Introduction to belief functions

Thierry Denœux[1]

[1]Université de Technologie de Compiègne
HEUDIASYC (UMR CNRS 6599)
http://www.hds.utc.fr/~tdenoeux

Spring School BFTA 2011
Autrans, April 4-8, 2011

Contents of this lecture

1. Context, position of belief functions with respect to classical theories of uncertainty.
2. Fundamental concepts: belief, plausibility, commonality, Conditioning, basic combination rules.
3. Some more advanced concepts: least commitment principle, cautious rule, multidimensional belief functions.

Uncertain reasoning

- In science and engineering we always need to reason with partial knowledge and uncertain information (from sensors, experts, models, etc.).
- Different kinds of uncertainty:
  - Aleatory uncertainty induced by the variability of entities in populations and outcomes of random (repeatable) experiments. Example: drawing a ball from an urn. Cannot be reduced;
  - Epistemic uncertainty, due to lack of knowledge. Example: inability to distinguish the color of a ball because of color blindness. Can be reduced.
- Classical frameworks for reasoning with uncertainty:
  1. Probability theory;
  2. Set-membership approach.

Probability theory
Interpretations

- Probability theory can be used to represent:
  - Aleatory uncertainty: probabilities are considered as objective quantities and interpreted as frequencies or limits of frequencies;
  - Epistemic uncertainty: probabilities are subjective, interpreted as degrees of belief.
- Main objections against the use of probability theory as a model epistemic uncertainty (Bayesian model):
  - Inability to represent ignorance;
  - Not a plausible model of how people make decisions based on weak information.

Inability to represent ignorance
The wine/water paradox

- Principle of Indifference (PI): in the absence of information about some quantity X, we should assign equal probability to any possible value of X.
- The wine/water paradox:

There is a certain quantity of liquid. All that we know about the liquid is that it is composed entirely of wine and water, and the ratio of wine to water is between 1/3 and 3. What is the probability that the ratio of wine to water is less than or equal to 2?

Inability to represent ignorance

The wine/water paradox (continued)

- Let X denote the ratio of wine to water. All we know is that $X \in [1/3, 3]$. According to the PI, $X \sim \mathcal{U}_{[1/3,3]}$. Consequently:

$$P(X \leq 2) = (2 - 1/3)/(3 - 1/3) = 5/8.$$

- Now, let $Y = 1/X$ denote the ratio of water to wine. Similarly, we only know that $Y \in [1/3, 3]$. According to the PI, $Y \sim \mathcal{U}_{[1/3,3]}$. Consequently:

$$\begin{aligned} P(X \leq 2) &= P(Y \geq 1/2) \\ &= (3 - 1/2)/(3 - 1/3) = 15/16. \end{aligned}$$

Decision making
Ellsberg's paradox

- Suppose you have an urn containing 30 red balls and 60 balls, either black or yellow. You are given a choice between two gambles:
  - A: You receive 100 euros if you draw a red ball;
  - B: You receive 100 euros if you draw a black ball.
- Also, you are given a choice between these two gambles (about a different draw from the same urn):
  - C: You receive 100 euros if you draw a red or yellow ball;
  - D: You receive 100 euros if you draw a black or yellow ball.
- Most people strictly prefer A to B, hence $P(red) > P(black)$, but they strictly prefer D to C, hence

$$P(black) + P(yellow) > P(red) + P(yellow)$$
$$\Rightarrow P(black) > P(red).$$

Set-membership approach

- Partial knowledge about some variable X is described by a set of possible values E (constraint).
- Example:
  - Consider a system described by the equation

    $$y = f(x_1, \ldots, x_n; \theta)$$

    where y is the output, $x_1, \ldots, x_n$ are the inputs and θ is a parameter.
  - Knowing that $x_i \in [\underline{x}_i, \overline{x}_i]$, $i = 1, \ldots, n$ and $\theta \in [\underline{\theta}, \overline{\theta}]$, find a set $\mathbb{X}$ surely containing x.
- Advantage: computationally simpler than the probabilistic approach in many cases (interval analysis).
- Drawback: no way to express doubt, conservative approach.

Theory of belief functions

- Alternative theories of uncertainty:
  - Possibility theory (Zadeh, 1978; Dubois and Prade 1980's-1990's);
  - Imprecise probability theory (Walley, 1990's);
  - Theory of belief functions (Dempster-Shafer theory, Evidence theory, Transferable Belief Model) (Dempster, 1968; Shafer, 1976; Smets 1980's-1990's).
- The theory of belief functions extends both the Set-membership approach and Probability Theory:
  - A belief function may be viewed both as a generalized set and as a non additive measure.
  - The theory includes extensions of probabilistic notions (conditioning, marginalization) and set-theoretic notions (intersection, union, inclusion, etc.)

Outline

Outline

1. Basics
   - Belief representation
   - Information fusion
   - Decision making
2. Selected advanced topics
   - Informational orderings
   - Cautious rule
   - Multidimensional belief functions

Mass function

Definition

- Let X be a variable taking values in a finite set Ω (frame of discernment).
- Mass function: $m : 2^{\Omega} \to [0, 1]$ such that

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

- Every A of Ω such that $m(A) > 0$ is a focal set of m.
- m is said to be normalized if $m(\emptyset) = 0$. This condition may be required or not.

Murder example

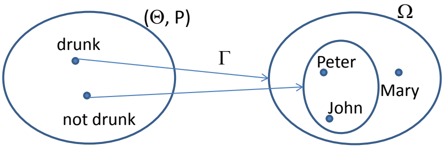
- A murder has been committed. There are three suspects: $\Omega = \{Peter, John, Mary\}$.
- A witness saw the murderer going away in the dark, and he can only assert that it was man. How, we know that the witness is drunk 20 % of the time.
- This piece of evidence can be represented by

$$m(\{Peter, John\}) = 0.8,$$

$$m(\Omega) = 0.2$$

- The mass 0.2 is not committed to $\{Mary\}$, because the testimony does not accuse Mary at all!

Mass function
Multi-valued mapping interpretation



- A mass function m on Ω may be viewed as arising from
  - A set $\Theta = \{\theta_1, \ldots, \theta_r\}$ of interpretations;
  - A probability measure P on Θ;
  - A multi-valued mapping $\Gamma : \Theta \rightarrow 2^\Omega$.
- Meaning: under interpretation θ_i, the evidence tells us that $X \in \Gamma(\theta_i)$, and nothing more. The probability $P(\{\theta_i\})$ is transferred to $A_i = \Gamma(\theta_i)$.
- $m(A)$ is the probability of knowing only that $X \in A$, given the available evidence.

Mass functions
Special cases

- Only one focal set:

$$m(A) = 1 \text{ for some } A \subseteq \Omega$$

  $\rightarrow$ categorical (logical) mass function ($\sim$ set). Special case: $A = \Omega$, vacuous mass function, represents total ignorance.

- All focal sets are singletons:

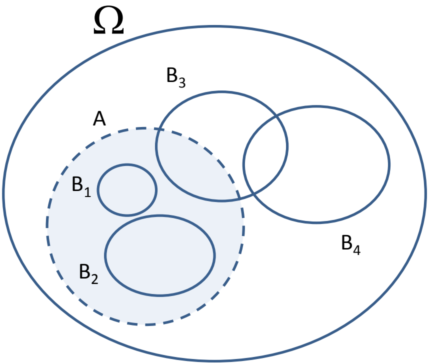
$$m(A) > 0 \Rightarrow |A| = 1$$

  $\rightarrow$ Bayesian mass function ($\sim$ probability mass function).

- A mass function can thus be seen as
  - a generalized set;
  - a generalized probability distribution.

Belief and plausibility functions

Definitions



$$bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B)$$

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B),$$

$$pl(A) \geq bel(A), \quad \forall A \subseteq \Omega.$$

Belief and plausibility functions

Interpretation and special cases

- Interpretations:
  - $bel(A) =$ degree to which the evidence supports A.
  - $pl(A) =$ upper bound on the degree of support that could be assigned to A if more specific information became available.
- Special case: if m is Bayesian, $bel = pl$ (probability measure).

Murder example

A	$\emptyset$	$\{P\}$	$\{J\}$	$\{P,J\}$	$\{M\}$	$\{P,M\}$	$\{J,M\}$	Ω
$m(A)$	0	0	0	0.8	0	0	0	0.2
$bel(A)$	0	0	0	0.8	0	0	0	1
$pl(A)$	0	1	1	1	0.2	1	1	1

- We observe that

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

$$pl(A \cup B) \leq pl(A) + pl(B) - bel(A \cap B)$$

- *bel* and *pl* are non additive measures.

Wine/water paradox revisited

- Let X denote the ratio of wine to water. All we know is that $X \in [1/3, 3]$. This is modeled by the categorical mass function m_X such that $m_X([1/3, 3]) = 1$. Consequently:

$$bel_X([2, 3]) = 0, \quad pl_X([2, 3]) = 1.$$

- Now, let $Y = 1/X$ denote the ratio of water to wine. All we know is that $Y \in [1/3, 3]$. This is modeled by the categorical mass function m_Y such that $m_Y([1/3, 3]) = 1$. Consequently:

$$bel_Y([1/3, 1/2]) = 0, \quad pl_Y([1/3, 1/2]) = 1.$$

Relations between *m*, *bel* et *pl*

- Relations:

$$bel(A) = pl(\Omega) - pl(\overline{A}), \quad \forall A \subseteq \Omega$$

$$m(A) = \begin{cases} \sum_{\emptyset \neq B \subseteq A} (-1)^{|A|-|B|} bel(B), & A \neq \emptyset \\ 1 - bel(\Omega) & A = \emptyset \end{cases}$$

- *m*, *bel* et *pl* are thus three equivalent representations of
  - a piece of evidence or, equivalently,
  - a state of belief induced by this evidence.

Relationship with Possibility theory

- Assume that the focal sets of m are nested: $A_1 \subset A_2 \subset \ldots \subset A_r \rightarrow m$ is said to be consonant.
- The following relations 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:

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

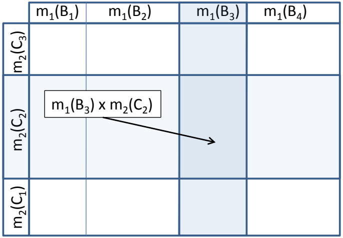
- The theory of belief function can thus be considered as more expressive than possibility theory.

Outline

1. Basics
   - Belief representation
   - Information fusion
   - Decision making
2. Selected advanced topics
   - Informational orderings
   - Cautious rule
   - Multidimensional belief functions

Conjunctive combination
Definitions

Let m_1 and m_2 be two mass functions on Ω induced by two independent items of evidence.



1. Unnormalized Dempster's rule

$$(m_1 \cap\!\!\!\!\bigcirc\, m_2)(A) = \sum_{B \cap C = A} m_1(B) m_2(C)$$

2. Normalized Dempster's rule

$$(m_1 \oplus m_2)(A) = \begin{cases} \frac{(m_1 \cap\!\!\!\!\bigcirc\, m_2)(A)}{1 - K_{12}} & \text{if } A \neq \emptyset \\ 0 & \text{if } A = \emptyset \end{cases}$$

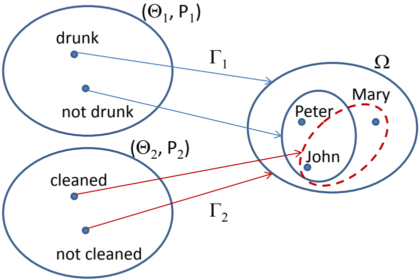
$K_{12} = (m_1 \cap\!\!\!\!\bigcirc\, m_2)(\emptyset)$: degree of conflict.

Dempster's rule
Example

- We have $m_1(\{Peter, John\}) = 0.8$, $m_1(\Omega) = 0.2$.
- New piece of evidence: a blond hair has been found. There is a probability 0.6 that the room has been cleaned before the crime $\rightarrow m_2(\{John, Mary\}) = 0.6$, $m_2(\Omega) = 0.4$.

	$\{Peter, John\}$ 0.8	Ω 0.2
$\{John, Mary\}$ 0.6	$\{John\}$ 0.48	$\{John, Mary\}$ 0.12
Ω 0.4	$\{Peter, John\}$ 0.32	Ω 0.08

Dempster's rule
Justification



- Let $(\Theta_1, P_1, \Gamma_1)$ and $(\Theta_2, P_2, \Gamma_2)$ be the multi-valued mappings associated to m_1 and m_2.
- If $\theta_1 \in \Theta_1$ and $\theta_2 \in \Theta_2$ both hold, then $X \in \Gamma_1(\theta_1) \cap \Gamma_2(\theta_2)$.
- If the two pieces of evidence are independent, then this happens with probability $P_1(\{\theta_1\})P_2(\{\theta_2\})$.
- The normalized rule is obtained after conditioning on the event $\{(\theta_1, \theta_2) | \Gamma_1(\theta_1) \cap \Gamma_2(\theta_2) \neq \emptyset\}$.

Dempster's rule
Properties

- Commutativity, associativity. Neutral element: m_Ω.
- Generalization of intersection: if m_A and m_B are categorical mass functions, then

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

- Generalization of probabilistic conditioning: if m is a Bayesian mass function and m_A is a categorical mass function, then $m \oplus m_A$ is a Bayesian mass function that corresponding to the conditioning of m by A.
- Notations for conditioning (special case):

$$m \textcircled{\cap} m_A = m(\cdot|A), \quad m \oplus m_A = m^*(\cdot|A).$$

Dempster's rule
Expression using commonalities

- 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,

$$m(A) = \sum_{B \supseteq A} (-1)^{|B \setminus A|} q(B), \quad \forall A \subseteq \Omega.$$

- Interpretation: $q(A) = m(A|A)$, for any $A \subseteq \Omega$.
- Expression of the unnormalized Dempster's rule using commonalities:

$$(q_1 \textcircled{\cap} q_2)(A) = q_1(A) \cdot q_2(A), \quad \forall A \subseteq \Omega.$$

TBM disjunctive rule

Definition and justification

- Let $(\Theta_1, P_1, \Gamma_1)$ and $(\Theta_2, P_2, \Gamma_2)$ be the multi-valued mapping frameworks associated to two pieces of evidence.
- If interpretation $\theta_k \in \Theta_k$ holds and piece of evidence k is reliable, then we can conclude that $X \in \Gamma_k(\theta_k)$.
- If interpretation $\theta_1 \in \Theta_1$ and $\theta_2 \in \Theta_2$ both hold and we assume that at least one of the two pieces of evidence is reliable, then we can conclude that $X \in \Gamma_1(\theta_1) \cup \Gamma_2(\theta_2)$.
- This leads to the TBM disjunctive rule:

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

TBM disjunctive rule
Properties

- Commutativity, associativity.
- Neutral element: $m_\emptyset$
- Let $b = bel + m(\emptyset)$ (implicability function). We have:

$$(b_1 \bigcirc\!\!\!\!\!\cup\, b_2) = b_1 \cdot b_2$$

- De Morgan laws for $\bigcirc\!\!\!\!\!\cap$ and $\bigcirc\!\!\!\!\!\cup$:

$$\begin{aligned}\overline{m_1 \bigcirc\!\!\!\!\!\cup\, m_2} &= \overline{m_1} \bigcirc\!\!\!\!\!\cap\, \overline{m_2},\\ \overline{m_1 \bigcirc\!\!\!\!\!\cap\, m_2} &= \overline{m_1} \bigcirc\!\!\!\!\!\cup\, \overline{m_2},\end{aligned}$$

where $\overline{m}$ denotes the complement of m defined by $\overline{m}(A) = m(\overline{A})$ for all $A \subseteq \Omega$.

Selecting a combination rule

- All three rules $\textcircled{\cap}$, $\oplus$ and $\textcircled{\cup}$ assume the pieces of evidence to be independent.
- The conjunctive rules $\textcircled{\cap}$ and $\oplus$ further assume that the pieces of evidence are both reliable;
- The TBM disjunctive rule $\textcircled{\cup}$ only assumes that at least one of the items of evidence combined is reliable (weaker assumption).
- $\textcircled{\cap}$ vs. $\oplus$:
  - $\textcircled{\cap}$ keeps track of the conflict between items of evidence: very useful in some applications.
  - $\textcircled{\cap}$ also makes sense under the open-world assumption.
  - The conflict increases with the number of combined mass functions: normalization is often necessary at some point.
- What to do with dependent items of evidence? → Cautious rule

Outline

Decision making

Problem formulation

- A decision problem can be formalized by defining:
  - A set of acts $\mathcal{A} = \{a_1, \ldots, a_s\}$;
  - A set of states of the world Ω;
  - A loss function $L : \mathcal{A} \times \Omega \rightarrow \mathbb{R}$, such that $L(a, \omega)$ is the loss incurred if we select act a and the true state is ω.
- Bayesian framework
  - Uncertainty on Ω is described by a probability measure P;
  - Define the risk of each act a as the expected loss if a is selected:

$$R(a) = \mathbb{E}_P[L(a, \cdot)] = \sum_{\omega \in \Omega} L(a, \omega) P(\{\omega\}).$$

  - Select an act with minimal risk.
- Extension to the belief function framework?

Decision making
Compatible probabilities

- Let m be a normalized mass function, and $\mathcal{P}(m)$ the set of compatible probability measures on Ω, i.e., the set of P verifying

$$bel(A) \leq P(A) \leq pl(A), \quad \forall A \subseteq \Omega.$$

- The lower and upper expected risk of each act a are defined, respectively, as:

$$\underline{R}(a) = \underline{\mathbb{E}}_m[L(a,\cdot)] = \inf_{P \in \mathcal{P}(m)} R_P(a) = \sum_{A \subseteq \Omega} m(A) \min_{\omega \in A} L(a,\omega)$$

$$\overline{R}(a) = \overline{\mathbb{E}}_m[L(a,\cdot)] = \sup_{P \in \mathcal{P}(m)} R_P(a) = \sum_{A \subseteq \Omega} m(A) \max_{\omega \in A} L(a,\omega)$$

Decision making

Strategies

- For each act a we have a risk interval $[\underline{R}(a), \overline{R}(a)]$. How to compare these intervals?
- Three strategies:
  1. a is preferred to a' iff $\overline{R}(a) \leq \underline{R}(a')$;
  2. a is preferred to a' iff $\underline{R}(a) \leq \underline{R}(a')$ (optimistic strategy);
  3. a is preferred to a' iff $\overline{R}(a) \leq \overline{R}(a')$ (pessimistic strategy).
- Strategy 1 yields only a partial preorder: a and a' are not comparable if $\overline{R}(a) > \underline{R}(a')$ and $\overline{R}(a') > \underline{R}(a)$.

Decision making
Special case

- Let $\Omega = \{\omega_1, \ldots, \omega_K\}$, $\mathcal{A} = \{a_1, \ldots, a_K\}$, where a_i is the act of selecting ω_i.
- Let

$$L(a_i, \omega_j) = \begin{cases} 0 & \text{if } i = j \text{ (the true state has been selected),} \\ 1 & \text{otherwise .} \end{cases}$$

- Then $\underline{R}(a_i) = 1 - pl(\omega_i)$ and $\overline{R}(a_i) = 1 - bel(\omega_i)$.
- The lower (resp., upper) risk is minimized by selecting the hypothesis with the largest plausibility (resp., degree of belief).

Decision making

Coming back to Ellsberg's paradox

We have $m(\{r\}) = 1/3$, $m(\{b, y\}) = 2/3$.

	r	b	y	$\underline{R}$	$\overline{R}$
A	-100	0	0	-100/3	-100/3
B	0	-100	0	-200/3	0
C	-100	0	-100	-100	-100/3
D	0	-100	-100	-200/3	-200/3

The observed behavior (preferring A to B and D to C) is explained by the pessimistic strategy.

Decision making

Other decision strategies

- How to find a compromise between the pessimistic strategy (minimizing the upper expected risk) and the optimistic one (minimizing the lower expected risk)?
- Two approaches:
  - Hurwicz criterion: a is preferred to a' iff $R_\rho(a) \leq R_\rho(a')$ with

    $$R_\rho(a) = (1-\rho)\underline{R}(a) + \rho\overline{R}(a).$$

    and $\rho \in [0, 1]$ is a pessimism index describing the attitude of the decision maker in the face of ambiguity.
  - Pignistic transformation (Transferable Belief Model).

Decision making
TBM approach

- The “Dutch book” argument: in order to avoid Dutch books (sequences of bets resulting in sure loss), we have to base our decisions on a probability distribution on Ω.
- The TBM postulates that uncertain reasoning and decision making are two fundamentally different operations occurring at two different levels:
  - Uncertain reasoning is performed at the credal level using the formalism of belief functions.
  - Decision making is performed at the pignistic level, after the m on Ω has been transformed into a probability measure.

Decision making
Pignistic transformation

- The pignistic transformation *Bet* transforms a normalized mass function m into a probability measure $P_m = Bet(m)$ as follows:

$$P_m(A) = \sum_{\emptyset \neq B \subseteq \Omega} m(B) \frac{|A \cap B|}{|B|}, \quad \forall A \subseteq \Omega.$$

- It can be shown that $bel(A) \leq P_m(A) \leq pl(A)$, hence $P_m \in \mathcal{P}(m)$. Consequently,

$$\underline{R}(a) \leq R_{P_m}(a) \leq \overline{R}(a), \quad \forall a \in \mathcal{A}.$$

Decision making
Example

- Let $m(\{John\}) = 0.48$, $m(\{John, Mary\}) = 0.12$, $m(\{Peter, John\}) = 0.32$, $m(\Omega) = 0.08$.
- We have

$$P_m(\{John\}) = 0.48 + \frac{0.12}{2} + \frac{0.32}{2} + \frac{0.08}{3} \approx 0.73,$$

$$P_m(\{Peter\}) = \frac{0.32}{2} + \frac{0.08}{3} \approx 0.19$$

$$P_m(\{Mary\}) = \frac{0.12}{2} + \frac{0.08}{3} \approx 0.09$$

Outline

Informational comparison of belief functions

- Let m_1 et 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 categorical mass functions.
  - m_A is more committed than m_B iff $A \subseteq B$.
- Extension to arbitrary mass functions?

Plausibility and commonality orderings

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

$$pl_1(A) \leq pl_2(A), \quad \forall A \subseteq \Omega.$$

- 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:

$$m_A \sqsubseteq_{pl} m_B \Leftrightarrow m_A \sqsubseteq_{q} m_B \Leftrightarrow A \subseteq B.$$

  - Greatest element: vacuous mass function m_Ω.

Strong (specialization) ordering

- 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_Ω;
  - $m_1 \sqsubseteq_s m_2 \Rightarrow m_1 \sqsubseteq_{pl} m_2$ and $m_1 \sqsubseteq_q m_2$.

Least Commitment Principle
Definition

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

Cautious rule

Motivations

- The standard rules $\textcircled{\cap}$, $\oplus$ 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.

Cautious rule

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 $q_{12} = q_1 \wedge q_2$.
- 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.
- This ordering will be based on the (conjunctive) canonical decomposition of belief functions.

Canonical decomposition

Simple and separable mass functions

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

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

with $A \subset \Omega$ and $w_A \in [0, 1]$.

- Notation: A^{w_A}.
- Property: $A^{w_1} \textcircled{\cap} A^{w_2} = A^{w_1 w_2}$.
- A mass function is separable if it can be written as the combination of simple mass functions:

$$
m = \textcircled{\cap}_{A \subset \Omega} A^{w(A)}
$$

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

Canonical decomposition

Subtracting evidence

- Let $m_{12} = m_1 \textcircled{\cap} m_2$. We have $q_{12} = q_1 \cdot q_2$.
- Assume we no longer trust m_2 and we wish to subtract it from m_{12}.
- If m_2 is non dogmatic (i.e. $m_2(\Omega) > 0$ or, equivalently, $q_2(A) > 0, \forall A$), m_1 can be retrieved as

$$q_1 = q_{12}/q_2.$$

- We note $m_1 = m_{12} \textcircled{\not\cap} m_2$.
- Remark: $m_1 \textcircled{\not\cap} m_2$ may not be a valid mass function!

Canonical decomposition

Theorem (Smets, 1995)

Any non dogmatic mass function ($m(\Omega) > 0$) can be canonically decomposed as:

$$m = \left(\bigcirc\!\!\!\!\cap_{A \subset \Omega} A^{w_C(A)} \right) \oslash \left(\bigcirc\!\!\!\!\cap_{A \subset \Omega} A^{w_D(A)} \right)$$

with $w_C(A) \in (0, 1]$, $w_D(A) \in (0, 1]$ *and* $\max(w_C(A), w_D(A)) = 1$ *for all* $A \subset \Omega$.

- Let $w = w_C / w_D$.
- Function $w : 2^\Omega \setminus \Omega \rightarrow \mathbb{R}_+^*$ is called the (conjunctive) weight function.
- It is a new equivalent representation of a non dogmatic mass function (together with *bel*, *pl*, *q*, *b*).

Properties of *w*

- Function *w* is directly available when *m* is built by accumulating simple mass functions (common situation).
- Calculation of *w* from *q*:

$$\ln w(A) = -\sum_{B \supseteq A} (-1)^{|B|-|A|} \ln q(B), \quad \forall A \subset \Omega.$$

- Conversely,

$$\ln q(A) = -\sum_{\Omega \supset B \not\supseteq A} \ln w(B), \quad \forall A \subseteq \Omega$$

- TBM conjunctive rule:

$$w_1 \cap\!\!\!\!\bigcirc\, w_2 = w_1 \cdot w_2.$$

w-ordering

- Let m_1 and m_2 be two non dogmatic mass functions. We say that m_1 is w-more committed than m_2 (denoted as $m_1 \sqsubseteq_w m_2$) if $w_1 \leq w_2$.
- Interpretation: $m_1 = m_2 \textcircled{\cap} m$ with m separable.
- Properties:
  - $m_1 \sqsubseteq_w m_2 \Rightarrow m_1 \sqsubseteq_s m_2 \Rightarrow \begin{cases} m_1 \sqsubseteq_{pl} m_2 \\ m_1 \sqsubseteq_q m_2, \end{cases}$
  - m_Ω is the only maximal element of $\sqsubseteq_w$:

$$m_\Omega \sqsubseteq_w m \Rightarrow m = m_\Omega.$$

Cautious rule
Definition

Theorem

Let m_1 and m_2 be two nondogmatic BBAs. 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 the following weight function:

$$w_{1 \textcircled{\wedge} 2}(A) = w_1(A) \wedge w_2(A), \quad \forall A \subset \Omega.$$

Definition (cautious conjunctive rule)

$$m_1 \textcircled{\wedge} m_2 = \textcircled{\cap}_{A \subset \Omega} A^{w_1(A) \wedge w_2(A)}.$$

Cautious rule
Definition

Theorem

Let m_1 and m_2 be two nondogmatic BBAs. 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 the following weight function:

$$w_{1 \owedge 2}(A) = w_1(A) \wedge w_2(A), \quad \forall A \subset \Omega.$$

Definition (cautious conjunctive rule)

$$m_1 \owedge m_2 = \bigcirc\!\!\!\!\cap_{A \subset \Omega} A^{w_1(A) \wedge w_2(A)}.$$

Cautious rule
Computation

Cautious rule computation

m-space		w-space
m_1	$\longrightarrow$	w_1
m_2	$\longrightarrow$	w_2
$m_1 \owedge m_2$	$\longleftarrow$	$w_1 \wedge w_2$

Cautious rule
Properties

- Commutative, associative
- Idempotent : $\forall m,\ m \textcircled{\scriptsize$\wedge$} m = m$
- Distributivity of $\textcircled{\scriptsize$\cap$}$ with respect to $\textcircled{\scriptsize$\wedge$}$:

$$(m_1 \textcircled{\scriptsize$\cap$} m_2) \textcircled{\scriptsize$\wedge$} (m_1 \textcircled{\scriptsize$\cap$} m_3) = m_1 \textcircled{\scriptsize$\cap$} (m_2 \textcircled{\scriptsize$\wedge$} m_3), \forall m_1, m_2, m_3.$$

  The same item of evidence m_1 is not counted twice!
- No neutral element, but $m_\Omega \textcircled{\scriptsize$\wedge$} m = m$ iff m is separable.

Related rules

- Normalized cautious rule:

$$(m_1 \owedge^* m_2)(A) = \begin{cases} \frac{(m_1 \owedge m_2)(A)}{1-(m_1 \owedge m_2)(\emptyset)} & \text{if } A \neq \emptyset \\ 0 & \text{if } A = \emptyset. \end{cases}$$

- Bold disjunctive rule:

$$m_1 \ovee m_2 = \overline{\overline{m}_1 \owedge \overline{m}_2}.$$

- Both $\owedge^*$ and $\ovee$ are commutative, associative and idempotent.

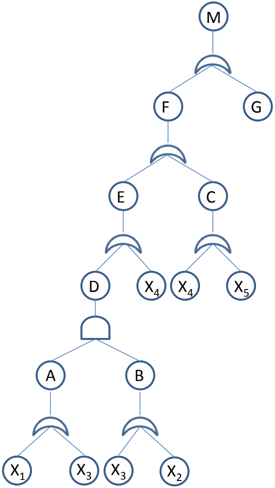
Global picture

- Six basic rules:

Sources		independent	dependent
All reliable	open world	$\textcircled{\cap}$	$\textcircled{\wedge}$
	closed world	$\oplus$	$\textcircled{\wedge}^*$
At least one reliable		$\textcircled{\cup}$	$\textcircled{\vee}$

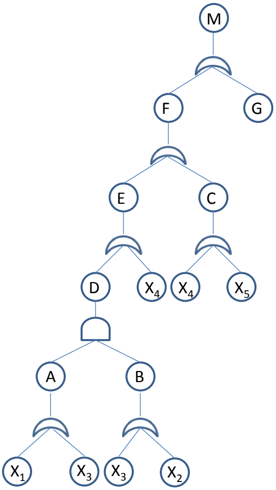
Outline

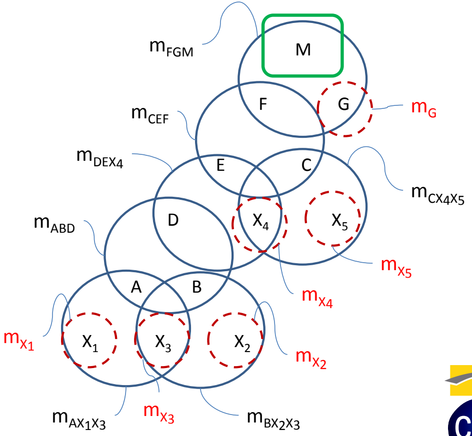
Multidimensional belief functions
Motivations



- In many applications, we need to express uncertain information about several variables taking values in different domains.
- Example: fault tree (logical relations between Boolean variables and probabilistic or evidential information about elementary events).

Fault tree example
(Dempster & Kong, 1988)





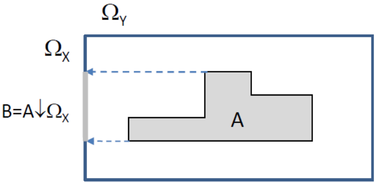
Hypergraph

Multidimensional belief functions

Marginalization, vacuous extension

- Let X and Y be two variables defined on frames Ω_X and Ω_Y.
- Let $\Omega_{XY} = \Omega_X \times \Omega_Y$ be the product frame.
- A mass function $m^{\Omega_{XY}}$ on Ω_{XY} can be seen as an uncertain relation between variables X and Y.
- Two basic operations on product frames:
  1. Express a joint mass function $m^{\Omega_{XY}}$ in the coarser frame Ω_X or Ω_Y (marginalization);
  2. Express a marginal mass function m^{Ω_X} on Ω_X in the finer frame Ω_{XY} (vacuous extension).

Marginalization



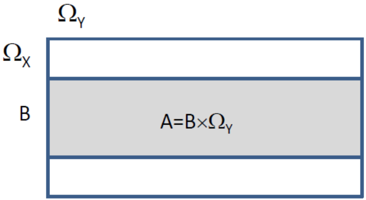
- Problem: express $m^{\Omega_{XY}}$ in Ω_X.
- Solution: transfer each mass $m^{\Omega_{XY}}(A)$ to the projection of A on Ω_X:

- Marginal mass function

$$m^{\Omega_{XY}\downarrow\Omega_X}(B) = \sum_{\{A\subseteq\Omega_{XY}, A\downarrow\Omega_X=B\}} m^{\Omega_{XY}}(A), \quad \forall B \subseteq \Omega_X.$$

- Generalizes both set projection and probabilistic marginalization.

Vacuous extension



- Problem: express m^{Ω_X} in Ω_{XY}.
- Solution: transfer each mass $m^{\Omega_X}(B)$ to the cylindrical extension of B: $B \times \Omega_Y$.

- Vacuous extension:

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

Operations in product frames

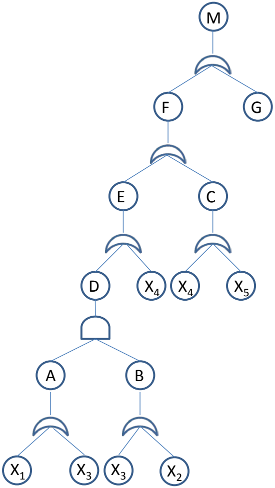
Application to approximate reasoning

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

$$m^{\Omega_Y} = \left(m^{\Omega_X \uparrow \Omega_{XY}} \cap\!\!\!\bigcirc\, m^{\Omega_{XY}} \right)^{\downarrow \Omega_Y}.$$

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

Fault tree example



Cause	$m(\{1\})$	$m(\{0\})$	$m(\{0,1\})$
X_1	0.05	0.90	0.05
X_2	0.05	0.90	0.05
X_3	0.005	0.99	0.005
X_4	0.01	0.985	0.005
X_5	0.002	0.995	0.003
G	0.001	0.99	0.009
M	0.02	0.951	0.029
F	0.019	0.961	0.02

Summary

- The theory of belief function: a very general formalism for representing imprecision and uncertainty that extends both probabilistic and set-theoretic frameworks:
  - Belief functions can be seen both as generalized sets and as generalized probability measures;
  - Reasoning mechanisms extend both set-theoretic notions (intersection, union, cylindrical extension, inclusion relations, etc.) and probabilistic notions (conditioning, marginalization, Bayes theorem, stochastic ordering, etc.).
- The theory of belief function can also be seen as more general than Possibility theory (possibility measures are particular plausibility functions).

References I

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.

References II

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

T. Denœux.

Analysis of evidence-theoretic decision rules for pattern classification.

Pattern Recognition, 30(7):1095-1107, 1997.

T. Denœux.

Conjunctive and Disjunctive Combination of Belief Functions Induced by Non Distinct Bodies of Evidence.

Artificial Intelligence, Vol. 172, pages 234-264, 2008.