Introduction to belief functions

Thierry Denœux

Université de technologie de Compiègne, France
Institut Universitaire de France
https://www.hds.utc.fr/~tdenoeux

6th School on Belief Functions and their Applications
Ishikawa, Japan, October 28, 2023

Contents of this lecture

1. Fundamental concepts: belief, plausibility, commonality, conditioning, basic combination rules.
2. Some more advanced concepts: informational ordering, cautious rule, compatible frames.

Theory of belief functions

History

- A formal framework for representing and reasoning with uncertain information.
- Also known as Dempster-Shafer (DS) theory or Evidence theory.
- Originates from the work of Dempster (1967)[1] in the context of statistical inference.
- Formalized by Shafer (1976)[2] as a theory of evidence.
- Popularized and developed by Smets in the 1980's and 1990's as the "Transferable Belief Model".
- Starting from the 1990's, growing number of applications in information fusion, knowledge representation, machine learning (classification, clustering), reliability and risk analysis, etc.

[1] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping, *Annals of Mathematical Statistics*, 38:325–339, 1967.

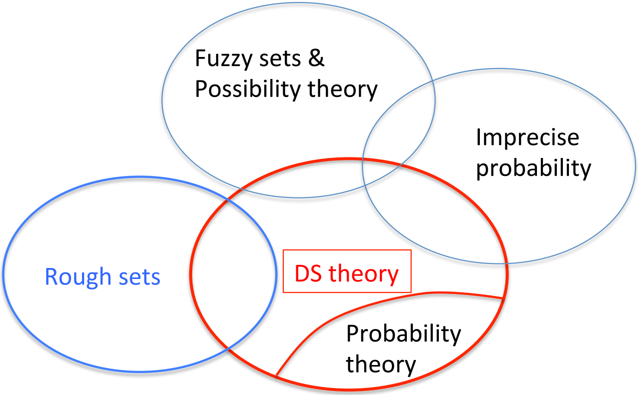
[2] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, Princeton, N.J., 1976.

Theory of belief functions

Main idea

- The theory of belief functions extends both logical/set-based formalisms (such as Propositional Logic and Interval Analysis) and Probability Theory:
  - A belief function may be viewed both as a generalized set and as a nonadditive measure
  - The theory includes extensions of probabilistic notions (conditioning, marginalization) and set-theoretic notions (intersection, union, inclusion, etc.).
- DS reasoning produces the same results as probabilistic reasoning or interval analysis when provided with the same information.
- However, the greater expressive power of the theory of belief functions allows us to represent what we know in a more faithful way.

Relationships with other theories



Outline

1. Basic notions
   - Mass functions
   - Belief and plausibility functions
   - Dempster's rule

2. Selected advanced topics
   - Informational orderings
   - Cautious rule
   - Compatible frames

Outline

Mass function

Definition

Definition (Frame of discernment, mass function, focal set)

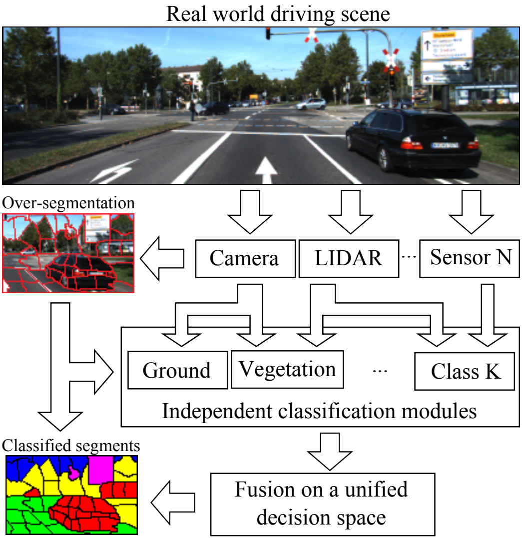
Let Ω be the finite set called a frame of discernment. A mass function on Ω is a mapping $m : 2^\Omega \to [0,1]$ such that

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

Every subset A of Ω such that $m(A) > 0$ is a focal set of m. If $m(\emptyset) = 0$, m is said to be normalized (assumed in this lecture).

In DS theory, a mass function is used to represent evidence about an uncertain variable X taking values in Ω.

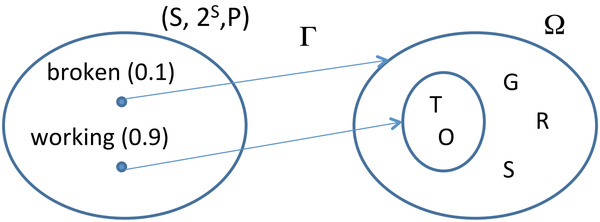
Example: road scene analysis



Example: road scene analysis (continued)

- Let X be the type of object in some region of the image, and $\Omega = \{G, R, T, O, S\}$, corresponding to the possibilities **G**rass, **R**oad, **T**ree/Bush, **O**bstacle, **S**ky.
- Assume that a lidar sensor (laser telemeter) returns the information $X \in \{T, O\}$, but we there is a probability $p = 0.1$ that the information is not reliable (because, e.g., the sensor is out of order).
- How to represent this information by a mass function?

Formalization



- Here, the probability p is not about X, but about the state of a sensor.
- Let $S = \{\text{working}, \text{broken}\}$ the set of possible sensor states.
  - If the state is "working", we know that $X \in \{T, O\}$.
  - If the state is "broken", we just know that $X \in \Omega$, and nothing more.
- This uncertain evidence can be represented by the following mass function m on Ω:

$$m(\{T, O\}) = 0.9, \quad m(\Omega) = 0.1$$

Meaning of a mass function

- In the previous example,
  - $m(\{T, O\}) = 0.9$ is the probability of knowing only that $X \in \{T, O\}$, and nothing more
  - $m(\Omega) = 0.1$ is the probability of knowing nothing at all.
- In general, what is the meaning (semantics) of a mass function in DS theory?
- A precise interpretation was proposed by Shafer (1981)[3]: random code semantics.

[3] G. Shafer. Constructive probability. *Synthese*, 48(1):1–60, 1981.

Random code semantics

- We consider a situation in which we receive a coded message containing reliable information about variable $X \in \Omega$.
- The message was encoded using some code in the set $S = \{c_1, \ldots, c_n\}$.
- There is a multi-valued mapping $\Gamma : S \rightarrow 2^\Omega \setminus \{\emptyset\}$ that defines the meaning of the message: if code c_i was used, then the meaning of the message is "$X \in \Gamma(c_i)$".
- We don't know which code was used, but we know that each code c_i had a chance p_i of being selected, with $\sum_{i=1}^n p_i = 1$.
- Then $m(A)$ is the probability that the meaning of the message is "$X \in A$":

$$m(A) = P(\{c \in S : \Gamma(c) = A\}) = \sum_{i:\Gamma(c_i)=A} p_i$$

Random code semantics (continued)

- In practice, we do not receive randomly coded messages, but we can construct a mass function by comparing our evidence about some variable X, to a hypothetical situation in which we receive a randomly coded message.
- A mass function m can be elicited by finding the "coded-message" canonical example that is the most similar to our evidence.
- Remark: The tuple $(S, 2^S, P, \Omega, 2^\Omega, \Gamma)$ is called a random set. This notion plays an important role for defining belief functions in infinite spaces. I will also introduce the more general notion of random fuzzy set in a later lecture.

Special mass functions

Definition (Logical mass function)

If a mass function has only one focal set $A \subseteq \Omega$., it is said to be logical; we denote it as $m_{[A]}$. It represents "infallible" evidence that tells us that $X \in A$ for sure and nothing more. (There is a one-to-one correspondence between logical mass functions and nonempty sets).

Definition (Vacuous mass function)

The vacuous mass function $m_?$ is the logical mass function such that $m_?(\Omega) = 1$. It represents total ignorance.

Definition (Bayesian mass function)

A mass function is Bayesian if its focal sets are singletons. It is equivalent to a probability distribution.

Outline

Definitions

Definition (Belief, plausibility, contour functions)

Given a mass function m on Ω, the corresponding belief and plausibility functions are mappings from 2^Ω to [0,1] defined as follows:

$$Bel(A) = \sum_{B \subseteq A} m(B)$$

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B) = 1 - Bel(\overline{A}).$$

The mapping $pl :\rightarrow \Omega$ such that $pl(\omega) = Pl(\{\omega\})$ is called the contour function associated to m.

Interpretation:

- $Bel(A)$ is a measure of the total support given to A
- $Pl(A)$ is a measure of the lack of support given to $\overline{A}$

Road scene analysis example

- We had $\Omega = \{G, R, T, O, S\}$ and

$$m(\{T, O\}) = 0.9, \quad m(\Omega) = 0.1$$

- Degrees of belief and plausibility of some subsets of Ω:

A	$\emptyset$	$\{T\}$	$\{O\}$	$\{T, O\}$	$\{T, O, R\}$	$\{T, R\}$	$\{R, S\}$	Ω
$Bel(A)$	0	0	0	0.9	0.9	0	0	1
$Pl(A)$	0	1	1	1	1	1	0.1	1

Elementary properties

- $Bel(\emptyset) = Pl(\emptyset) = 0$
- $Bel(\Omega) = Pl(\Omega) = 1$
- Superadditivity of *Bel*:

$$Bel(A \cup B) \geq Bel(A) + Bel(B) - Bel(A \cap B)$$

- Subadditivity of *Pl*:

$$Pl(A \cup B) \leq Pl(A) + Pl(B) - Pl(A \cap B)$$

- When m is Bayesian, the two mappings *Bel* and *Pl* are equal and additive:

$$Bel(A) = Pl(A) = \sum_{\omega \in A} m(\{\omega\})$$

for all $A \subseteq \Omega$.

Characterization of belief functions

- Function $Bel : 2^\Omega \to [0,1]$ is completely monotone: for any $k \geq 2$ and for any family $A_1, \ldots, A_k$ in 2^Ω:

$$Bel\left(\bigcup_{i=1}^{k} A_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} Bel\left(\bigcap_{i \in I} A_i\right)$$

- Conversely, to any completely monotone set function Bel such $Bel(\emptyset) = 0$ and $Bel(\Omega) = 1$ corresponds a unique mass function m such that:

$$m(A) = \sum_{\emptyset \neq B \subseteq A} (-1)^{|A|-|B|} Bel(B), \quad \forall A \subseteq \Omega.$$

Relations between *m*, *Bel* and *Pl*

- Let m be a mass function, Bel and Pl the corresponding belief and plausibility functions.
- For all $A \subseteq \Omega$,

$$Bel(A) = 1 - Pl(\overline{A})$$

$$m(A) = \sum_{\emptyset \neq B \subseteq A} (-1)^{|A|-|B|} Bel(B)$$

$$m(A) = \sum_{B \subseteq A} (-1)^{|A|-|B|+1} Pl(\overline{B})$$

- m, Bel and Pl are thus three equivalent representations of a piece of evidence.

Relationship with Possibility Theory

- When the focal sets of m are nested: $A_1 \subset A_2 \subset \ldots \subset A_r$, m is said to be consonant.
- The following relations then hold:

$$Pl(A \cup B) = \max\left(Pl(A), Pl(B)\right), \quad \forall A, B \subseteq \Omega$$

  and the plausibility function can be computed from the contour function as

$$Pl(A) = \max_{\omega \in A} pl(\omega), \quad \forall A \subseteq \Omega$$

- Pl is then called a possibility measure, and Bel is the dual necessity measure.
- In a sense, the theory of belief functions can thus be considered as more expressive than possibility theory (but the combination operations are different, as we will see later).

Relation with imprecise probabilities

- A probability measure P on Ω is said to be compatible with m if

$$\forall A \subseteq \Omega, \quad Bel(A) \leq P(A) \leq Pl(A)$$

- The set $\mathcal{P}(m)$ of probability measures compatible with m is called the credal set of m

$$\mathcal{P}(m) = \{P : \forall A \subseteq \Omega, Bel(A) \leq P(A)\}$$

- *Bel* is the lower envelope of $\mathcal{P}(m)$

$$\forall A \subseteq \Omega, \quad Bel(A) = \min_{P \in \mathcal{P}(m)} P(A)$$

- Not all lower envelopes of sets of probability measures are belief functions.
- The theory of belief functions is not a theory of imprecise probabilities (the two theories have different conditioning operations).

Outline

Road scene example continued

- Variable X was defined as the type of object in some region of the image, and the frame was $\Omega = \{G, R, T, O, S\}$, corresponding to the possibilities **G**rass, **R**oad, **T**ree/Bush, **O**bstacle, **S**ky
- A lidar sensor gave us the following mass function:

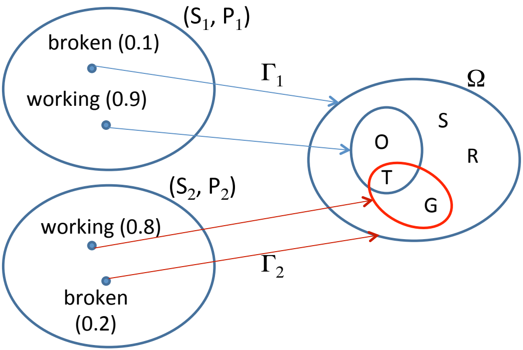
$$m_1(\{T, O\}) = 0.9, \quad m_1(\Omega) = 0.1$$

- Now, assume that a camera returns the mass function:

$$m_2(\{G, T\}) = 0.8, \quad m_2(\Omega) = 0.2$$

- How to combine these two pieces of evidence?

Analysis



- If the two sensors are in states s_1 and s_2, then $X \in \Gamma_1(s_1) \cap \Gamma_2(s_2)$.
- If the two pieces of evidence are independent, then the probability that the sensors are in states s_1 and s_2 is $P_1(\{s_1\})P_2(\{s_2\})$.

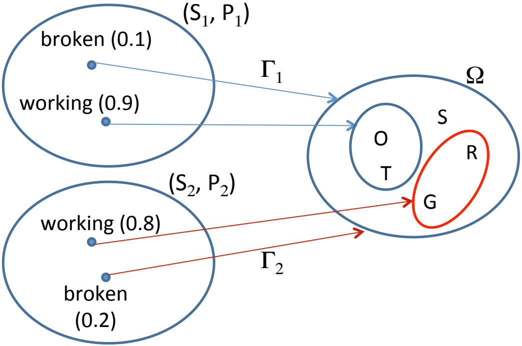
Computation

$m_1 \backslash m_2$	$\{T, G\}$ (0.8)	Ω (0.2)
$\{O, T\}$ (0.9)	$\{T\}$ (0.72)	$\{O, T\}$ (0.18)
Ω (0.1)	$\{T, G\}$ (0.08)	Ω (0.02)

We then get the following combined mass function:

$$
\begin{aligned}
m(\{T\}) &= 0.72 \\
m(\{O, T\}) &= 0.18 \\
m(\{T, G\}) &= 0.08 \\
m(\Omega) &= 0.02
\end{aligned}
$$

Case of conflicting pieces of evidence



- If $\Gamma_1(s_1) \cap \Gamma_2(s_2) = \emptyset$, we know that the pair of states (s_1, s_2) cannot have occurred.
- The joint probability distribution on $S_1 \times S_2$ must be conditioned to eliminate such pairs.

Computation

$m_1 \backslash m_2$	$\{G, R\}$ (0.8)	Ω (0.2)
$\{O, T\}$ (0.9)	$\emptyset$ (0.72)	$\{O, T\}$ (0.18)
Ω (0.1)	$\{G, R\}$ (0.08)	Ω (0.02)

We then get the following combined mass function,

$$
\begin{aligned}
m(\emptyset) &= 0 \\
m(\{O, T\}) &= 0.18/0.28 = 9/14 \\
m(\{G, R\}) &= 0.08/0.28 = 4/14 \\
m(\Omega) &= 0.02/0.28 = 1/14
\end{aligned}
$$

Dempster's rule

Definition (Degree of conflict)

Let m_1 and m_2 be two mass functions. Their degree of conflict is

$$\kappa = \sum_{B \cap C = \emptyset} m_1(B) m_2(C)$$

Definition (Orthogonal sum)

Let m_1 and m_2 be two mass functions such that $\kappa < 1$. Their orthogonal sum is the mass function defined by

$$(m_1 \oplus m_2)(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - \kappa}$$

for all $A \neq \emptyset$ and $(m_1 \oplus m_2)(\emptyset) := 0$.

Properties

Proposition

1. *If several pieces of evidence are combined, the order does not matter:*
$$m_1 \oplus m_2 = m_2 \oplus m_1$$
$$m_1 \oplus (m_2 \oplus m_3) = (m_1 \oplus m_2) \oplus m_3$$
2. *A mass function m is not changed if combined with the vacuous mass function $m_?$:*
$$m \oplus m_? = m$$
3. *Let pl_1, pl_2 and pl_{12} be the contour functions associated with, respectively, m_1, m_2 and $m_1 \oplus m_2$. We have*
$$pl_{12} = \frac{1}{1-\kappa} pl_1 pl_2$$

Misconception about Dempster's rule

- Following a 1979 report by Zadeh, it is repeated that "Dempster's rule yields counterintuitive results" (which is usually used as a justification to introduce new combination rules)
- Zadeh's example: $\Omega = \{a, b, c\}$, two experts

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

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

  We get $(m_1 \oplus m_2)(\{b\}) = 1$, which is claimed to be "counterintuitive" because both experts considered b as very unlikely.
- But Expert 1 claims that c is absolutely impossible, and Expert 2 claims that a is absolutely impossible, so b is the only remaining possibility!
- Dempster's rule does produce sound results when used and interpreted correctly.

Dempster's conditioning

- Conditioning is a special case of Dempster's rule, where a mass function m is combined with a logical mass function $m_{[A]}$. Notation:

$$m \oplus m_{[A]} = m(\cdot \mid A)$$

- It can be shown that

$$Pl(B \mid A) = \frac{Pl(A \cap B)}{Pl(A)}.$$

- Generalization of Bayes' conditioning: if m is a Bayesian mass function and $m_{[A]}$ is a logical mass function, then $m \oplus m_{[A]}$ is a Bayesian mass function corresponding to the conditioning of m by A.

Commonality function

- Commonality function: let $Q : 2^\Omega \to [0,1]$ be defined as

$$Q(A) = \sum_{B \supseteq A} m(B), \quad \forall A \subseteq \Omega$$

- Conversely,

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

- Q is another equivalent representation of a belief function.

Commonality function and Dempster's rule

- Let Q_1 and Q_2 be the commonality functions associated to m_1 and m_2.
- Let Q_{12} be the commonality function associated to $m_1 \oplus m_2$.
- We have

$$Q_{12}(A) = \frac{1}{1-\kappa} Q_1(A) Q_2(A), \quad \forall A \subseteq \Omega, A \neq \emptyset$$

$$(Q_1 \oplus Q_2)(\emptyset) = 1$$

Smets' disjunctive rule

- Let m_1 and m_2 be two mass functions induced by random messages/sets $(S_1, 2^{S_1}, P_1, \Omega, 2^\Omega, \Gamma_1)$ and $(S_2, 2^{S_2}, P_2, \Omega, 2^\Omega, \Gamma_2)$.
- Previously, we have assumed that both messages were reliable, i.e., if the true codes are $c_1 \in S_1$ and $c_2 \in S_2$, we can conclude that $X \in \Gamma_1(c_1) \cap \Gamma_2(c_2)$ for sure.
- We can weaken this assumption by supposing only that at least one of the two messages is reliable, i.e., if the true codes are $c_1 \in S_1$ and $c_2 \in S_2$, we can only conclude that $X \in \Gamma_1(c_1) \cup \Gamma_2(c_2)$ for sure.
- This leads to the Smets' disjunctive rule:

$$(m_1 \textcircled{\cup} m_2)(A) = \sum_{B \cup C = A} m_1(B) m_2(C), \quad \forall A \subseteq \Omega$$

- $Bel_1 \textcircled{\cup} Bel_2 = Bel_1 \cdot Bel_2$

Outline

Informational comparison of belief functions

- Let m_1 and m_2 be two mass functions on Ω
- In what sense can we say that m_1 is more informative (committed) than m_2?
- Special case:
  - Let $m_{[A]}$ and $m_{[B]}$ be two logical mass functions
  - $m_{[A]}$ is more committed than $m_{[B]}$ iff $A \subseteq B$
- Extension to arbitrary mass functions?

Plausibility ordering

Definition

m_1 is pl-more committed than m_2 (noted $m_1 \sqsubseteq_{pl} m_2$) if

$$Pl_1(A) \leq Pl_2(A), \quad \forall A \subseteq \Omega$$

or, equivalently,

$$Bel_1(A) \geq Bel_2(A), \quad \forall A \subseteq \Omega.$$

- Imprecise probability interpretation:

$$m_1 \sqsubseteq_{pl} m_2 \Leftrightarrow \mathcal{P}(m_1) \subseteq \mathcal{P}(m_2)$$

- Properties:
  - Extension of set inclusion:

$$m_{[A]} \sqsubseteq_{pl} m_{[B]} \Leftrightarrow A \subseteq B$$

  - Greatest element: vacuous mass function $m_?$

Commonality ordering

- If $m_1 = m \oplus m_2$ for some m, and if there is no conflict between m and m_2, then $Q_1(A) = Q(A)Q_2(A) \leq Q_2(A)$ for all $A \subseteq \Omega$
- This property suggests that smaller values of the commonality function are associated with richer information content of the mass function

Definition

m_1 is q-more committed than m_2 (noted $m_1 \sqsubseteq_q m_2$) if

$$Q_1(A) \leq Q_2(A), \quad \forall A \subseteq \Omega$$

Properties:

- Extension of set inclusion:

$$m_{[A]} \sqsubseteq_q m_{[B]} \Leftrightarrow A \subseteq B$$

- Greatest element: vacuous mass function $m_?$

Strong (specialization) ordering

Definition

m_1 is a specialization of m_2 (noted $m_1 \sqsubseteq_s m_2$) if m_1 can be obtained from m_2 by distributing each mass $m_2(B)$ to subsets of B:

$$m_1(A) = \sum_{B \subseteq \Omega} S(A,B) m_2(B), \quad \forall A \subseteq \Omega,$$

where $S(A,B) =$ proportion of $m_2(B)$ transferred to $A \subseteq B$.

- S is called a specialization matrix
- Properties:
  - Extension of set inclusion
  - Greatest element: $m_?$
  - $m_1 \sqsubseteq_s m_2 \Rightarrow \begin{cases} m_1 \sqsubseteq_{pl} m_2 \\ m_1 \sqsubseteq_q m_2 \end{cases}$

Least Commitment Principle

Definition (Least Commitment Principle)

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

A very powerful method for constructing belief functions!

Outline

Motivation

- The basic rules $\oplus$ and $\textcircled{\cup}$ assume the sources of information to be independent, e.g.
  - experts with nonoverlapping experience/knowledge
  - nonoverlapping datasets
- What to do in case of dependent/overlapping evidence?
  - Describe the nature of the interaction between sources (difficult, requires a lot of information)
  - Use a combination rule that tolerates redundancy in the combined information
- Such rules can be derived from the LCP using suitable informational orderings.

Principle

- Two sources provide mass functions m_1 and m_2, and the sources are both considered to be reliable.
- After receiving these m_1 and m_2, the agent's state of belief should be represented by a mass function m_{12} more committed than m_1, and more committed than m_2.
- Let $\mathcal{S}_x(m)$ be the set of mass functions m' such that $m' \sqsubseteq_x m$, for some $x \in \{pl, q, s, \cdots\}$. We thus impose that

$$m_{12} \in \mathcal{S}_x(m_1) \cap \mathcal{S}_x(m_2).$$

- According to the LCP, we should select the x-least committed element in $\mathcal{S}_x(m_1) \cap \mathcal{S}_x(m_2)$, if it exists.

Need for a new ordering relation

- The above approach works for special cases.
- Example[4]: if m_1 and m_2 are consonant, then the q-least committed element in $\mathcal{S}_q(m_1) \cap \mathcal{S}_q(m_2)$ exists and it is unique: it is the consonant mass function with commonality function $Q_{12} = \min(Q_1, Q_2)$.
- In general, neither existence nor uniqueness of a solution can be guaranteed with any of the x-orderings, $x \in \{pl, q, s\}$.
- We need to define a new ordering relation.

[4] D. Dubois and H. Prade and Ph. Smets. New Semantics for Quantitative Possibility Theory. Proc. of ECSQARU 2001, pp 410–421, Springer Verlag, 2001.

Simple mass functions

- Definition: m is simple mass function if it has the following form

$$m(A) = 1 - \delta(A)$$
$$m(\Omega) = \delta(A)$$

  for some $A \subset \Omega$, $A \neq \emptyset$ and $\delta(A) \in (0, 1]$.
- The quantity $w(A) = -\ln \delta(A) \geq 0$ is called the weight of evidence for A. Mass function m is denoted by $A^{w(A)}$.
- Property:

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

Separable mass functions

Definition (Separable mass function)

A (normalized) mass function is separable if it can be written as the orthogonal sum of simple mass functions:

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

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

The w-ordering

Definition

Let m_1 and m_2 be two mass functions. We say that m_1 is w-more committed than m_2 (denoted by $m_1 \sqsubseteq_w m_2$) if

$$m_1 = m_2 \oplus m.$$

for some separable mass function m.

How to check this condition?

Weight function

- If m is separable, the corresponding weights of evidence can be recovered as
$$w(A) = \sum_{B \supseteq A} (-1)^{|B|-|A|} \ln Q(B) \quad (1)$$
for all $A \subseteq \Omega$.
- For any nondogmatic mass function m, (i.e., such that $m(\Omega) > 0$), we can still define "weights" from (1), but we can have $w(A) < 0$.
- Function w is called the weight function.
- m can be computed from w by
$$m = \bigoplus_{\emptyset \neq A \subset \Omega} A^{w(A)},$$
although $A^{w(A)}$ is not a proper mass function when $w(A) < 0$.

Properties of the weight function

- m is separable iff

$$w(A) \geq 0, \quad \forall A \subset \Omega, A \neq \emptyset$$

- Dempster's rule can be computed using the w-function by

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

- Equivalent definition of the w-ordering[5]

$$\boxed{m_1 \sqsubseteq_w m_2 \Leftrightarrow w_1(A) \geq w_2(A), \quad \forall A \subset \Omega, A \neq \emptyset}.$$

[5] T. Denoeux. Conjunctive and Disjunctive Combination of Belief Functions Induced by Non Distinct Bodies of Evidence. *Artificial Intelligence*, 172:234–264, 2008.

Cautious rule

Proposition

Let m_1 and m_2 be two nondogmatic mass functions with weight functions w_1 and w_2. The w-least committed element in $\mathcal{S}_w(m_1) \cap \mathcal{S}_w(m_2)$ exists and is unique. It is defined by:

$$m_1 \owedge m_2 = \bigoplus_{\emptyset \neq A \subset \Omega} A^{\max(w_1(A), w_2(A))}$$

Operator $\owedge$ is called the (normalized) cautious rule.

Computation

Cautious rule computation

m-space		w-space
m_1	$\longrightarrow$	w_1
m_2	$\longrightarrow$	w_2
$m_1 \textcircled{\wedge} m_2$	$\longleftarrow$	$\max(w_1, w_2)$

Remark: we often have simple mass functions in the first place, so that the *w* function is readily available.

Properties of the cautious rule

- Commutative, associative
- Idempotent :

$$\forall m, \quad m \textcircled{\wedge} m = m$$

- Distributivity of $\oplus$ with respect to $\textcircled{\wedge}$

$$\forall m_1, m_2, m_3, \quad (m_1 \oplus m_2) \textcircled{\wedge} (m_1 \oplus m_3) = m_1 \oplus (m_2 \textcircled{\wedge} m_3)$$

The common item of evidence m_1 is not counted twice!

- No neutral element, but $m_? \textcircled{\wedge} m = m$ iff m is separable.

Basic rules

The four basic rules

Sources	independent	dependent
All reliable	$\oplus$	$\textcircled{\wedge}$
At least one reliable	$\textcircled{\cup}$	$\textcircled{\vee}$

$\textcircled{\vee}$ is the bold disjunctive rule

Outline

Refinement and coarsening

Example

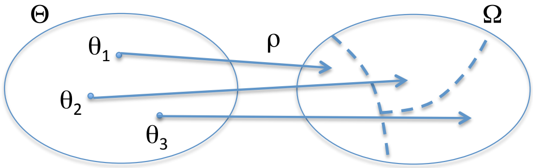
- Let us come back to the road scene analysis example, with $\Omega = \{G, R, T, O, S\}$.
- Assume that we have a vegetation detector, which can determine if a region of the image contains vegetation or not. For this detector, the frame of discernment is $\Theta = \{V, \neg V\}$, where V means that there is vegetation, and $\neg V$ means that there is no vegetation.
- We have the correspondence

$$\begin{aligned} V &\rightarrow \{G, T\} \\ \neg V &\rightarrow \{R, O, S\} \end{aligned}$$

- The elements of Ω can be obtained by splitting some or all of the elements of Θ. We say that Ω is a refinement of Θ, and Θ is a coarsening of Ω

Refinement and coarsening

General definition



Definition

A frame Ω is a refinement of a frame Θ iff there is a mapping $\rho : 2^\Theta \rightarrow 2^\Omega$ (called a refining) such that:

- $\{\rho(\{\theta\}), \theta \in \Theta\} \subseteq 2^\Omega$ is a partition of Ω, and
- For all $A \subseteq \Omega$, $\rho(A) = \bigcup_{\theta \in A} \rho(\{\theta\})$.

Vacuous extension

- In the road scene example, assume that the vegetation detector provides the following mass function on Θ:

$$m^{\Theta}(\{V\}) = 0.6, \quad m^{\Theta}(\{\neg V\}) = 0.3, \quad m^{\Theta}(\Theta) = 0.1$$

- How to express m^{Θ} in Ω?
- Solution: for all $A \subseteq \Theta$, we transfer the mass $m^{\Theta}(A)$ to $\rho(A)$. Here,

$$\begin{array}{rcl} m^{\Theta}(\{V\}) = 0.6 & \rightarrow & \rho(\{V\}) = \{G, T\} \\ m^{\Theta}(\{\neg V\}) = 0.3 & \rightarrow & \rho(\{\neg V\}) = \{R, O, S\} \\ m^{\Theta}(\Theta) = 0.1 & \rightarrow & \rho(\Theta) = \Omega \end{array}$$

- We finally get the following mass function on Ω,

$$m^{\Theta\uparrow\Omega}(\{G, T\}) = 0.6, \quad m^{\Theta\uparrow\Omega}(\{R, O, S\}) = 0.3, \quad m^{\Theta\uparrow\Omega}(\Omega) = 0.1.$$

- $m^{\Theta\uparrow\Omega}$ is called the vacuous extension of m^{Θ} in Ω.

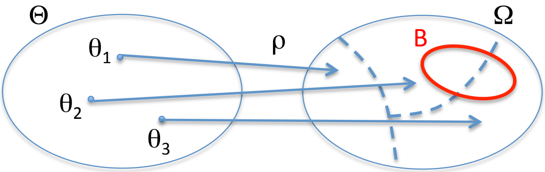
Expression of information in a coarser frame

- Let us now assume that we have the following mass function on Ω,

$$m^{\Omega}(\{T\}) = 0.4, \quad m^{\Omega}(\{T, O\}) = 0.3, \quad m^{\Omega}(\{R, S\}) = 0.3.$$

- How to express m^{Ω} in Θ?
- We cannot do it without loss of information, because, for instance, there is no $A \subseteq \Theta$ such that $\rho(A) = \{T\}$: the mapping ρ does not have an inverse.

Inner and outer reductions



- We can approximate any subset B of Ω by two subsets in Θ:
  - The inner reduction of B:

$$\underline{\rho}^{-1}(B) = \{\theta \in \Theta : \rho(\{\theta\}) \subseteq B\}$$

  - The outer reduction of B:

$$\overline{\rho}^{-1}(B) = \{\theta \in \Theta : \rho(\{\theta\}) \cap B \neq \emptyset\}.$$

- In the example:

$$\underline{\rho}^{-1}(\{T\}) = \underline{\rho}^{-1}(\{T, O\}) = \underline{\rho}^{-1}(\{R, S\}) = \emptyset$$

$$\overline{\rho}^{-1}(\{T\}) = \{V\}, \quad \overline{\rho}^{-1}(\{T, O\}) = \{V, \neg V\}, \quad \overline{\rho}^{-1}(\{R, S\}) = \{\neg V\}$$

Restriction

Definition

The restriction of m^Ω in Θ is obtained by transferring each mass $m^\Omega(B)$ to the outer reduction of B: for all subset A of Θ,

$$m^{\Omega\downarrow\Theta}(A) = \sum_{\overline{\rho}^{-1}(B)=A} m^\Omega(B)$$

- In the example, we thus have

$$m^{\Omega\downarrow\Theta}(\{V\}) = 0.4, \quad m^{\Omega\downarrow\Theta}(\Theta) = 0.3, \quad m^{\Omega\downarrow\Theta}(\{\neg V\}) = 0.3$$

- Remark: the vacuous extension of $m^{\Omega\downarrow\Theta}$ is

$$m^{(\Omega\downarrow\Theta)\uparrow\Omega}(\{G,T\}) = 0.4, \quad m^{(\Omega\downarrow\Theta)\uparrow\Omega}(\Omega) = 0.3$$

$$m^{(\Omega\downarrow\Theta)\uparrow\Omega}(\{R,S,O\}) = 0.3$$

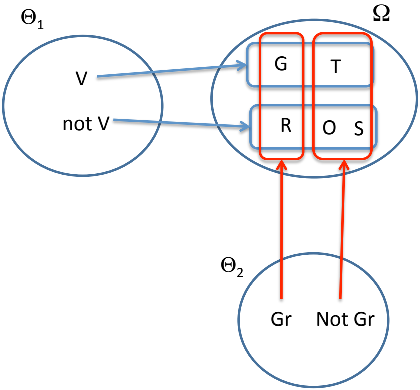
It is less precise that m^Ω: we have lost information when expressing m^Ω in a coarser frame.

Compatible frames of discernment

Definition

Two frames are compatible if they have a common refinement.

Example:



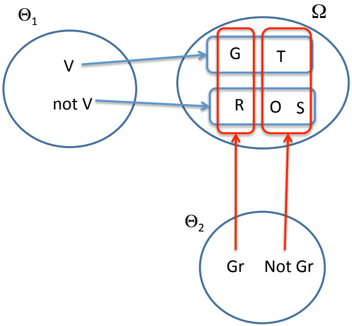
Combination of mass functions on compatible frames

Definition

Let m^{Θ_1} and m^{Θ_2} be two mass functions defined on compatible frames Θ_1 and Θ_2 with common refinement Ω. Their orthogonal sum in Ω is

$$m^{\Theta_1} \oplus m^{\Theta_2} = m^{\Theta_1 \uparrow \Omega} \oplus m^{\Theta_2 \uparrow \Omega}$$

Example



Let

$$m^{\Theta_1}(\{V\}) = 0.3, m^{\Theta_1}(\{\neg V\}) = 0.5,$$

$$m^{\Theta_1}(\{V, \neg V\}) = 0.2$$

and

$$m^{\Theta_2}(\{Gr\}) = 0.4, m^{\Theta_2}(\{\neg Gr\}) = 0.5,$$

$$m^{\Theta_2}(\{Gr, \neg Gr\}) = 0.1$$

Their extensions are

$$m^{\Theta_1 \uparrow \Omega}(\{G, T\}) = 0.3, m^{\Theta_1 \uparrow \Omega}(\{R, O, S\}) = 0.5, m^{\Theta_1 \uparrow \Omega}(\Omega) = 0.2$$

and

$$m^{\Theta_2 \uparrow \Omega}(\{G, R\}) = 0.4, m^{\Theta_2 \uparrow \Omega}(\{T, O, S\}) = 0.5, m^{\Theta_2 \uparrow \Omega}(\Omega) = 0.1$$

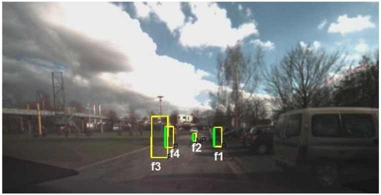
Example (continued)

Calculation of the orthogonal sum:

		$m^{\Theta_2 \uparrow \Omega}$		
		$\{G, R\}, 0.4$	$\{T, O, S\}, 0.5$	$\Omega, 0.1$
	$\{G, T\}, 0.3$	$\{G\}, 0.12$	$\{T\}, 0.15$	$\{G, T\}, 0.03$
$m^{\Theta_1 \uparrow \Omega}$	$\{R, O, S\}, 0.5$	$\{R\}, 0.2$	$\{O, S\}, 0.25$	$\{R, O, S\}, 0.05$
	$\Omega, 0.2$	$\{G, R\}, 0.08$	$\{T, O, S\}, 0.1$	$\Omega, 0.02$

Example: object association

- Let $E = \{e_1, \ldots, e_n\}$ and $F = \{f_1, \ldots, f_p\}$ be two sets of objects perceived by two sensors, or by one sensor at two different times.
- Problem: given information about each object (position, velocity, class, etc.), find a matching between the two sets, in such a way that each object in one set is matched with at most one object in the other set.



Method of approach

1. For each pair of objects $(e_i, f_j) \in E \times F$, use sensor information to build a pairwise mass function $m^{\Theta_{ij}}$ on the frame $\Theta_{ij} = \{h_{ij}, \overline{h}_{ij}\}$, where
   - $h_{ij} \equiv$ "e_i and f_j are the same object", and
   - $\overline{h}_{ij} \equiv$ "e_i and f_j are different objects"
2. Combine the np mass functions $m^{\Theta_{ij}}$
3. Find the matching relation with the highest plausibility.

Building the pairwise mass functions

Using position information

- Assume that each sensor provides an estimated position for each object. Let d_{ij} denote the distance between the estimated positions of e_i and f_j, computed using some distance measure.
- A small value of d_{ij} supports hypothesis h_{ij}, while a large value of d_{ij} supports hypothesis $\overline{h}_{ij}$. Depending on sensor reliability, a fraction of the unit mass should also be assigned to $\Theta_{ij} = \{h_{ij}, \overline{h}_{ij}\}$.
- Model:

$$m_p^{\Theta_{ij}}(\{h_{ij}\}) = \alpha\varphi(d_{ij})$$
$$m_p^{\Theta_{ij}}(\{\overline{h}_{ij}\}) = \alpha\,(1 - \varphi(d_{ij}))$$
$$m_p^{\Theta_{ij}}(\Theta_{ij}) = 1 - \alpha$$

where $\alpha \in [0,1]$ is a degree of confidence in the sensor information and φ is a decreasing function taking values in $[0,1]$.

Building the pairwise mass functions

Using velocity information

- Let us now assume that each sensor returns a velocity vector for each object. Let d'_{ij} denote the distance between the velocities of objects e_i and f_j.
- Here, a large value of d'_{ij} supports the hypothesis $\overline{h}_{ij}$, whereas a small value of d'_{ij} does not support specifically h_{ij} or $\overline{h}_{ij}$, as two distinct objects may have similar velocities.
- Model:

$$m_v^{\Theta_{ij}}(\{\overline{h}_{ij}\}) = \alpha' \psi(d'_{ij})$$
$$m_v^{\Theta_{ij}}(\Theta_{ij}) = 1 - \alpha' \psi(d'_{ij}),$$

where $\alpha' \in [0,1]$ is a degree of confidence in the sensor information and ψ is an increasing function taking values in $[0,1]$.

Building the pairwise mass functions

Using class information

- Let us assume that the objects belong to classes. Let Ω be the set of possible classes, and let m_i and m_j denote mass functions representing evidence about the class membership of objects e_i and f_j.
- If e_i and f_j do not belong to the same class, they cannot be the same object. However, if e_i and f_j do belong to the same class, they may or may not be the same object.
- We can show that the mass function $m_c^{\Theta_{ij}}$ on Θ_{ij} derived from m_i and m_j has the following expression:

$$m_c^{\Theta_{ij}}(\{\overline{h}_{ij}\}) = \kappa_{ij}$$
$$m_c^{\Theta_{ij}}(\Theta_{ij}) = 1 - \kappa_{ij},$$

where κ_{ij} is the degree of conflict between m_i and m_j

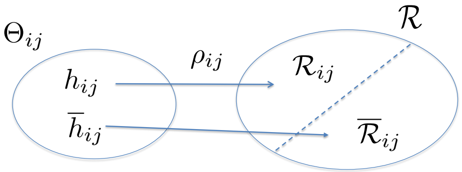
Combination

- For each object pair (e_i, f_j), a pairwise mass function $m^{\Theta_{ij}}$ representing all the available evidence about Θ_{ij} can finally be obtained as:

$$\boxed{m^{\Theta_{ij}} = m_p^{\Theta_{ij}} \oplus m_v^{\Theta_{ij}} \oplus m_c^{\Theta_{ij}}}$$

- How to combine the np mass functions $m^{\Theta_{ij}}$?
- Does there exist a common refinement of the frames Θ_{ij} for (i, j)?

Common refinement



- Let $\mathcal{R}$ be the set of all admissible matching relations, and let $\mathcal{R}_{ij} \subseteq \mathcal{R}$ be the subset of relations R such that $(e_i, f_j) \in R$.
- We can define a refining ρ_{ij} from $2^{\Theta_{ij}}$ to $2^{\mathcal{R}}$. The frames Θ_{ij} are compatible.
- Vacuously extending $m^{\Theta_{ij}}$ in $\mathcal{R}$ yields the following mass function:

$$
\begin{aligned}
m^{\Theta_{ij}\uparrow\mathcal{R}}(\mathcal{R}_{ij}) &= m^{\Theta_{ij}}(\{h_{ij}\}) = \alpha_{ij} \\
m^{\Theta_{ij}\uparrow\mathcal{R}}(\overline{\mathcal{R}_{ij}}) &= m^{\Theta_{ij}}(\{\overline{h}_{ij}\}) = \beta_{ij} \\
m^{\Theta_{ij}\uparrow\mathcal{R}}(\mathcal{R}) &= m^{\Theta_{ij}}(\Theta_{ij}) = 1 - \alpha_{ij} - \beta_{ij}.
\end{aligned}
$$

Combination of contour functions

- The frame $\mathcal{R}$ is very big and computing the orthogonal sum of the *np* mass functions
$$m^{\mathcal{R}} = \bigoplus_{i,j} m^{\Theta_{ij}\uparrow\mathcal{R}}$$
has exponential complexity.
- Instead, we will only compute the combined contour function *pl* corresponding to $m^{\mathcal{R}}$. We recall that
$$pl \propto \prod_{i,j} pl_{ij},$$
where pl_{ij} denote the contour function corresponding to $m^{\Theta_{ij}\uparrow\mathcal{R}}$.

Expression of contour functions

- We have

$$m^{\Theta_{ij}\uparrow\mathcal{R}}(\mathcal{R}_{ij}) = \alpha_{ij}, \quad m^{\Theta_{ij}\uparrow\mathcal{R}}(\overline{\mathcal{R}_{ij}}) = \beta_{ij}, \quad m^{\Theta_{ij}\uparrow\mathcal{R}}(\mathcal{R}) = 1 - \alpha_{ij} - \beta_{ij}.$$

- For all $R \in \mathcal{R}$,

$$pl_{ij}(R) = \begin{cases} 1 - \beta_{ij} & \text{if } R \in \mathcal{R}_{ij}, \\ 1 - \alpha_{ij} & \text{otherwise,} \end{cases}$$
$$= (1 - \beta_{ij})^{R_{ij}} (1 - \alpha_{ij})^{1 - R_{ij}},$$

  where $R_{ij} = 1$ if e_i and f_j are matched and $R_{ij} = 0$ otherwise.
- Consequently, the combined contour function is

$$pl(R) \propto \prod_{i,j} (1 - \beta_{ij})^{R_{ij}} (1 - \alpha_{ij})^{1 - R_{ij}}.$$

Finding the most plausible matching

- We have

$$\ln pl(R) = \sum_{i,j} [R_{ij} \ln(1-\beta_{ij}) + (1-R_{ij}) \ln(1-\alpha_{ij})] + C$$
$$= \sum_{i,j} R_{ij} \ln \frac{1-\beta_{ij}}{1-\alpha_{ij}} + C'$$

- The most plausible relation R^* can thus be found by solving the following binary linear optimization problem:

$$\max \sum_{i,j} R_{ij} \ln \frac{1-\beta_{ij}}{1-\alpha_{ij}}$$

subject to $R_{ij} \in \{0,1\}, \forall (i,j)$, $\sum_{j=1}^{p} R_{ij} \leq 1, \forall i$ and $\sum_{i=1}^{n} R_{ij} \leq 1, \forall j$.

- This problem can be shown to be equivalent to a linear assignment problem and can be solved in $O(\max(n,p)^3)$ time.

References

cf. http://www.hds.utc.fr/~tdenoeux

T. Denœux, D. Dubois and H. Prade.
Representations of Uncertainty in Artificial Intelligence: Beyond Probability and Possibility. In P. Marquis et al. (Eds), "A Guided Tour of Artificial Intelligence Research", Volume 1, Chapter 4, Springer Verlag, pages 119-150, 2020.

T. Denœux.
Conjunctive and Disjunctive Combination of Belief Functions Induced by Non Distinct Bodies of Evidence. *Artificial Intelligence*, Vol. 172, pages 234-264, 2008.

T. Denœux, N. El Zoghby, V. Cherfaoui and A. Jouglet.
Optimal object association in the Dempster-Shafer framework. *IEEE Transactions on Cybernetics*, Vol. 44, Issue 22, pages 2521-2531, 2014.