Representation of evidence

Workshop on belief functions

Thierry Denœux

April, 2016

This chapter

- In this chapter, we define some of the main concepts of Dempster-Shafer theory in the finite case.
- These notions are sufficient to cope with a large number of applications.
- The extension to infinite spaces involves some mathematical intricacies and is technically more difficult, except in some simple (and practically important) cases; it will be addressed later.

Overview

Mass function
- Definitions
- Semantics

Belief and plausibility functions
- Definitions
- Properties
- Vector representation

Special cases and related theories
- Consonant mass functions
- Relation with imprecise probabilities

Overview

Mass function
- Definitions
- Semantics

Belief and plausibility functions
- Definitions
- Properties
- Vector representation

Special cases and related theories
- Consonant mass functions
- Relation with imprecise probabilities

Frame of discernment

- Let Ω be a finite set of possible answers to some question Q, one and only one of which is true.
- The true answer will be denoted by $\boldsymbol{\omega}$, and an arbitrary element of Ω by ω.
- Shafer (1976) calls such a space a frame of discernment, to emphasize the fact that it is not a set of "states of nature" objectively given, but a subjective construction based on our state of knowledge.
- For instance, if Q relates to a person's state of health, Ω might contain only the diseases known at a certain time. This set could be later refined or extended if new knowledge became available.

Mass function

- A piece of evidence about Q will be represented by a mass function, defined as a mapping m from the power set 2^Ω to the interval $[0,1]$, such that $m(\emptyset)=0$ and

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

- Each number $m(A)$ represents the probability that the evidence supports exactly the proposition $\omega\in A$, and no more specific proposition.
- Any subset A of Ω such that $m(A)>0$ is called a focal set of m. The union of the focal sets of a mass function is called its core.

Special cases

1. If m has only one focal set, it is said to be logical. Logical mass functions are in one-to-one correspondence with subsets of Ω: consequently, general mass functions can be viewed as generalized sets. A particular logical mass function plays a special role in the theory; it is the vacuous mass function $m_?$ defined by $m_?(\Omega) = 1$; such a mass function corresponds to a totally uninformative piece of evidence.
2. If all focal sets are singletons (i.e., sets of cardinality one), m is said to be Bayesian. To each Bayesian mass function can be associated a probability distribution $p : \Omega \to [0, 1]$ such that $p(\omega) = m(\{\omega\})$ for all $\omega \in \Omega$.

Example

- Consider the mass on functions on $\Omega = \{a, b, c\}$ shown in below. Mass function m_1 is Bayesian, m_2 is logical, $m_?$ is vacuous, and m_3 has no special form.

A	$\emptyset$	$\{a\}$	$\{b\}$	$\{a, b\}$	$\{c\}$	$\{a, c\}$	$\{b, c\}$	$\{a, b, c\}$
$m_1(A)$	0	0.2	0.5	0	0.3	0	0	0
$m_2(A)$	0	0	0	1	0	0	0	0
$m_?(A)$	0	0	0	0	0	0	0	1
$m_3(A)$	0	0.1	0.05	0.2	0.15	0.3	0.1	0.1

Overview

Mass function
Definitions
Semantics

Belief and plausibility functions
Definitions
Properties
Vector representation

Special cases and related theories
Consonant mass functions
Relation with imprecise probabilities

The murder example

- A murder has been committed and there are three suspects: Peter, John and Mary.
- The question Q of interest is the identity of the murderer and the frame of discernment is $\Omega = \{\mathrm{Peter}, \mathrm{John}, \mathrm{Mary}\}$.
- The piece of evidence under study is a testimony: a witness saw the murderer. However, this witness is short-sighted and he can only report that he saw a man.
- Unfortunately, this testimony is also not fully reliable, because we know that the witness is drunk 20 % of the time.
- How can such a piece of evidence be encoded in the language of mass functions?

Formalization

- We can see here that what the testimony tells us about Q depends on the answer to another question Q': Was the witness drunk at the time of the murder?
- If he was not drunk, we know that the murderer is Peter or John. Otherwise, we know nothing.
- Since there is 80% chance that the former hypothesis holds, we may assign a 0.8 mass to the set $\{\text{Peter}, \text{John}\}$, and 0.2 to Ω:

$$m(\{\text{Peter}, \text{John}\}) = 0.8, \quad m(\Omega) = 0.2$$

A message with random meaning

- In the above example, we receive a message (a testimony) about Q, whose meaning depends on the answer to a related question Q' for which we have a chance model (a probability distribution).
- We can compare our evidence to a canonical example where we know that the outcomes of a random experiment are s_1 and s_2 with corresponding chances $p_1 = 0.8$ and $p_2 = 0.2$, and the message can only be interpreted with knowledge of the outcome.
- If the outcome is s_1, then the meaning is $\omega \in \{\mathrm{Peter}, \mathrm{John}\}$, otherwise the meaning is $\omega \in \Omega$, i.e., the message is totally uninformative.

Canonical examples

- We have seen that, in the constructive approach, probability judgements can be made by comparing the available evidence to some canonical example involving a chance setup.
- In the Bayesian theory, we compare our evidence to a situation where the truth is governed by chance (e.g., by thinking of the murderer as having been selected at random).
- In the belief function approach, the canonical example describes a situation where the meaning of the evidence is governed by chance.
- Two scenarios are specially useful to construct canonical examples for mass functions.

The unreliable machine

- The first scenario involves a machine that has two modes of operation, normal and faulty. We know that in the normal mode it broadcasts true messages, but we are completely unable to predict what it does in the faulty mode.
- We further assume that the operating mode of the machine is random and there a chance p that it is in the normal mode.
- It is then natural to say that a message $\omega \in A$ produced by the machine has a chance p of meaning what it says and a chance $1 - p$ of meaning nothing.
- This leads to the mass function $m(A) = p$ and $m(\Omega) = 1 - p$.
- Such a mass function, with two focal sets including Ω, is called a simple mass function.

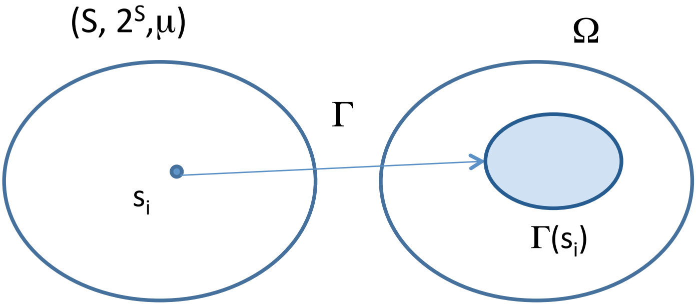
Random code

- The above story is not general enough to cover all kinds of evidence.
- More sophisticated scenario: a source holds some true information of the form $\omega \in A^*$ for some $A^* \subseteq \Omega$.
- It sends us this information as an encoded message using a code chosen at random from a set of codes $S = \{s_1, \ldots, s_r\}$, according to some known probability measure μ.
- We know the set of codes as well as the chances of each code to be selected. If we decode the message using code s, we get a decoded message of the form $\omega \in \Gamma(s)$ for some subset $\Gamma(s)$ of Ω. Then,

$$m(A) = \mu(\{s \in S | \Gamma(s) = A\}) \quad (2)$$

is the chance that the original message was "$\omega \in A$", i.e., the probability of knowing that $\omega \in A$, and nothing more.

The random code – continued



Random set

- In the above framework, the mapping $\Gamma : S \rightarrow 2^{\Omega} \setminus \{\emptyset\}$ is called a multi-valued mapping and the 4-tuple $(S, 2^S, \mu, \Gamma)$ is called a source.
- We can observe that a source corresponds formally to a random set.
- However, the term "random set" may be misleading here, because we are not interested in situations where a set is selected at random (such as, e.g., drawing a handful of marbles from a bag).
- Here, the true answer to the question of interest is a single element of Ω and it is not assumed to have been selected at random. Instead, chances are introduced when comparing our evidence to a situation where the meaning of a message depends on the result of a random experiment.

Relation between random sets and mass functions

- It is clear that a source $(S, 2^S, \mu, \Gamma)$ always induces a mass function.
- Conversely, any mass function can be seen as generated by a source. For instance, if $A_1, \ldots, A_n$ are the focal sets of a mass function m, we may set $S = \{1, \ldots, n\}$ and $\mu(\{i\}) = m(A_i)$ for $1 \leq i \leq n$.
- However, as we will see later, the concept of a source is more general than that of mass function, because a source can be used in the infinite case to generate a belief function even when a mass function does not exist.

Overview

Mass function
- Definitions
- Semantics

Belief and plausibility functions
- Definitions
- Properties
- Vector representation

Special cases and related theories
- Consonant mass functions
- Relation with imprecise probabilities

Overview

Mass function
Definitions
Semantics

Belief and plausibility functions
Definitions
Properties
Vector representation

Special cases and related theories
Consonant mass functions
Relation with imprecise probabilities

Overview

Mass function
- Definitions
- Semantics

Belief and plausibility functions
- Definitions
- Properties
- Vector representation

Special cases and related theories
- Consonant mass functions
- Relation with imprecise probabilities

Overview

Mass function
- Definitions
- Semantics

Belief and plausibility functions
- Definitions
- Properties
- Vector representation

Special cases and related theories
- Consonant mass functions
- Relation with imprecise probabilities

Overview

Mass function
- Definitions
- Semantics

Belief and plausibility functions
- Definitions
- Properties
- Vector representation

Special cases and related theories
- Consonant mass functions
- Relation with imprecise probabilities

Overview

Mass function
- Definitions
- Semantics

Belief and plausibility functions
- Definitions
- Properties
- Vector representation

Special cases and related theories
- Consonant mass functions
- Relation with imprecise probabilities

Overview

Mass function
- Definitions
- Semantics

Belief and plausibility functions
- Definitions
- Properties
- Vector representation

Special cases and related theories
- Consonant mass functions
- Relation with imprecise probabilities