with $\theta^{(i)} = \theta_1$ if $v^{(i)} \geq 0$ and $\theta^{(i)} = \theta_2$ otherwise, and where $w^{(i)} = s^{(i)}|v^{(i)}|$ with $s^{(i)} = \exp\left(-\gamma^{(i)} d_i^2\right)$.
- The evidence supporting θ_1 over all prototypes is
$$\begin{aligned} m^+ &= \oplus_{i:v^{(i)}\geq 0} m^{(i)} \\ &= \{\theta_1\}^{\exp(-w^+)} \end{aligned}$$
with $w^+ = \sum_{i:v^{(i)}\geq 0} w^{(i)}$.
- Similarly, the overall evidence supporting θ_2 is $m^- = \{\theta_2\}^{\exp(-w^-)}$ with $w^- = \sum_{i:v^{(i)}<0} w^{(i)}$
- The output mass function $m_{\mathbf{z}}$ of r-ERBFN is then:
$$m_{\mathbf{z}} = m^+ \oplus_r m^-$$

$m_{\mathbf{z}}(\{\theta_1\})$ (top) and $m_{\mathbf{z}}(\Theta)$ (bottom) vs (w^+, w^-)

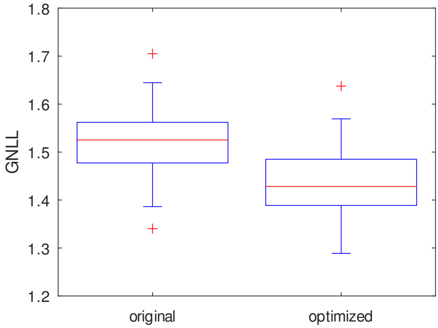


Learning of the parameters

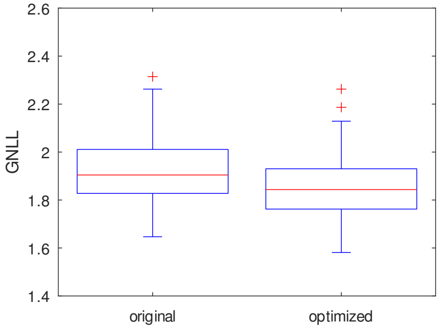
- The dependence parameter is the correlation r (one correlation per class if multiclass).
- Other ('RBFN') parameters:
  - The prototypes $\mathbf{p}^{(i)}$, $i = 1, \ldots, \ell$
  - The parameters $\gamma^{(i)}$ and $v^{(i)}$, $i = 1, \ldots, \ell$
- If these other parameters are identified to that of a trained RBFN with a logistic output unit, then the normalized contour function $p_{m_{\mathbf{z}}}$ of the output mass function $m_{\mathbf{z}}$ is nothing but the probabilistic output of this trained network.
- The dependence parameter can then be learnt by minimizing the GNLL.

Experiments

Binary classification



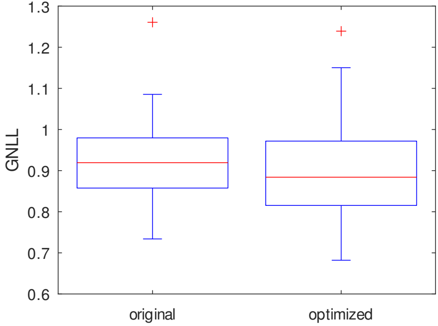
Pima



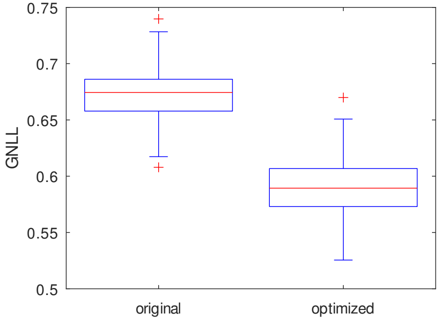
Ionosphere

Experiments

Multi-class classification



Glass



Vowel

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.
- It is adapted when the pieces of evidence are reliable and independent.
- There exist alternative and sound combination rules, corresponding to other assumptions.
- If unknown, the reliability and dependence of the pieces of evidence can be determined using several means, depending on the available information.
- Partially reliable and dependent pieces of evidence about continuous variables in [Denœux, 2024].

References I

Dempster's rule

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.

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.

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.

References III

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.

Extension of Dempster's rule

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.

References IV

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

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.

D. Mercier, B. Quost, and T. Denœux.
Refined modeling of sensor reliability in the belief function framework using contextual discounting.
Information Fusion, 9(2):246–258, 2008.

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.

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 V

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.

T. Denœux.
Combination of dependent and partially reliable Gaussian random fuzzy numbers.
Information Sciences, 681:121208, 2024.

N. Pellicanò, S. Le Hégarat-Mascle, and E. Aldea
2CoBel: A scalable belief function representation for 2D discernment frames.
Int. J. of Approximate Reasoning, 103:320–342, 2018.

G. Shafer.
Probability judgment in artificial intelligence.
Uncertainty in Artificial Intelligence, Machine Intelligence and Pattern Recognition, 4:127–135, 1986.

H. Zhang, B. Quost, and M.-H. Masson.
Cautious weighted random forests.
Expert Systems With Applications, 213:118883, 2023.

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.
A neural network classifier based on Dempster-Shafer theory.
IEEE Trans. on Syst., Man, and Cybernetics - Part A, 30(2):131–150, 2000.

References VI

Rule Selection

S. Destercke, D. Dubois and E. Chojnacki.
Cautious conjunctive merging of belief functions.
Proc. European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2007), Lecture Notes in Computer Science, 4724:332–343, Springer, 2007.

M. Cattaneo.
Combining belief functions issued from dependent sources.
Proc. Third International Symposium on Imprecise Probabilities and Their Application (ISIPTA'03), Lugano, Switzerland, pages 133–147, 2003.

F. Pichon, S. Diène, T. Denœux, S. Ramel, and D. Mercier.
r-ERBFN : an Extension of the Evidential RBFN Accounting for the Dependence Between Positive and Negative Evidence.
16th International Conference on Scalable Uncertainty Management (SUM 2024), Palermo, Italy, November 27–29, 2024, volume 15350 of Lecture Notes in Artifical Intelligence, pages 354-368, Springer, 2024.

S. Diène, S. Ramel, F. Pichon, and D. Mercier.
An Extension of the Evidential Neural Network Classifier based on Contextual Discounting.
37th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2025), Athens, Greece, November 03–05, 2025, accepted for publication, 2025.

P. Smets.
The application of the matrix calculus to belief functions.
International Journal of Approximate Reasoning 31:1–30, 2002.

References VII

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.

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.

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

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.

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.

References VIII

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.

L. Huang, S. Ruan, P. Decazes, and T. Denœux.
Deep evidential fusion with uncertainty quantification and reliability learning for multimodal medical image segmentation.
Information Fusion, Vol. 113, 102648, 2025.

Thank you for your attention.

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 defined on $\Theta = \{p, c, h, o\}$ such that

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

- We proceed in four steps:
  1. input m_1 and m_2 using the "focal set format";
  2. compute $m_1 \textcircled{\scriptsize$\cap$} m_2$ using the commonality-based approach;
  3. compute $m_1 \oplus m_2$ by normalizing $m_1 \textcircled{\scriptsize$\cap$} m_2$;
  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 $\Theta = \{\theta_1, \ldots, \theta_p\}$, 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 $\mathbf{F}$ is a $r \times p$ binary matrix such that

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

- This format is convenient to input a mass function.

Step 1: input m_1 and m_2

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

```
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., $m_2(\{c, h\}) = 0.8, m_2(\Theta) = 0.2$.

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

Step 2: compute $m_1 \textcircled{\cap} m_2$ using the commonality-based approach

- A commonality function q is specified with the $2^{|\Theta|}$ numbers $q(A), A \subseteq \Theta$.
- It can be represented by a $2^{|\Theta|}$-dimensional column vector $\mathbf{q}$ whose element j stores $q(A_j)$ with A_j the subset of Θ such that $\theta_i \in A_j$ if the i-th bit in the binary representation of $j-1$ equals 1.
- Example for $\Theta = \{\theta_1, \theta_2, \theta_3\}$

Position	$\theta_3\ \theta_2\ \theta_1$	$\mathbf{q}$
1	0 0 0	$q(\emptyset)$
2	0 0 1	$q(\{\theta_1\})$
3	0 1 0	$q(\{\theta_2\})$
4	0 1 1	$q(\{\theta_1, \theta_2\})$
5	1 0 0	$q(\theta_3)$
6	1 0 1	$q(\{\theta_1, \theta_3\})$
7	1 1 0	$q(\{\theta_2, \theta_3\})$
8	1 1 1	$q(\{\theta_1, \theta_2, \theta_3\})$

Step 2: compute $m_1 \textcircled{\cap} m_2$ 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 $2^{|\Theta|}$-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 $m_1 \textcircled{\cap} m_2$ 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 $q_{1 \textcircled{\cap} 2}$

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

- Computation of $m_{1 \textcircled{\cap} 2}$, 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 $m_{1 \cap\!\!\!\bigcirc} m_2$

- 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
```