Reliability and Dependence in the Combination of Evidence

Frédéric Pichon

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

SIPTA Seminar
December 12, 2024

Contents

- Dempster's rule of combination is the cornerstone of Shafer's theory of evidence [Shafer, 1976].
- It allows the combination of independent and reliable pieces of evidence.
- This talk: extensions of Dempster's rule allowing us to account for various assumptions with respect to the reliability and dependence of the pieces of evidence

Contents

- Dempster's rule of combination is the cornerstone of Shafer's theory of evidence [Shafer, 1976].
- It allows the combination of independent and reliable pieces of evidence.
- This talk: extensions of Dempster's rule allowing us to account for various assumptions with respect to the reliability and dependence of the pieces of evidence
  - A general approach for the fusion of independent pieces of evidence, which permits refined forms of the lack of reliability;
  - A means to specify the dependence when combining reliable pieces of evidence, in the particular yet important case where they are elementary (i.e., represented by simple mass functions).
  - Some theoretical and practical interests of these extensions.

Outline

Outline

Mass function

- A piece of evidence about a variable X taking values in a finite set $\Theta = \{\theta_1, \ldots, \theta_K\}$ (frame of discernment) is represented by a mass function $m : 2^\Theta \to [0, 1]$ such that $m(\emptyset) = 0$ and

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

- Mass $m(A)$ represents the probability that the evidence supports exactly the proposition $X \in A$.
- Any $A \subseteq \Theta$ such that $m(A) > 0$ is a focal set of m.
- m is said to be:
  - non dogmatic if Θ is a focal set;
  - vacuous if Θ is the only focal set (total ignorance);
  - Bayesian if its focal sets are singletons (probability distribution);
  - simple if it has two focal sets: Θ and A for some $A \subset \Theta$.

Semantics

[Shafer, 1981]

- Suppose we receive an encoded message about X.
- The actual code used is unknown, but we know :
  - it was one in a finite set Ω;
  - the chance $P(\omega)$ of each code $\omega \in \Omega$ being selected.
- Furthermore, we know that the meaning of the message is $X \in \Gamma(\omega)$, with $\Gamma(\omega)$ a nonempty subset of Θ, if code ω was used.
- The tuple (Ω, P, Γ) represents then the available information.
- The probability that the message means $X \in A$ is:

$$m(A) := P(\{\omega \in \Omega : \Gamma(\omega) = A\}), \quad \forall A \in 2^{\Theta} \setminus \{\emptyset\}.$$

$\rightarrow$ A mass function is obtained by fitting a piece of evidence to such message (Ω, P, Γ).

Semantics

[Shafer, 1981]

- Suppose we receive an encoded message about X.
- The actual code used is unknown, but we know :
  - it was one in a finite set Ω;
  - the chance $P(\omega)$ of each code $\omega \in \Omega$ being selected.
- Furthermore, we know that the meaning of the message is $X \in \Gamma(\omega)$, with $\Gamma(\omega)$ a nonempty subset of Θ, if code ω was used.
- The tuple (Ω, P, Γ) represents then the available information.
- The probability that the message means $X \in A$ is:

$$m(A) := P(\{\omega \in \Omega : \Gamma(\omega) = A\}), \quad \forall A \in 2^{\Theta} \setminus \{\emptyset\}.$$

$\rightarrow$ A mass function is obtained by fitting a piece of evidence to such message (Ω, P, Γ).

- Remark: (Ω, P, Γ) is formally a random set.

Dempster's rule

- Let $(\Omega_1, P_1, \Gamma_1)$ and $(\Omega_2, P_2, \Gamma_2)$, with $\Gamma_i : \Omega_i \rightarrow 2^\Theta \backslash \{\emptyset\}$, $i = 1, 2$, 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., the chance $P_{12}(\omega_1, \omega_2)$ that the pair of codes $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$ was chosen is equal to $P_1(\omega_1) \cdot P_2(\omega_2)$.
- Assume further that they are reliable: if the actual codes were ω_1 and ω_2, we know for sure that $X \in \Gamma_\cap(\omega_1, \omega_2) := \Gamma_1(\omega_1) \cap \Gamma_2(\omega_2)$
  - if $\Gamma_\cap(\omega_1, \omega_2) = \emptyset$, then we know that (ω_1, ω_2) could not be the pair of codes actually used.

$\rightarrow$ We must condition the chance distribution on the event $\Theta_\cap = \{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \Gamma_\cap(\omega_1, \omega_2) \neq \emptyset\}$.

Dempster's rule (continued)

- Let $P_\cap$ be the probability measure on $\Omega_1 \times \Omega_2$ resulting from the conditioning of P_{12} on the event $\Theta_\cap$.
- Under the assumptions that the pieces of evidence represented by mass functions m_1 and m_2 are independent and reliable, our knowledge about X is represented by the mass function denoted $m_1 \oplus m_2$, called the orthogonal sum of m_1 and m_2, and induced by the random set $(\Omega_1 \times \Omega_2, P_\cap, \Gamma_\cap)$, i.e., the probability of knowing that $X \in A$ is

$$(m_1 \oplus m_2)(A) := P_\cap(\{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \Gamma_\cap(\omega_1, \omega_2) = A\}).$$

- The orthogonal sum is well defined if $P_{12}(\Theta_\cap) > 0$.
- It is easy to show that

$$(m_1 \oplus m_2)(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - \sum_{B \cap C = \emptyset} m_1(B) m_2(C)}, \quad \forall A \in 2^\Theta \setminus \{\emptyset\}.$$

- The binary operation $\oplus$ is called Dempster's rule.

Outline

Outline

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 ten-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 $P^{\mathcal{R}}(unrel) = \alpha$.
- What can then be inferred about X?
- If the actual code was $\omega \in \Omega$ and
  - the message is reliable, we know that $X \in \Pi_{\Gamma(\omega)}(rel)$
  - the message is unreliable, we know that $X \in \Pi_{\Gamma(\omega)}(unrel) = \Theta$
- Hence, the probability to know $X \in A \subset \Theta$ is

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

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 nothing but the result of Shafer's discounting with discount rate α of mass function m.

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 nothing but the result of Shafer's discounting with discount rate α of mass function m.
- Unreliability can be refined into contextual unreliability, leading to a more general model that includes also the contextual discounting of [Mercier et al., 2008].

Truthfulness

- 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 $\Pi_A^{\mathcal{T}} : \mathcal{R}^{\mathcal{T}} \rightarrow 2^{\Theta}$ 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

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

where
  - $\Gamma^{\boldsymbol{\mathcal{R}}}(\omega_1, \omega_2, \mathbf{r}) := \Gamma_1^{\mathcal{R}}(\omega_1, r_1) \cap \Gamma_2^{\mathcal{R}}(\omega_2, r_2)$ for all $\mathbf{r} = (r_1, r_2) \in \boldsymbol{\mathcal{R}}$
  - $P_{\boldsymbol{\mathcal{R}}}$: probability measure $P_{12} \times P^{\boldsymbol{\mathcal{R}}}$ conditioned on $\Theta_{\boldsymbol{\mathcal{R}}} = \{(\omega_1, \omega_2, \mathbf{r}) \in \Omega_1 \times \Omega_2 \times \boldsymbol{\mathcal{R}} : \Gamma^{\boldsymbol{\mathcal{R}}}(\omega_1, \omega_2, \mathbf{r}) \neq \emptyset\}$

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 reliabilities

$\rightarrow$ Discount and combine

- $\alpha m_1 + (1-\alpha)m_2$ 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

Imprecise reliability

- Assume the reliability is known in the form of $\mathbf{R} \subseteq \mathcal{R}$.
- Then we obtain 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 $\Theta_{\mathbf{R}} = \{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \Gamma_{\mathbf{R}}(\omega_1, \omega_2) \neq \emptyset\}$
- Particular cases:
  - $m_1 \textcircled{\cup} m_2$ for $\mathbf{R} = \{(rel_1, rel_2), (rel_1, unrel_2), (unrel_1, rel_2)\}$.

  $\rightarrow$ Disjunctive rule
  - $\mathbf{R} =$ "$N - Q$ out of the N messages are reliable".

  $\rightarrow$ Q-relaxation rule

Imprecise reliability

- Assume the reliability is known in the form of $\mathbf{R} \subseteq \mathcal{R}$.
- Then we obtain 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 $\Theta_{\mathbf{R}} = \{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \Gamma_{\mathbf{R}}(\omega_1, \omega_2) \neq \emptyset\}$
- Particular cases:
  - $m_1 \cup\!\!\!\!\bigcirc m_2$ for $\mathbf{R} = \{(rel_1, rel_2), (rel_1, unrel_2), (unrel_1, rel_2)\}$.

  $\rightarrow$ Disjunctive rule
  - $\mathbf{R} =$ "$N - Q$ out of the N messages are reliable".

  $\rightarrow$ Q-relaxation rule
- 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.

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

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

→ 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 (continued)

- 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.
  - $P_{\mathcal{B}}$: probability measure $P_{12} \times P^{\mathcal{B}}$ conditioned on $\Theta_{\mathcal{B}} = \{(\omega_1, \omega_2, b) \in \Omega_1 \times \Omega_2 \times \mathcal{B} : \Gamma^{\mathcal{B}}(\omega_1, \omega_2, b) \neq \emptyset\}$.

Theorem

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

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

$\rightarrow$ Prism to understand and select alternatives to Dempster's rule

Applications

- Alternatives 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) classification rule [Denœux, 1995]
  - 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

- $m_1 \oplus m_2$ relies 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]: $\Omega_i = \{0, 1\}$
  - $\Gamma_1(0) = A$, $\Gamma_1(1) = \Theta$ and $P_1(1) = 0.2$
  - $\Gamma_2(0) = \overline{A}$, $\Gamma_2(1) = \Theta$ and $P_2(1) = 0.01$
  - Let S_i be the random variable, with state space Ω_i, representing the interpretation for the i-th message. Dependence specified by $P_{12}(S_1 = 1|S_2 = 1) = 0.9$.
  - We have $P_{12} \neq P_1 \times P_2$.
- Remark: it is an example of non independence between messages inducing simple mass functions.

$\rightarrow$ Focus on the combination of such dependent and elementary pieces of evidence (assumed throughout to be reliable)

Why such focus?

- Recall that 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

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

for some $d \in [0, 1]$.
  - 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.
- It is the simplest kind of evidence [Shafer, 1976].
  - [Shafer, 1976] and [Smets, 1995] proposed solutions to view belief functions as resulting (in part) from the combination of such elementary pieces of evidence, assumed reliable and independent.
  - In applications, belief functions often result from such combination.

$\rightarrow$ Important both theoretically and in practice.

Outline

Setting

- Let $m_i, i = 1, \ldots, N$, be simple mass functions.
- Mass function m_i is induced by message $(\Omega_i, P_i, \Gamma_i)$ with $\Omega_i = \{0, 1\}, P_i(1) = d_i$, and

$$\begin{aligned} \Gamma_i(0) &= A_i, \\ \Gamma_i(1) &= \Theta, \end{aligned}$$

  for some $A_i \subset \Theta$ and $d_i \in [0, 1]$.
- Hence, m_i is of the form $m_i(A_i) = 1 - d_i, m_i(\Theta) = d_i$, which may be denoted $A_i^{d_i}$ for short.
- Let S_i be the random variable, with state space Ω_i, representing the interpretation for the i-th message.
- Assume the messages have some dependence structure, described by a joint probability distribution $P_{1\ldots N}$ for variables $S_1, \ldots, S_N$, defined on $\boldsymbol{\Omega} := \times_{i=1}^{N} \Omega_i$ and having $P_1, \ldots, P_N$, as marginals.
- Assume these messages are reliable.

Resulting mass function

- Under the preceding conditions, knowledge about X is represented by mass function $m_{1\dots N}$ induced by the random set

$$(\boldsymbol{\Omega}, P_\cap, \Gamma_\cap)$$

  where
  - $\Gamma_\cap(\boldsymbol{\omega}) := \bigcap_{i=1}^{N} \Gamma_i(\omega_i)$ for all $\boldsymbol{\omega} = (\omega_1, \dots, \omega_N) \in \boldsymbol{\Omega}$
  - $P_\cap$ is $P_{1\dots N}$ conditioned on the event $\Theta_\cap = \{\boldsymbol{\omega} \in \boldsymbol{\Omega} : \Gamma_\cap(\boldsymbol{\omega}) \neq \emptyset\}$.
- If $P_{1\dots N} = \times_{i=1}^{N} P_i$ (independent messages), then

$$m_{1\dots N} = \bigoplus_{i=1}^{N} A_i^{d_i}$$

  and is thus separable[1].

[1] A mass function is separable if it can be obtained as the combination by Dempster's rule of simple mass functions.

Characterization of the dependence structure

- $P_{1\ldots N}$ is a multivariate Bernoulli distribution
- It is characterized by [Teugels, 1990]:

$$d_i = \mathbb{E}[S_i]$$

$$\sigma_{\boldsymbol{\omega}} := \mathbb{E}\left[\prod_{i=1}^{K}(S_i - d_i)^{\omega_i}\right]$$

for all $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_N) \in \boldsymbol{\Omega}$ such that $\sum_{i=1}^{N} \omega_i > 1$

- There are $2^N - N - 1$ central moments $\sigma_{\boldsymbol{\omega}}$. They represent the dependencies between any subset (of at least two) of all the S_i.
- Notation: $\boldsymbol{\sigma}$ vector whose elements are the dependencies $\sigma_{\boldsymbol{\omega}}$

→ Any dependence structure between some messages $(\Omega_i = \{0, 1\}, P_i, \Gamma_i)$, $i = 1, \ldots, N$, is fully described by a vector $\boldsymbol{\sigma}$ of central moments

σ-sum

Definition

- Let $(\Omega_i, P_i, \Gamma_i)$ be messages representing reliable and elementary pieces of evidence about X, inducing simple mass functions $m_i = A_i^{d_i}$, $i = 1, \ldots, N$.
- Assume they have a dependence structure described by some $\boldsymbol{\sigma}$.
- The mass function $m_{1\ldots N}$ is then induced by the random set $(\boldsymbol{\Omega}, P_\cap, \Gamma_\cap)$ with $P_\cap$ the probability distribution $P_{1\ldots N}$ specified by $\boldsymbol{\sigma}$ and conditioned on $\Theta_\cap$.

Definition

Let $A_1^{d_1}, \ldots, A_N^{d_N}$ be simple mass functions. Their $\boldsymbol{\sigma}$-sum is the mass function denoted $\oplus_{\boldsymbol{\sigma}}(A_1^{d_1}, \ldots, A_N^{d_N})$ and defined as

$$\oplus_{\boldsymbol{\sigma}}(A_1^{d_1}, \ldots, A_N^{d_N}) := m_{1\ldots N}.$$

- Remark: for $\boldsymbol{\sigma} = \mathbf{0}$, we have $P_{1\ldots N} = \times_{i=1}^N P_i$ and thus $\oplus_{\boldsymbol{\sigma}}$ reduces to $\oplus$, i.e., the **0**-sum is the orthogonal sum

σ-sum

Example

- The pieces of evidence in [Shafer, 1986] are represented by simple mass functions $A^{0.2}$ and $\overline{A}^{0.01}$.
- Their dependence is characterized by covariance

$$\sigma_{(1,1)} = \mathbb{E}\left[(S_1 - d_1)(S_2 - d_2)\right] = 0.007.$$

- Knowledge about X is represented by mass function

$$\oplus_{(0.007)}(A^{0.2}, \overline{A}^{0.01})$$

- We have

$$\begin{aligned} \left(\oplus_{(0.007)}(A^{0.2}, \overline{A}^{0.01})\right)(A) &\approx 0.005, \\ \left(\oplus_{(0.007)}(A^{0.2}, \overline{A}^{0.01})\right)(\overline{A}) &\approx 0.95, \\ \left(\oplus_{(0.007)}(A^{0.2}, \overline{A}^{0.01})\right)(\Theta) &\approx 0.045. \end{aligned}$$

Outline

Canonical decomposition

Theorem

Any mass function m on $\Theta = \{\theta_1, \ldots, \theta_K\}$ *satisfies*

$$m = \oplus_{\boldsymbol{\sigma}}(\overline{\{\theta_1\}}^{d_1}, \ldots, \overline{\{\theta_K\}}^{d_K})$$

with d_i, $1 \leq i \leq K$, *the means and* $\boldsymbol{\sigma}$ *the dependence vector of the* K*-variate Bernoulli distribution* $P_{1\ldots K}$ *such that*

$$P_{1\ldots K}(S_1 = \omega_1, \ldots, S_K = \omega_K) := m(A_\omega)$$

with A_ω *the subset of* Θ *such that* $\theta_i \in A_\omega$ *if* $\omega_i = 1$ *and* $\theta_i \notin A_\omega$ *if* $\omega_i = 0$, *for all* $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_K) \in \boldsymbol{\Omega}$.

- Remark: $d_i = pl(\theta_i)$, $1 \leq i \leq K$, with *pl* the contour function associated to *m* such that $pl(\theta_i) = \sum_{\theta_i \in A} m(A)$ for all $\theta_i \in \Theta$.

Example

- Mass function m defined on $\Theta = \{\theta_1, \theta_2, \theta_3\}$ by

$$\begin{aligned} m(\{\theta_1, \theta_2\}) &= 0.5, \\ m(\{\theta_3\}) &= 0.2, \\ m(\{\theta_2, \theta_3\}) &= 0.3 \end{aligned}$$

- Contour function: $pl(\theta_1) = 0.5, pl(\theta_2) = 0.8, pl(\theta_3) = 0.5$.
- m satisfies

$$m = \oplus_{(0.1,-0.25,-0.1,0)} \left(\overline{\{\theta_1\}}^{0.5}, \overline{\{\theta_2\}}^{0.8}, \overline{\{\theta_3\}}^{0.5} \right)$$

Comparison with previous solutions

- Alternative solution to that of [Shafer, 1976] and [Smets, 1995] for recovering belief functions from (reliable) elementary pieces of evidence.
- Since it is not possible to recover all belief functions merely from independent pieces of evidence (which leads only to the class of separable belief functions), our approach is to consider that they may be dependent.
- A simple yet arguably natural approach.
- More conventional than Smets', involving "debt of belief" represented by generalized simple mass functions, whose masses may lie outside the unit interval.
- Quite different from that of Shafer's, involving coarsening and limits, criticized by Smets.

Outline

Special case of only two elementary pieces of evidence

- 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 specified by providing some $\sigma_{(1,1)}$.
- Alternatively, P_{12} may be specified simply by providing

$$P_{12}(S_1 = 1, S_2 = 1).$$

- Choosing $P_{12}(S_1 = 1, S_2 = 1)$, given $d_1 = P_1(S_1 = 1)$ and $d_2 = P_2(S_2 = 1)$, actually amounts to specifying the dependence between events $S_1 = 1$ and $S_2 = 1$.

Correlation-based specification of the dependence

- This dependence can be completely characterized by a scalar $r \in [-1, 1]$ representing the correlation between the events.
- A model of correlation between two events of probabilities p_1 and p_2, with correlation r, is provided in [Ferson et al., 2004]: the probability of their conjunction is equal to $F(p_1, p_2, r)$ with

$$F(p_1, p_2, r) = \begin{cases} \min(p_1, p_2) & \text{if } r = 1, \\ p_1 \cdot p_2 & \text{if } r = 0, \\ \max(0, p_1 + p_2 - 1) & \text{if } r = -1, \\ \log_s[1 + (s^{p_1} - 1)(s^{p_2} - 1)/(s - 1)] & \text{otherwise,} \end{cases}$$

where $s = \tan(\pi(1 - r)/4)$.

- $r = 0$ corresponds to independence.

$\rightarrow$ The dependence between two elementary pieces of evidence can be characterized by a correlation $r \in [-1, 1]$.

r-sum

- Let $P_{\cap}^{r}$ be the result of conditioning P_{12}, specified by $P_{12}(S_1 = 1, S_2 = 1) := F(d_1, d_2, r)$ for some $r \in [-1, 1]$, on the event $\Theta_{\cap}$.
- Then, knowledge about X given messages $(\Omega_1, P_1, \Gamma_1)$ and $(\Omega_2, P_2, \Gamma_2)$ inducing $m_1 = A_1^{d_1}$ and $m_2 = A_2^{d_2}$, assumed to be reliable and with dependence characterized by r, is represented by the mass function induced by the random set $(\Omega_1 \times \Omega_2, P_{\cap}^{r}, \Gamma_{\cap})$.
- This mass function is called the r-sum of $A_1^{d_1}$ and $A_2^{d_2}$ and denoted $A_1^{d_1} \oplus_r A_2^{d_2}$.
- Binary operation $\oplus_r$ is a generalization of Dempster's rule for the combination of two simple mass functions ($\oplus$ recovered for $r = 0$).

Dependent positive and negative evidence

Definition

Positive and negative pieces of evidence with respect to a proposition $X \in A$ are elementary pieces of evidence inducing (non dogmatic) simple mass functions with focal set A and focal set $\overline{A}$, respectively.

- Our running example from [Shafer, 1986] is a case of (dependent) positive and negative evidence with respect to a proposition.

Proposition

Let A^{d_1} and $\overline{A}^{d_2}$ such that $d_i \in (0,1]$, $i = 1, 2$. We have

$$\begin{aligned}
(A^{d_1} \oplus_r \overline{A}^{d_2})(A) &= (d_2 - F(d_1, d_2, r))/(d_1 + d_2 - F(d_1, d_2, r)), \\
(A^{d_1} \oplus_r \overline{A}^{d_2})(\overline{A}) &= (d_1 - F(d_1, d_2, r))/(d_1 + d_2 - F(d_1, d_2, r)), \\
(A^{d_1} \oplus_r \overline{A}^{d_2})(\Theta) &= F(d_1, d_2, r)/(d_1 + d_2 - F(d_1, d_2, r)), \\
(A^{d_1} \oplus_r \overline{A}^{d_2})(B) &= 0, \quad \forall B \in 2^{\Theta} \setminus \{A, \overline{A}, \Theta\}.
\end{aligned}$$

Evidential RBF network

- [Huang et al., 2022] introduced an alternative evidential classifier to the prototype-based improvement [Denoeux, 2000] of the EkNN, having similar properties.
- Obtained by applying ideas developed in [Denoeux, 2019], to a radial basis function network (RBFN) with a softmax output layer (or a logistic output unit in the case of binary classification).

$\rightarrow$ Evidential RBFN (ERBFN).

- Used to enhance the predictions of a UNet model for a task of lymphoma segmentation from 3D PET-CT images.

Principle

- The ERBFN reveals a predictive, so-called latent, mass function $m_{\mathbf{u}}$ underlying the probabilistic prediction $P_{\mathbf{u}}$ of a given (trained) RBFN with a softmax output layer, with respect to the unknown class $X \in \Theta$ of an instance with feature vector **u**.
- $m_{\mathbf{u}}$ underlies $P_{\mathbf{u}}$ in the sense that its approximation by a Bayesian mass function using the plausibility transformation method [Cobb and Shenoy, 2006] is exactly $P_{\mathbf{u}}$.

Definition (Plausibility transformation method)

Let m be a mass function with contour function pl. Its approximation is the Bayesian mass function p_m defined as

$$p_m(\{\theta_k\}) := \frac{pl(\theta_k)}{\sum_{\ell=1}^{K} pl(\theta_\ell)}, \quad k = 1, \ldots, K.$$

Latent mass function $m_{\mathbf{u}}$

In a nutshell

- $m_{\mathbf{u}}$ is obtained by:
  1. defining positive and negative pieces of evidence, denoted m_k^+ and m_k^-, for each class θ_k, based on the parameters of the RBFN and on $\mathbf{u}$;
  2. pooling them by Dempster's rule.
- We have thus

$$m_{\mathbf{u}} := \bigoplus_{k=1}^{K} \left(m_k^+ \oplus m_k^- \right).$$

(In)Dependence between positive and negative evidence

- In the ERBFN, positive and negative evidence for a given class are considered independent.
- However, they are obtained from the same set of values and therefore the independence assumption may be questioned.

$\rightarrow$ When pooling the positive and negative evidence for θ_k, it seems safer to assume that there is some dependence between them.

- Such a dependence can be characterized by a correlation $r_k \in [-1, 1]$.

Existence of a set of latent mass functions for a RBFN

Theorem

Let $m_{\mathbf{u},\mathbf{r}}$, for some $\mathbf{r} = (r_1 \ldots, r_K) \in [-1, 1]^K$, be the mass function defined as

$$m_{\mathbf{u},\mathbf{r}} := \bigoplus_{k=1}^{K} \left(m_k^+ \oplus_{r_k} m_k^- \right).$$

We have

$$p_{m_{\mathbf{u},\mathbf{r}}}(\{\theta_k\}) = P_{\mathbf{u}}(\theta_k), \quad \forall \theta_k \in \Theta,$$

with $p_{m_{\mathbf{u},\mathbf{r}}}$ the approximation of $m_{\mathbf{u},\mathbf{r}}$

→ The independence assumption made in the ERBFN for each class is actually inconsequential insofar as any possible dependence structure yields a predictive latent mass function.

Identification of the correlations

- Which **r** to select to compute the predictive latent mass function for any given test instance?
- Assume some learning data $\{\mathbf{u}_i, x_i\}_{i=1}^n$, where $\mathbf{u}_i$ is the feature vector of instance i and x_i is its true class, are available.
- We may fit **r** over this learning set, i.e., we can search for the correlations $\hat{\mathbf{r}}$ that minimizes the loss over this learning data:

$$\hat{\mathbf{r}} = \arg \min_{\mathbf{r} \in [-1,1]^K} \sum_{i=1}^{n} \mathcal{L}(x_i, m_{\mathbf{u}_i,\mathbf{r}}).$$

  for some loss $\mathcal{L}(x, m_\mathbf{u})$ of an evidential prediction $m_\mathbf{u}$ for an instance with feature vector **u** and whose true class is x.
- Following [Denœux, 2024], an appropriate choice is the generalized negative log-likelihood (GNLL) criterion:

$$\mathcal{L}(x, m_\mathbf{u}) = -\frac{1}{2} \ln Bel_\mathbf{u}(\{x\}) - \frac{1}{2} \ln Pl_\mathbf{u}(\{x\}).$$

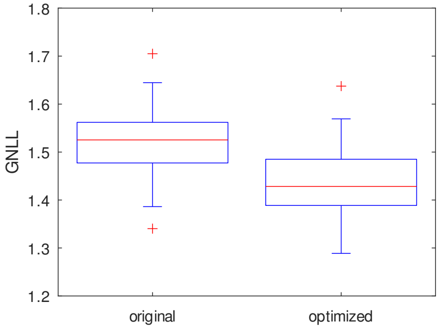
Experiments

Protocol

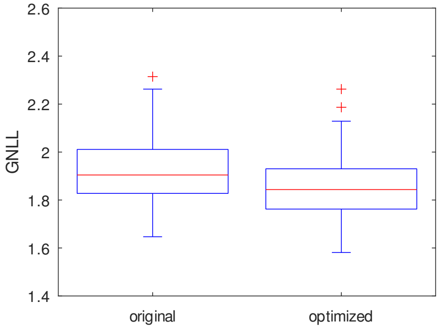
- Datasets: Pima (2 classes), Ionosphere (2 classes), Glass (6 classes), Vowel (6 classes).
- For each dataset, the data were split randomly into training, validation and test sets containing, respectively, 60%, 20% and 20% of the instances.
- The training set was used to learn the RBFN, the validation set was used to optimize $\mathbf{r}$, and the test set was used to evaluate the performance, according to the average GNLL, of $\mathbf{r} = \hat{\mathbf{r}}$ as well as of $\mathbf{r} = \mathbf{0}$ (original proposal from [Huang et al., 2022]).
- This process was repeated 50 times.

Experiments

Results, binary classification



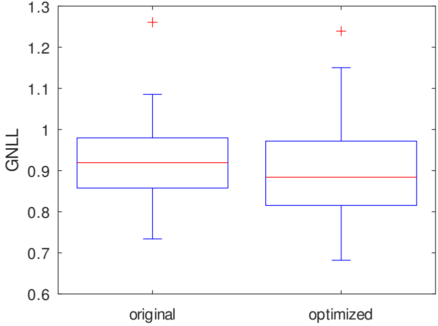
Pima



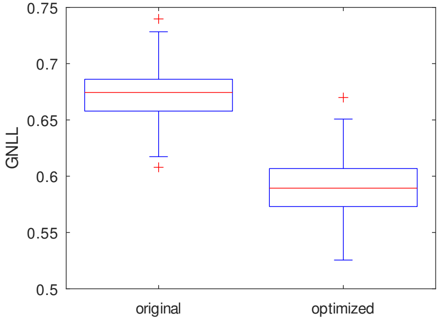
Ionosphere

Experiments

Results, multi-class classification



Glass



Vowel

Summary

- Reliability and independence $\rightarrow$ Dempster's rule.
- A general approach for the fusion of partially reliable pieces of evidence, allowing us:
  - To account for various forms of the lack of reliability;
  - To obtain an interpretation of Dempster's rule generalized to all logical connectives.
- An approach for the fusion of dependent elementary pieces of evidence, allowing us:
  - To account for all possible dependence structures through some dependence quantities;
  - To obtain an interpretation of belief functions.
- Usefulness of these extensions in applications.
- Partially reliable and dependent pieces of evidence about continuous variables in [Denœux, 2024].

References[2] on reliability and dependence

Reliability

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.

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.

Dependence

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

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, pages 354-368, vol. 15350 of LNAI, 2024.

[2] The first two theorems presented in this talk can be seen as "normalized" versions of results in these references. Their full proofs are provided in the appendix to my BELIEF 2024 talk available at https://www.lgi2a.univ-artois.fr/~pichon/.

References I

- B.R. Cobb and P. P. Shenoy.
  On the plausibility transformation method for translating belief function models to probability models.
  Int. J. of Approximate Reasoning, 41(3):314–330, 2006.

- 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.

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

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

- S. Ferson, R. Nelsen,J. Hajagos, D. Berleant, J. Zhang, W.T. Tucker, L. Ginzburg, and W.L. Oberkampf.
  Dependence in probabilistic modeling, Dempster-Shafer theory, and probability bounds analysis.
  Sandia Nat. Lab., Tech. Rep. SAND2004-3072, 2004.

- L. Huang, S. Ruan, P. Decazes, and T. Denœux.
  Lymphoma segmentation from 3D PET-CT images using a deep evidential network.
  Int. J. of Approximate Reasoning, 149:39–60, 2022.

References II

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, 102648, 2025.

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.

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.
A mathematical theory of evidence.
Princeton University Press, 1976.

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

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

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

References III

J.L. Teugels.
Some representations of the multivariate Bernoulli and binomial distributions.
Journal of Multivariate Analysis, 32(2):256–268, 1990.

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

Thank you for your attention.