International Journal of Approximate Reasoning 71 (2016) 15–33

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Interpreting evidential distances by connecting them to partial orders: Application to belief function approximation

John Klein [a,*], Sebastien Destercke [b], Olivier Colot [a]

[a] Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL, Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France

[b] Technologic University of Compiegne, UMR 7253 - Heudiasyc, Centre de Recherche de Royallieu, Compiègne, France

ARTICLE INFO

Article history:
Received 31 July 2015
Received in revised form 29 December 2015
Accepted 4 January 2016
Available online 14 January 2016

Keywords:
Belief functions
Distance
Metric
Partial order
Semantics
Convex optimization

ABSTRACT

Distances between mass functions are instrumental tools in evidence theory, yet it is not always clear in which situation a particular distance should be used. Indeed, while the mathematical properties of distances have been well studied, how to interpret them is still a largely open issue. As a step towards answering this question, we propose to interpret distances by looking at their compatibility with partial orders. We formalize this compatibility through some mathematical properties thereby allowing to combine the advantages of both partial orders (clear semantics) and distances (richer structure and access to numerical tools). We explore in particular the case of informational partial orders, and how distances compatible with such orders can be used to approximate initial belief functions by simpler ones through the use of convex optimization. We finish by discussing some perspectives of the current work.

1. Introduction

The theory of belief functions is a flexible framework to model uncertainty in the presence of imprecision. This framework mixes set and probabilistic representations. It was initially proposed to model imprecise statistical observations [1], and this initial work was then extended [2] to include subjective and non-statistical uncertainty (*e.g.*, when a variable has a fixed, yet ill-known value). This latter view was then pursued by Smets [3], who dissociated belief functions from any probabilistic interpretation. They include many other representations proposed in the literature, such as sets, probability measures or possibility measures. In this paper, we will use the term evidence theory as a generic term for frameworks relying on belief functions.

Among the tools developed to work with belief functions, distances have recently received a growing attention. They have been proposed as tools to achieve various tasks: measuring conflict [4,5], measuring dependencies [6], learning models from data [7,8], or belief function approximation [9–15]. Jousselme and Maupin [16] surveyed evidential distances and classified them with respect to their mathematical properties and to show some correlated behaviors among them. Following Jousselme and Maupin's analysis, Loudahi et al. [17,18] formalized some properties with intuitive interpretations in the framework of evidence theory: compatibility of distances with some combination rules and with the set-inclusion. Despite these efforts, providing evidential distances with clear interpretations remains an open problem.

* Corresponding author.
E-mail addresses: john.klein@univ-lille1.fr (J. Klein), sebastien.destercke@hds.utc.fr (S. Destercke), olivier.colot@univ-lille1.fr (O. Colot).

http://dx.doi.org/10.1016/j.ijar.2016.01.001
0888-613X/

In this paper, we start in Section 3 by proposing a new answer to this problem: we interpret a distance by its compatibility or incompatibility with some partial order possessing a clear semantics. More precisely, we say that a distance is compatible with a partial order if, given a set of three belief functions forming a chain within this partial order, the distance between the minimal and the maximal belief functions should be greater than the distance between any other pair. We prove that this compatibility property holds for several infinite families of evidential distances when informational partial orders are considered, thereby formally bridging two important notions in the theory of belief functions.

We then study in Section 4 the problem of belief function approximation, in which partial orders related to informative content play a specific role. The combination of both distances and orders is very interesting in this problem, as the partial orders allow us to select those distances fitted to the approximation problem, while the use of specific distances (within the selected subset) allows us to take advantage of their mathematical properties to find unique solutions (when partial orders only offer sets of incomparable solutions). Indeed, these approximation problems are convex and can be easily solved using quadratic programming for instance.

Finally, Section 5 provides some discussion about the presented results and ideas. We briefly review the basic concepts of the theory of belief functions in Section 2.

2. Basics of evidence theory

This section reminds the notions of evidence theory used in this paper. More details about the various tools used in evidence theory can be found in [19], for example. After providing this necessary background, we will give a more detailed presentation of partial orders and distances used in evidence theory in section 3.

Let $\Omega = \{\omega_1, \ldots, \omega_n\}$ be a finite space over which a given ill-known variable θ takes its values. In evidence theory, a **mass function** $m : 2^\Omega \to [0, 1]$ defined over the power set of Ω represents our uncertainty about the value of θ. The mass $m(A)$ can be given several interpretations depending on the chosen interpretation:

- amount of evidence given to the fact that A contains the true value [2,3],
- or the frequency of the imprecise observation A [1].

Mass functions sum to one, *i.e.*, $\sum_{A \in 2^\Omega} m(A) = 1$. A set A receiving a positive mass $m(A) > 0$ is called a **focal element**. We will denote by $|A|$ the **cardinality** of a set A. In particular, $|\Omega| = n$ and $|2^\Omega| = N = 2^n$.

Several alternative set-functions can then be defined to represent the same information as the one encoded in a mass function. The main ones are the plausibility, belief, implacability and commonality functions. The **plausibility** function $pl : 2^\Omega \to [0, 1]$ is defined as

$$pl(A) = \sum_{E \cap A \neq \emptyset} m(E) \tag{1}$$

and evaluates how much event A (being true) is consistent with the current evidence. The **belief** $bel : 2^\Omega \to [0, 1]$ and **implicability** $b : 2^\Omega \to [0, 1]$ functions are defined as

$$bel(A) = \sum_{E \subseteq A, E \neq \emptyset} m(E), \tag{2}$$

$$b(A) = \sum_{E \subseteq A} m(E) = bel(A) + m(\emptyset). \tag{3}$$

Both evaluate how much event A (being true) is implied by the current evidence, with the implicability assuming that $\emptyset$ can imply anything, and the belief discarding $\emptyset$ from valid hypotheses. We have $pl(A) = 1 - b(A^c)$, A^c being the complement of A. Also, we always have $bel(A) \leq pl(A)$.

When $m(\emptyset) = 0$, we have $bel = b$, and the couple belief/plausibility can be interpreted as bounds of an ill-known probability, in the sense that they induce a non-empty set

$$\mathcal{P}(m) = \{P | bel(A) \leq P(A) \leq pl(A), \forall A \subseteq \Omega\}$$

where P are probability measures over the probability space $\left(\Omega, 2^\Omega, P\right)$. Also, in this case, the value $pl(A) - bel(A)$ measures the imprecision of the information contained in m. When $pl(A) = bel(A)$ for all A, the set $\mathcal{P}(m)$ contains only one probability measure. This is a fully precise situation and $m(E) > 0$ only if $|E| = 1$. When $pl(A) - bel(A) = 1$ for all A, $\mathcal{P}(m)$ is the set of all probability measures. This is a maximally imprecise situation and $m(\Omega) = 1$.

Requiring $m(\emptyset) = 0$ can therefore be seen as a consistency constraint, while allowing for $m(\emptyset) \neq 0$ means that m is allowed to encode some self-contradiction. A positive mass for $\emptyset$ features an underlying conflict between the pieces of evidence encoded by m. This conflict can have various origins: (i) untruthfulness of some of these pieces of evidence, (ii) the fact that the true value is not in Ω (open world assumption). We will call **normalized** those masses such that $m(\emptyset) = 0$.