Introduction to belief functions

Thierry Denœux

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

Fifth School on Belief Functions and their Applications
Sienna, Italy, October 27, 2019

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

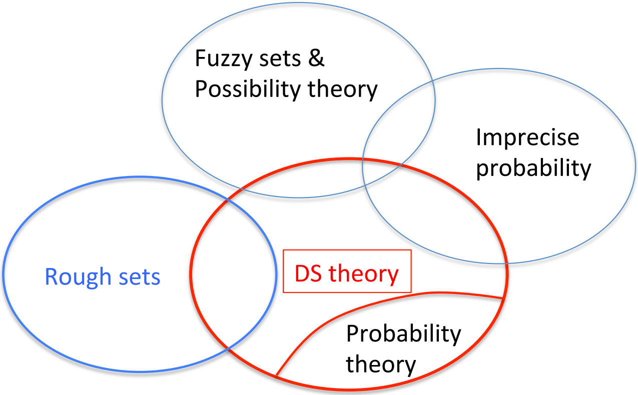
- A formal framework for representing and reasoning with uncertain information.
- Also known as Dempster-Shafer (DS) theory or Evidence theory.
- Originates from the work of Dempster (1968) in the context of statistical inference.
- Formalized by Shafer (1976) as a theory of evidence.
- Popularized and 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.

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

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

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

Mass function

Definition

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

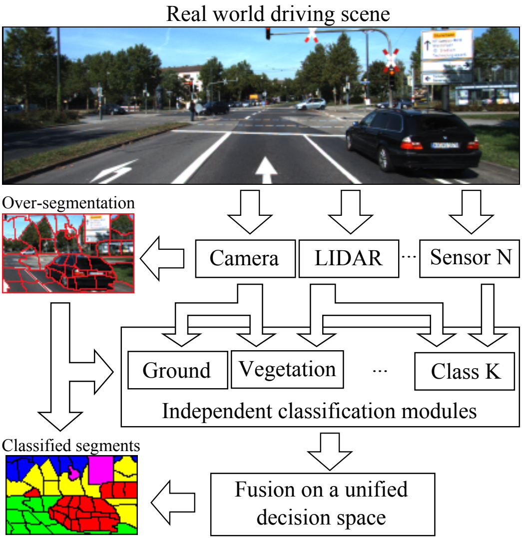
Let Ω be a finite set called a frame of discernment. A mass function on Ω is a mapping $m : 2^\Omega \to [0, 1]$ such that

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

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.

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

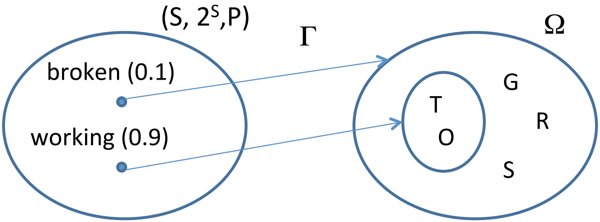
Example: road scene analysis



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
  - $m(\Omega) = 0.1$ is the probability of knowing nothing.
- In general, what is the meaning (semantics) of a mass function in DS theory?
- A precise interpretation was proposed by Shafer (1981): random code semantics.

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 "$X \in A$":

$$m(A) = P(\{c \in S \mid \Gamma(c) = A\}) = \sum_{i=1}^n p_i I(\Gamma(c_i) = A),$$

where $I(\cdot)$ is the indicator function.

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 is elicited by finding the “coded-message” canonical example that is the most similar to our evidence.

Random set

- The tuple $(S, 2^S, P, \Gamma)$, where
  - $(S, 2^S, P)$ is a probability space and
  - Γ is a mapping from S to 2^Ω

  is called a random set.
- We have seen that, given the random set $(S, 2^S, P, \Gamma)$, we can define the mass function $m : 2^\Omega \to [0, 1]$ such that

$$m(A) = P(\{c \in S \mid \Gamma(c) = A\})$$

- Conversely, given any mass function $m : 2^\Omega \to [0, 1]$, we can define the random set $(S, 2^S, P, \Gamma)$ with

$$S = 2^\Omega,$$

$$P(\{A\}) = m(A), \quad A \subseteq \Omega,$$

  and

$$\Gamma(A) = A, \quad A \subseteq \Omega.$$

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.

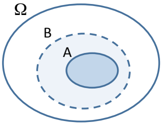
Definition (Bayesian mass function)

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

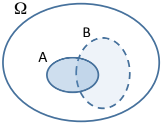
Outline

Certainty and possibility

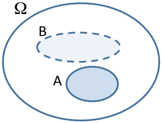
- Assume our evidence tells us that $X \in A$ for sure and nothing more, for some $A \subseteq \Omega$. It is represented by the logical mass function $m_{[A]}$.
- Let $B \subseteq \Omega$. What can we say about the proposition "$X \in B$"?



- If $A \subseteq B$, we know for sure that $X \in B$. This proposition is said to be certain. (It is supported/implied by the evidence)



- If $A \cap B \neq \emptyset$, we cannot exclude that $X \in B$. This proposition is said to be possible. (It is consistent with the evidence)



- If $A \cap B = \emptyset$, the proposition "$X \in B$" is impossible. (It is inconsistent with the evidence)

Belief function

- Let us now consider an arbitrary mass function m with (nonempty) focal sets $A_1, \ldots, A_n$.
- Let $B \subseteq \Omega$. If we know for sure that $X \in A_i$, the proposition $X \in B$ is supported by the evidence whenever $A_i \subseteq B$.
- The probability that the proposition $X \in B$ is supported by the evidence is

$$Bel(B) = \sum_{i=1}^{n} m(A_i) I(A_i \subseteq B).$$

- The number $Bel(B)$ is called the credibility of (degree of belief in) B, and the mapping $Bel : 2^\Omega \to [0,1]$ is called the belief function induced by m.
- Elementary properties: $Bel(\emptyset) = 0$, $Bel(\Omega) = 1$.

Plausibility function

- We can also compute the probability that the proposition $X \in B$ is consistent with the evidence as

$$Pl(B) = \sum_{i=1}^{n} m(A_i) I(A_i \cap B \neq \emptyset).$$

- The number $Pl(B)$ is called the plausibility of B, and the mapping $Pl : 2^\Omega \rightarrow [0,1]$ is called the plausibility function induced by m.
- Elementary properties:
  - $Pl(\emptyset) = 0$, $Pl(\Omega) = 1$
  - For all $B \subseteq \Omega$, $Bel(B) \leq Pl(B)$
  - For any $A, B \subseteq \Omega$, $(A \cap B = \emptyset \Leftrightarrow A \subseteq \overline{B})$. Consequently,

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

- Function $pl : \Omega \rightarrow [0,1]$ such that $pl(\omega) = Pl(\{\omega\})$ is called the contour function of m.

Two-dimensional representation

- The uncertainty on a proposition B is represented by two numbers: $Bel(B)$ and $Pl(B)$, with $Bel(B) \leq Pl(B)$.
- The intervals $[Bel(B), Pl(B)]$ have maximum length when m is the vacuous mass function. Then,

$$[Bel(B), Pl(B)] = [0, 1]$$

  for all subset B of Ω, except $\emptyset$ and Ω.
- The intervals $[Bel(B), Pl(B)]$ are reduced to points when m is Bayesian. Then,

$$Bel(B) = Pl(B)$$

  for all B, and $Bel = Pl$ is a probability measure.

Broken sensor example

- From
$$m(A) = 0.9, \quad m(\Omega) = 0.1$$
we get

	A	$\overline{A}$	Ω
Bel	0.9	0	1
Pl	1	0.1	1

- We observe that
$$Bel(\Omega) = Bel(A \cup \overline{A}) \geq Bel(A) + Bel(\overline{A})$$
and
$$Pl(\Omega) = Pl(A \cup \overline{A}) \leq Pl(A) + Pl(\overline{A})$$
- *Bel* and *Pl* are nonadditive measures. (*Bel* is superadditive and *Pl* is subadditive).

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:

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

- Pl is this a possibility measure, and Bel is the dual necessity measure.
- The possibility distribution is the contour function:

$$pl(x) = Pl(\{x\}), \quad \forall x \in \Omega$$

- The theory of belief function can thus be considered as more expressive than possibility theory (but the combination operations are different, as we will see later).

Credal set

- 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!

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:

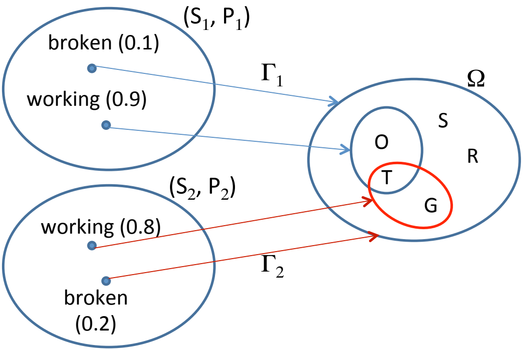
$$m_1(\{T, O\}) = 0.9, \quad m_1(\Omega) = 0.1$$

- 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 $P_1(\{s_1\})P_2(\{s_2\})$.

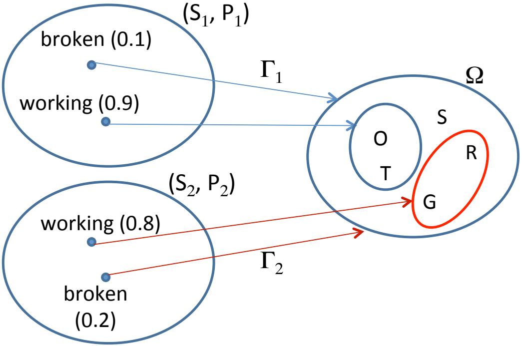
Computation

$m_1 \backslash m_2$	$\{T, G\}$ (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

- The orthogonal sum of two mass functions m_1 and m_2 on Ω is the mass function $m_1 \oplus m_2$ defined as $(m_1 \oplus m_2)(\emptyset) = 0$ and

$$(m_1 \oplus m_2)(A) = \frac{1}{1-\kappa} \sum_{B \cap C = A} m_1(B) m_2(C), \quad \forall A \neq \emptyset,$$

where

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

is the degree of conflict between m_1 and m_2.

- If $\kappa = 1$, m_1 and m_2 are not combinable.

Properties of Dempster's rule

- Commutativity, associativity. Neutral element: $m_?$
- Generalization of intersection: if $m_{[A]}$ and $m_{[B]}$ are logical mass functions and $A \cap B \neq \emptyset$, then
$$m_{[A]} \oplus m_{[B]} = m_{[A \cap B]}$$
- If either m_1 or m_2 is Bayesian, then so is $m_1 \oplus m_2$ (as the intersection of a singleton with another subset is either a singleton, or the empty set).

Dempster's conditioning

- Conditioning is a special case, 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 $Q_1 \oplus Q_2$ be the commonality function associated to $m_1 \oplus m_2$.
- We have

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

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

- In particular, $pl(\omega) = Q(\{\omega\})$. Consequently,

$$pl_1 \oplus pl_2 = (1-\kappa)^{-1} pl_1 pl_2.$$

Remarks on normalization

- Mass functions expressing pieces of evidence are always normalized.
- Smets introduced the unnormalized Dempster's rule (TBM conjunctive rule $\textcircled{\cap}$), which may yield an unnormalized mass function.
- He proposed to interpret $m(\emptyset)$ as the mass committed to the hypothesis that X might not take its value in Ω (open-world assumption).
- I now think that this interpretation is problematic, as $m(\emptyset)$ increases "mechanically" when combining more and more items of evidence.
- Claim: unnormalized mass functions are convenient mathematically as equivalent representations of normalized mass functions, but only normalized mass functions make sense.
- In particular, *Bel* and *Pl* should always be computed from normalized mass functions.

TBM disjunctive rule

- Let m_1 and m_2 be two mass functions induced by random messages (S_1, P_1, Γ_1) and (S_2, P_2, Γ_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 TBM disjunctive rule:

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

- $Bel_1 \bigcirc\!\!\!\!\!\cup\, Bel_2 = Bel_1 \cdot Bel_2$

Outline

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

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

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: 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

Motivations

- The basic rules $\oplus$ and $\textcircled{\cup}$ assume the sources of information to be independent, e.g.
  - experts with non overlapping experience/knowledge
  - non overlapping datasets
- What to do in case of non independent 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.

Cautious rule

Problem

- The above approach works for special cases.
- Example (Dubois, Prade, Smets 2001): 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.

Simple mass functions

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

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

  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:

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

- Remark: In earlier work, following Smets' terminology, I used the term "weight" for $\delta(A)$. I now think it is better to reserve the term "weight" for additive quantities. In recent work, Faux and Dubois use the term "diffidence" for $\delta(A)$.

Separable mass functions

Definition (Separable mass function)

A (normalized) mass function is separable *if it can be written as the $\oplus$ combination of simple mass functions:*

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

with $w(A) \geq 0$ 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 obtained as
$$w(A) = \sum_{B \supseteq A} (-1)^{|B|-|A|} \ln Q(B) \quad (1)$$
for all $A \subseteq \Omega$.
- For any non dogmatic 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 also be recovered 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
$$m_1 \oplus m_2 = \bigoplus_{\emptyset \neq A \subset \Omega} A^{w_1(A)+w_2(A)}.$$
- Characterization of the w-ordering
$$m_1 \sqsubseteq_w m_2 \Leftrightarrow w_1(A) \geq w_2(A), \quad \forall A \subset \Omega, A \neq \emptyset.$$

Cautious rule

Definition

Let m_1 and m_2 be two non dogmatic mass functions with weight functions w_1 and w_2.

Proposition

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
$m_1 \owedge m_2$	$\longleftarrow$	$\max(w_1, w_2)$

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

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

The common item of evidence m_1 is not counted twice!

- No neutral element, but $m_? \owedge m = m$ iff m is separable

Basic rules

Sources	independent	dependent
All reliable	$\oplus$	$\owedge$
At least one reliable	$\textcircled{\cup}$	$\ovee$

$\ovee$ is the bold disjunctive rule

Outline

Refinement and coarsening

Example

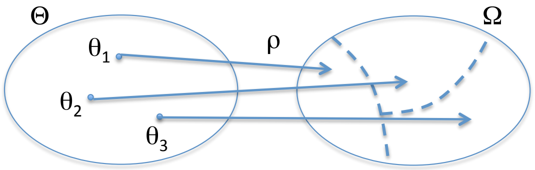
- Let us come back to the road scene analysis example, with $\Omega = \{G, R, T, O, S\}$.
- Assume that we have a vegetation detector, which can determine if a region of the image contains vegetation or not. For this detector, the frame of discernment is $\Theta = \{V, \neg V\}$, where V means that there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

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

- 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 $\rho : 2^\Theta \to 2^\Omega$ (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,

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

- We finally 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.

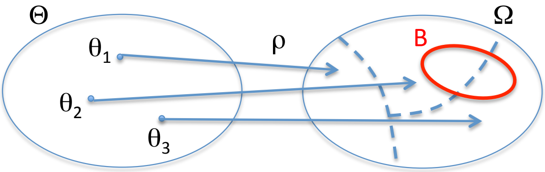
Expression of information in a coarser frame

- 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:

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

  - The outer reduction of B:

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

- 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

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

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

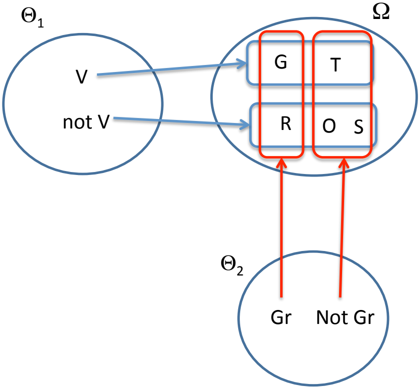
It is less precise that m^Ω: we have lost information when expressing m^Ω in a coarser frame.

Compatible frames of discernment

Definition

Two frames are compatible if they have a common refinement.

Example:



Combination of mass functions on compatible frames

- 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.
- The orthogonal sum of m^{Θ_1} and m^{Θ_2} in Ω is

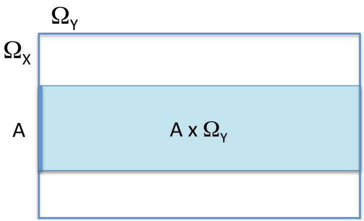
$$m^{\Theta_1} \oplus m^{\Theta_2} = m^{\Theta_1 \uparrow \Omega} \oplus m^{\Theta_2 \uparrow \Omega}$$

- Example: assume that $m^{\Theta_1}(\{V\}) = 0.3$, $m^{\Theta_1}(\{\neg V\}) = 0.5$, $m^{\Theta_1}(\{V, \neg V\}) = 0.2$, and $m^{\Theta_2}(\{Gr\}) = 0.4$, $m^{\Theta_2}(\{\neg Gr\}) = 0.5$, $m^{\Theta_2}(\{Gr, \neg Gr\}) = 0.1$. Compute $m^{\Theta_1} \oplus m^{\Theta_2}$.

Case of product frames

Cylindrical extension

- Let us now assume that we have two frames Ω_X and Ω_Y related to two different questions about, e.g., the values of two unknown variables X and Y.
- Let $\Omega_{XY} = \Omega_X \times \Omega_Y$ be the product space. It is a refinement of both Ω_X and Ω_Y.



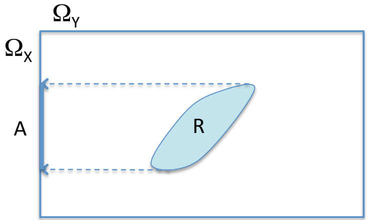
- We can define the following refining ρ from 2^{Ω_X} to $2^{\Omega_{XY}}$:

$$\rho(A) = A \times \Omega_Y,$$

for all $A \subseteq \Omega_X$. The set $\rho(A)$ is called the cylindrical extension of A in Ω_{XY} and is denoted by $A \uparrow \Omega_{XY}$.

Case of product frames

Projection



- Conversely, let R be a subset of Ω_{XY}.
- Its outer reduction is

$$\begin{aligned}\overline{\rho}^{-1}(R) &= \{x \in \Omega_X \mid \rho(\{x\}) \cap R \neq \emptyset\} \\ &= \{x \in \Omega_X \mid \exists y \in \Omega_Y, (x,y) \in R\}.\end{aligned}$$

- This set is denoted by $R \downarrow \Omega_X$ and is called the projection of R on Ω_X

Case of product frames

Vacuous extension and marginalization

- The vacuous extension of a mass function m^X from Ω_X to Ω_{XY} is obtained by transferring each mass $m^X(B)$ for any subset B of Ω_X to the cylindrical extension of B:

$$m^{X\uparrow XY}(A) = \begin{cases} m^X(B) & \text{if } A = B \times \Omega_Y \\ 0 & \text{otherwise.} \end{cases}$$

- Conversely, the restriction of a joint mass function m^{XY} on Ω_{XY} is

$$m^{XY\downarrow X}(A) = \sum_{B\downarrow\Omega_X = A} m^{XY}(B),$$

for all $A \subseteq \Omega_X$. The mass functions $m^{XY\downarrow X}$ and $m^{XY\downarrow Y}$ are called the marginals of m^{XY} and the operation that computes the marginals from a joint mass function is called marginalization. This operation extends both set projection and probabilistic marginalization.

Application to approximate reasoning

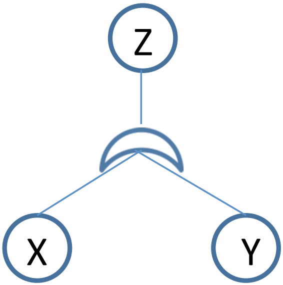
- Assume that we have:
  - Partial knowledge of X formalized as a mass function m^X
  - A joint mass function m^{XY} representing an uncertain relation between X and Y
- What can we say about Y?
- Solution:

$$m^Y = \left(m^{X\uparrow XY} \oplus m^{XY}\right)^{\downarrow Y}.$$

- Infeasible with many variables and large frames of discernment, but efficient algorithms exist to carry out the operations in frames of minimal dimensions.

Example

- A machine fails if any one of two components fails.
- Let Z, X and Y be the binary variables describing the states of the two components, and the machine.



- We have the following prior knowledge about the states of the components:

$$m^X(\{1\}) = 0.1, m^X(\{0\}) = 0.3,$$

$$m^X(\{0,1\}) = 0.6$$

$$m^Y(\{0,1\}) = 1$$

- We observe that the machine fails. What are our beliefs about the states of the two components?

Solution

- Pieces of evidence:

$$m_0^{XYZ}(\{(1,1,1),(1,0,1),(0,1,1),(0,0,0)\}) = 1$$

$$m^{X\uparrow XYZ}(\{1\}\times\Omega_{YZ}) = 0.1,\ m^{X\uparrow XYZ}(\{0\}\times\Omega_{YZ}) = 0.3,\ m^{X\uparrow XYZ}(\Omega_{XYZ}) = 0.6$$

$$m^{Y\uparrow XYZ}(\Omega_{XYZ}) = 1, \quad m^{Z\uparrow XYZ}(\Omega_{XY} \times \{1\}) = 1$$

- Let $m_1^{XYZ} = m_0^{XYZ} \oplus m^{X\uparrow XYZ} \oplus m^{Z\uparrow XYZ}$. We have

$$m_1^{XYZ}(\{(1,1,1),(1,0,1)\}) = 0.1,\ m_1^{XYZ}(\{(0,1,1)\}) = 0.3,$$

$$m_1^{XYZ}(\{(1,1,1),(1,0,1),(0,1,1)\}) = 0.6$$

- Marginalizing on X and Y, we get

$$m_1^{XYZ\downarrow X}(\{1\}) = 0.1, m_1^{XYZ\downarrow X}(\{0\}) = 0.3, m_1^{XYZ\downarrow X}(\{0,1\}) = 0.6$$

$$m_1^{XYZ\downarrow Y}(\{1\}) = 0.3, m_1^{XYZ\downarrow Y}(\{0,1\}) = 0.7$$

References

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

G. Shafer.
A mathematical theory of evidence. Princeton University Press, Princeton, N.J., 1976.

Ph. Smets and R. Kennes.
The Transferable Belief Model.
Artificial Intelligence, 66:191-243, 1994.

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

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