Information Sciences 639 (2023) 119027

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

A numerical comparative study of uncertainty measures in the Dempster–Shafer evidence theory

Michele Urbani *, Gaia Gasparini, Matteo Brunelli

Department of Industrial Engineering, University of Trento, via Sommarive 9, Trento, 38123, Italy

ARTICLE INFO

Keywords:
Evidence theory
Uncertainty measure
Entropy
Similarity

ABSTRACT

We consider a wide range of measures of uncertainty that have been proposed within the Dempster–Shafer evidence theory. All these measures aim to quantify the uncertainty associated with a given basic probability assignment. As a preliminary step, we offer a study of the literature, which shows a recent resurgence of interest in the quantification of uncertainty in the evidence theory. Then, we compare a number of uncertainty measures by means of numerical simulations and analyze their similarities and differences using rank correlation coefficients, hierarchical clustering, and centrality analysis. The results show that uncertainty measures with similar formulations do not necessarily have similar numerical properties, and some original results are obtained. In particular, we demonstrate that numerical studies on uncertainty measures are necessary to obtain more insight and to enhance the interpretability of the values returned by the measures.

> Any expression of probability is a claim to a knowledge of the underlying issue which, by the nature of the problem, the speaker cannot have. In these circumstances, it may often make sense to describe the degree of uncertainty in non-probabilistic ways.
>
> [John Kay and Mervyn King]

1. Introduction

According to some widely accepted taxonomies [1, p. 53], uncertain phenomena can either be classified as aleatoric or epistemic, depending on the nature of the underlying uncertain event. Aleatoric uncertainty refers to events with inherent randomness, while epistemic uncertainty refers to events for which there is a lack of knowledge. An example of epistemic uncertainty is in the exact time at which you first saw this paper: there is no randomness, but only a lack of knowledge of the exact timing of the event. Whereas probability has been the main paradigm for dealing with aleatoric uncertainty, other theories have been proposed to represent uncertain phenomena of the epistemic type. Among these theories one can find fuzzy sets theory—and its interpretation in terms of possibility theory [2]—and evidence theory, also called Dempster–Shafer theory, proposed by Dempster [3] as a formalization of lower and upper probabilities and then developed by Shafer [4]. In fact, evidence theory can be seen as a generalization of both probability and possibility theories [1] and great effort has been made to make it operational and applicable in decision analysis.

* Corresponding author.
E-mail addresses: michele.urbani@unitn.it (M. Urbani), gaia.gasparini@unitn.it (G. Gasparini), matteo.brunelli@unitn.it (M. Brunelli).

https://doi.org/10.1016/j.ins.2023.119027
Received 11 October 2022; Received in revised form 17 April 2023; Accepted 24 April 2023

Available online 28 April 2023
0020-0255/

If we turn our attention to the special case of probability theory, the uncertainty of probability distributions has generated an important stream of literature connected with the relationship between uncertainty and information. Clearly, the more uncertain a probability distribution is, the more information one can gain by learning the realization of the event. This insight is the basis for the interpretation of Shannon entropy [5], which is still the most widely known measure of uncertainty/information.

If we consider the more general framework of evidence theory, the last 40 years—and especially the last 5—have witnessed the proposal of a wide number of uncertainty measures. However, it remains difficult to pin down a single measure that can be considered superior to all the others and, therefore, most of the proposed measures are still competitive when it comes to choosing among them in real-world applications. While some studies have attempted to find an outstanding measure that satisfies a number of desirable theoretical properties, they have been inconclusive and therefore, at present, none of the proposed measures should be considered superior. Other studies have compared measures with respect to specific case studies, so that some of their properties, such as monotonicity, can be verified empirically. However, such studies cannot offer a broad overview of the similarities and differences between uncertainty measures.

Let us note that the matter of the fair quantification of uncertainty within evidence theory is far from being a mere theoretical exercise, as the choice of uncertainty measure can have repercussions for real-world applications. Furthermore, the use of uncertainty measures has gone beyond the mere quantification of uncertainty, as shown by, for instance, their recent application in the evaluation of approximations of conjoint belief distributions [6].

The scope of this paper is to offer a presentation of uncertainty measures, in conjunction with a recap of their mathematical properties and a brief historical perspective of their development. However, since studies of the properties of such measures have already been presented in the literature, the main goal of this paper is to present the results of a numerical study that highlights possible similarities, differences, and mutual support between uncertainty measures in the evidence theory. So far, numerical comparative studies have been performed in some specific instances to show that some proposed measures, unlike some old ones, do satisfy some desirable properties.

For example, Wen et al. [7] considered only 9 representative mass assignments to compare 4 uncertainty measures and Pan et al. [8] verified the satisfaction of the property of monotonicity by applying 5 uncertainty measures to a family of 14 mass assignments for which, *ceteris paribus*, the cardinality of an element of the frame of discernment is increasing. A similar analysis of the monotonicity of 9 measures was offered by Wang et al. [9].

We conclude that the scope of these analyses remains limited and they cannot be considered full-fledged comparative studies of uncertainty measures. Hopefully, in addition to the existing literature, a more comprehensive numerical study will help choose among the many uncertainty measures.

The paper is organized as follows. Section 2 presents the basics of evidence theory and a brief introduction to the concept of entropy in probability theory. Section 3 is an exposition of a number of uncertainty measures that can be used in the evidence theory, with an eye on their formal properties and their historical development. Section 4 presents the methodological tools used in the analysis, and the results. Finally, in Section 5 we draw some conclusions.

2. Evidence theory and uncertainty

Let $X = \{x_1, x_2, \dots, x_n\}$ be a non-empty set containing n exhaustive and mutually exclusive elements; this set is called the *frame of discernment* (FOD). The power set of X is denoted as 2^X. A *basic probability assignment* (BPA), or mass assignment, is a function

$$m : 2^X \to [0,1] \tag{1}$$

that satisfies

$$\sum_{A \in 2^X} m(A) = 1 \text{ and } m(\emptyset) = 0, \tag{2}$$

where $A \subseteq X$ is called a *focal element* if $m(A) > 0$. The set of all the focal elements is denoted as $\mathcal{F} = \{A \subseteq X | m(A) > 0\}$. The set of focal elements together with their masses is called the *body of evidence* (BOE). Indeed, the BPA can be seen as a generalization of the common assignments of probabilities on singletons, which is now done on all the elements of 2^X. In fact, the BPA assigns a value between 0 and 1 to every element of the power set, so that $m(A)$ can be interpreted as the normalized amount of evidence that a given event belongs to A.

Two relevant measures can be associated with a BPA. First, given a BPA, the *belief* of A is defined as the sum of all the masses of the subsets of A (and of A itself),

$$\mathrm{Bel}(A) = \sum_{B \subseteq A} m(B), \quad \forall A \in 2^X, \tag{3}$$

and is a superadditive measure. This measure can be interpreted as the quantification of the level of confidence in knowing that the event is within A: through the sum of the masses, the belief measures the strength of the evidence in favor of the presence of the element in A or in a subset thereof. Second, the *plausibility* of A is the sum of the masses of all the subsets that intersect A,

$$\mathrm{Pl}(A) = \sum_{B \cap A \neq \emptyset} m(B), \quad \forall A \in 2^X. \tag{4}$$

Unlike belief, plausibility is a subadditive measure of the strength of the evidence that does not contradict the presence of the event in A.

The two measures $\mathrm{Bel}(A)$ and $\mathrm{Pl}(A)$ represent, respectively, the lower and the upper bounds of the probability of A, that is, $\exists\,\mathrm{Pr}$ such that

$$\mathrm{Bel}(A) \leq \mathrm{Pr}(A) \leq \mathrm{Pl}(A), \quad \forall A \subseteq X. \tag{5}$$

The set of probability functions compatible with Eq. (5) is a *credal set*, that is, a convex set of probability distributions. Hereafter, we denote by $\mathcal{P}(m)$ the credal set induced by m by means of the condition in Eq. (5). Note that (i) this also highlights the connection between evidence theory and the theory of imprecise probabilities and (ii) when the mass assignment involves only $\{x_i\}$, we have $\mathrm{Bel}(A) = \mathrm{Pr}(A) = \mathrm{Pl}(A)$, $\forall A \subseteq X$, which casts probability as a special case of the evidence theory.

Another important concept is Shannon entropy. Within the framework of probability theory, Shannon [5] introduced the concept of entropy as a measure of the amount of information associated with knowing the outcome of an uncertain event. Formally, Shannon entropy is defined as

$$\mathsf{H}_s(p) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i), \tag{6}$$

where $p(x_i)$ is the probability of $x_i \in X$. Indeed, it attains its maximum when the probability is uniform and the uncertainty is maximum. Similarly, in set theory, set uncertainty depends on the cardinality of the set and is usually measured by means of the Hartley entropy, according to which the uncertainty of a set A is $\log_2 |A|$.

According to Ayyub and Klir [1, Ch. 4], a sufficiently general mathematical representation of uncertainty, as the evidence theory, should be able to accommodate two types of uncertainty: *conflict* and *non-specificity*. In fact, in the evidence theory, different types of uncertainties coexist within the BPA and therefore it is now accepted that, when we want to measure the (total) uncertainty of the assignment m, we ought to consider both facets of uncertainty.

The situation is made even more complex by the fact that conflict and non-specificity have multiple interpretations. For example, non-specificity, which should measure the divergence of a BPA from being Bayesian (i.e., from having all the evidence assigned to singletons), has at least three interpretations [10]: (i) one in light of Hartley entropy, (ii) one related to the width of the so-called belief intervals $[\mathrm{Bel}(A), \mathrm{Pl}(A)]$, and (iii) one connected to the information volume of the credal set $\mathcal{P}(m)$. Further complexities emerge when one considers conflict. In fact, axiomatic studies such as the one by Bronevich and Lepskiy [11] explain that there are multiple ways to quantify the conflict in m that “do not coincide with the Shannon entropy on probability measures.” This richness of interpretations has provided fertile ground for the development of uncertainty measures.

3. Uncertainty measures in evidence theory

Several measures of uncertainty for basic probability assignments have been proposed in the evidence theory. Those studied in this paper are listed and commented on here, using the names of their proponents. Unless otherwise stated, in the following definitions, we will assume A and B to be subsets of the FOD, that is, $A, B \subseteq X$, and we use $|A|$ to denote the cardinality of set A. In addition, for simplicity, functions of singletons are written without curly brackets, for example, $\mathrm{Pl}(\{x_1\}) = \mathrm{Pl}(x_1)$. We shall see that the oldest measures of uncertainty are, at times, simplistic and cannot fully capture all facets of uncertainty.

Höhle

According to Klir and Ramer [12], the investigation of entropy in the evidence theory started in the 1980s, with the first measure, proposed by Höhle [13], defined as

$$\mathsf{H}_O(m) = -\sum_{A \in \mathcal{F}} m(A) \log_2 \mathrm{Bel}(A). \tag{7}$$

This attains its minimum if and only if the BPA is deterministic, that is, there exists $A \subseteq X$ such that $m(A) = 1$. In the literature, this measure is often called *confusion*.

Yager

Similarly to Höhle, Yager [14] proposed a measure of uncertainty starting again from the definition of Shannon entropy. Like the preceding one, this one by Yager,

$$\mathsf{H}_y(m) = -\sum_{A \in \mathcal{F}} m(A) \log_2 \mathrm{Pl}(A), \tag{8}$$

aimed at capturing conflict. In the literature, this measure is often called *dissonance*.

Smets

Smets [15] proposed a conflict measure with the use of the commonality function. The entropy, in this case, measures the information content of a BOE, and its formulation is

$$\mathsf{H}_t(m) = -\sum_{A\in\mathcal{F}} c(A)\log_2 Q(A), \tag{9}$$

where $Q(A) = \sum_{B|A\subseteq B} m(B)$ is the so-called commonality function already defined by Shafer [4]. Furthermore, we set $c(A) = 1$, which is, according to Smets [15, p. 37], the "most natural choice."

Nguyen

The measure proposed by Nguyen [16], defined as

$$\mathsf{H}_n(m) = -\sum_{A\in\mathcal{F}} m(A)\log_2 m(A), \tag{10}$$

is the most straightforward extension of Shannon entropy.

Dubois and Prade

It can be observed that all the previously recalled uncertainty measures do not consider the cardinality of sets A. To give a concrete example, if there is $A \subseteq X$ with $m(A) = 1$, they all reach their minimum, regardless of whether A is a singleton or the FOD X itself. Nonetheless, it is intuitive to consider the former case as less "uncertain" than the latter. Dubois and Prade [17] introduced the first measure of non-specificity, defined as a weighted average of Hartley entropies with respect to each focal element $A \in \mathcal{F}$,

$$\mathsf{H}_d(m) = \sum_{A\in\mathcal{F}} m(A)\log_2(|A|)\,. \tag{11}$$

Klir and Ramer

The measure presented by Klir and Ramer [12] was introduced to improve the measures H_O and H_y, and is defined as

$$\mathsf{H}_k(m) = -\sum_{A\in\mathcal{F}} m(A)\log_2\left(\sum_{B\in\mathcal{F}} m(B)\frac{|A\cap B|}{|B|}\right). \tag{12}$$

Klir and Parviz

An improvement of the previous measure was studied by Klir and Parviz [18]. They proposed the formulation

$$\mathsf{H}_p(m) = -\sum_{A\in\mathcal{F}} m(A)\log_2\left(\sum_{B\in\mathcal{F}} m(B)\frac{|A\cap B|}{|A|}\right). \tag{13}$$

Let us note that Klir and Ramer [12] and Klir and Parviz [18] also defined two measures of total uncertainty that are not usually cited in the literature.

Lamata and Moral

The uncertainty measures considered until now have the shortcoming of considering only some facets of uncertainty. Lamata and Moral [19] paved the way for more complex approaches when they defined a *global measure* to account for the two aspects of uncertainty. In a nutshell, Lamata and Moral [19] proposed summing the uncertainty measures of Yager [14] and Dubois and Prade [17]:

$$\mathsf{H}_l(m) = \overbrace{-\sum_{A\in\mathcal{F}} m(A)\log_2\left(\frac{\mathrm{Pl}(A)}{|A|}\right)}^{\mathsf{H}_y(m)+\mathsf{H}_d(m)}. \tag{14}$$

Pal et al.

Pal et al. [20] defined a new measure of total uncertainty,

$$\mathsf{H}_b(m) = \overbrace{-\sum_{A\in\mathcal{F}} m(A)\log_2\left(\frac{m(A)}{|A|}\right)}^{\mathsf{H}_d(m)+\mathsf{H}_n(m)}. \tag{15}$$

Even in this case, the measure can be decomposed into two separate measures of conflict and non-specificity.

Harmanec and Klir

Harmanec and Klir [21] considered the credal set $\mathcal{P}(m)$ and suggested taking the probability distribution $p \in \mathcal{P}(m)$ that maximizes Shannon entropy. Namely, they solved the following optimization problem:

$$\mathsf{AU}(m) = \max_{p\in\mathcal{P}(m)}\left\{-\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)\right\}, \tag{16}$$

where AU is the aggregate uncertainty. While being a potentially difficult optimization problem, some algorithmic procedures have been proposed to efficiently solve it and we adopted the one proposed by Huynh and Nakamori [22].

George and Pal

George and Pal [23] defined discord as the average of the conflict between each pair of pieces of evidence:

$$\mathsf{TC}(m) = \sum_{A\in\mathcal{F}} m(A) \sum_{B\in\mathcal{F}} m(B) \left(1 - \frac{|A\cap B|}{|A\cup B|}\right). \tag{17}$$

Jousselme et al.

Jousselme et al. [24] considered a set of reasonable properties and proposed another measure,

$$\mathsf{AM}(m) = -\sum_{i=1}^{n} \mathrm{Bet}(x_i)\log_2 \mathrm{Bet}(x_i), \tag{18}$$

where AM is the ambiguity measure, and $\mathrm{Bet}(x_i)$ is the pignistic transformation,

$$\mathrm{Bet}(x_i) = \sum_{A|x_i\in A} \frac{m(A)}{|A|} \quad \forall i = 1,\dots,n,$$

whose goal is to uniformly distribute the mass function of every focal element A to its elements $x_i \in A$.

Yang and Han

Yang and Han [25] considered the *belief interval*, $[\mathrm{Bel}(A), \mathrm{Pl}(A)]$, as representative of the uncertainty, and took into consideration the distance between the belief interval of a focal element and the most uncertain case $[0,1]$. Their uncertainty measure is

$$\mathsf{TU}^I(m) = 1 - \frac{1}{n}\sqrt{3}\sum_{i=1}^{n} d^I([\mathrm{Bel}(x_i), \mathrm{Pl}(x_i)], [0,1]), \tag{19}$$

where $\sqrt{3}$ is the normalization factor and

$$d^I\left([\mathrm{Bel}(x_i), \mathrm{Pl}(x_i)], [0,1]\right) = \sqrt{\left[\frac{\mathrm{Bel}(x_i)+\mathrm{Pl}(x_i)}{2} - \frac{0+1}{2}\right]^2 + \frac{1}{3}\left[\frac{\mathrm{Pl}(x_i)-\mathrm{Bel}(x_i)}{2} - \frac{1-0}{2}\right]^2}.$$

Deng and Wang

In an attempt to improve TU^I, Deng and Wang [26] used the Hellinger distance to measure the distance between intervals and proposed a new measure, which eventually collapses into

$$\mathsf{DU}(m) = \sum_{i=1}^{n}\left(1 - \sqrt{\left(\sqrt{\mathrm{Bel}(x_i)} - 0\right)^2 + \left(1 - \sqrt{\mathrm{Pl}(x_i)}\right)^2}\right). \tag{20}$$

Li et al.

Li et al. [27] used the Euclidean distance between intervals, and defined

$$d_E\left(\left[\mathrm{Bel}(x_i), \mathrm{Pl}(x_i)\right], [0,1]\right) = \sqrt{[\mathrm{Bel}(x_i)-0]^2 + [\mathrm{Pl}(x_i)-1]^2}.$$

Then, a transformation of this distance was considered and summed with respect to all the elements of the FOD so that it can be interpreted as a measure of uncertainty,

$$\mathsf{TU}(m) = \sum_{i=1}^{n}\left(\frac{2}{1+d_E} - 1\right). \tag{21}$$

Deng et al. [28] also employed the Euclidean distance to propose a similar uncertainty measure.

Wang and Song

Wang and Song [29] considered the relationship between all central values,

$$\frac{\mathrm{Bel}(x_i)+\mathrm{Pl}(x_i)}{2},$$

as a measure of conflict, and $\mathrm{Pl}(x_i) - \mathrm{Bel}(x_i)$ as a measure of non-specificity. The authors used the Shannon entropy to define conflict and included the non-specificity term in the definition,

$$\mathsf{SU}(m) = \sum_{i=1}^{n}\left[-\frac{\mathrm{Bel}(x_i)+\mathrm{Pl}(x_i)}{2}\log_2\frac{\mathrm{Bel}(x_i)+\mathrm{Pl}(x_i)}{2} + \frac{\mathrm{Pl}(x_i)-\mathrm{Bel}(x_i)}{2}\right]. \tag{22}$$

Deng

Deng [30] considered the measure proposed by Pal et al. [20] and modified it to give more importance to the relation between $|A|$ and the increase of non-specificity [31]:

$$\mathsf{H}_g(m) = -\sum_{A\in\mathcal{F}} m(A)\log_2\left(\frac{m(A)}{2^{|A|}-1}\right). \tag{23}$$

Formal properties of this measure have been critically inspected by Abellán [31], Moral-García and Abellán [32].

Pan and Deng

Pan and Deng [33] also used the central values of belief intervals proposed by Wang and Song [29], mixing this approach with the scaling factor proposed by Deng [30]. The resulting measure of uncertainty is[1]

$$\mathsf{H}_{bel}(m) = -\sum_{A\in\mathcal{F}} \frac{\mathrm{Bel}(A)+\mathrm{Pl}(A)}{2}\log_2\frac{\mathrm{Bel}(A)+\mathrm{Pl}(A)}{2(2^{|A|}-1)}. \tag{24}$$

Jiroušek and Shenoy

Jiroušek and Shenoy [6] proposed the measure

$$\mathsf{H}_j(m) = \overbrace{\sum_{x_i\in X}\mathrm{Pl}_{\mathrm{T}}(x_i)\log_2\left(\frac{1}{\mathrm{Pl}_{\mathrm{T}}(x_i)}\right)}^{\mathsf{H}_s(\mathrm{Pl}_{\mathrm{T}})} + \overbrace{\sum_{A\in\mathcal{F}} m(A)\log_2(|A|)}^{\mathsf{H}_d(m)}. \tag{25}$$

It may appear similar to the Jousselme et al. [24] ambiguity measure but, as argued by Jiroušek and Shenoy [6], the pignistic transformation used by Jousselme et al. [24] cannot satisfy one of the defined requirements. The plausibility transformation

$$\mathrm{Pl}_{\mathrm{T}}(x_i) = \frac{\mathrm{Pl}(x_i)}{\sum_{j=1}^{n}\mathrm{Pl}(x_j)}\ \forall x_i\in X$$

is another method to transform a belief function model into a corresponding probability function model.

Zhou et al.

Zhou et al. [34] defined entropy including, for the first time, the cardinality of the FOD. Compared to Deng entropy, in this measure,

$$\mathsf{E}_{Md}(m) = -\sum_{A\subseteq X} m(A)\log_2\left(\frac{m(A)}{2^{|A|}-1}e^{\frac{|A|-1}{|X|}}\right), \tag{26}$$

there is an additional term: the factor $\frac{|A|-1}{|X|}$ normalizes the cardinality of every focal element to the FOD.

Cui et al.

Other authors have started by including $|X|$ in their definition and also introduced the element of intersection between focal elements. Cui et al. [35] defined a novel uncertainty measure as follows:

$$\mathsf{E}(m) = -\sum_{A\subseteq X} m(A)\log_2\left(\frac{m(A)}{2^{|A|}-1}\exp\left(\sum_{\substack{B\in\mathcal{F}\\ B\neq A}}\frac{|A\cap B|}{2^{|X|}-1}\right)\right). \tag{27}$$

If there are no intersections, the equation reduces to Deng [30] entropy.

Yan and Deng

Yan and Deng [36] modified the definition of Zhou et al. [34] by combining it with the belief measure and by considering, instead of $|X|$, the number of singletons with $\mathrm{Pl} > 0$ indicated with $|S|$:

$$\mathsf{H}_{Md}(m) = -\sum_{A\subseteq X} m(A)\log_2\left(\frac{m(A)+\mathrm{Bel}(A)}{2(2^{|A|}-1)}e^{\frac{|A|-1}{|S|}}\right). \tag{28}$$

Qin et al.

Other authors have also had the idea of taking into account the cardinality $|X|$ of the FOD. Qin et al. [37] proposed a total uncertainty measure

[1] Let us note that, in the original formulation by Pan and Deng, this measure was defined as a sum over all $A\subseteq X$, but in the same paper the illustrative examples considered the sum over all $A\in\mathcal{F}$. We consider this latter formulation, which fits the numerical examples presented in the original paper [33].

$$\mathsf{H}_q(m) = \sum_{A\in\mathcal{F}} \frac{|A|}{|X|} m(A)\log_2(|A|) + \sum_{A\in\mathcal{F}} m(A)\log_2\left(\frac{1}{m(A)}\right), \tag{29}$$

where the entropy of Nguyen is used to measure the conflict and the entropy of Dubois and Prade measures non-specificity. The latter is scaled by the relative weight $|A|/|X|$.

Li and Pan

Unlike the preceding authors, Li and Pan [38] modified the non-specificity term, which is the uncertainty measure by Dubois and Prade [17] multiplied by $|X|$,

$$\mathsf{H}_{B\&F}(m) = \sum_{A\in\mathcal{F}} m(A)\log_2 \frac{|A|^{|X|}}{m(A)}. \tag{30}$$

Zhou and Deng

Zhou and Deng [39] proposed a measure of entropy based on a probability transformation that takes inspiration from the theory of fractals. More specifically, the fractal-based (FB) entropy measure is

$$\mathsf{E}_{FB} = -\sum_{A\in\mathcal{F}} m_F(A)\log_2 m_F(A) \tag{31}$$

where m_F is defined as follows:

$$m_F(A) = \sum_{B|A\subseteq B} \frac{m(B)}{2^{|B|}-1}.$$

Other measures

While the description of uncertainty measures has been, so far, quite extensive, we remark that some measures have not been analyzed in detail. For example, the formulations of the measures proposed by Wen et al. [7] and Zhao et al. [40] are extremely technical, while the two measures proposed by Zhang et al. [41] and Wang et al. [9] are parametric, and any comparison of them with other measures would be limited by having to fix a value for the parameters.

3.1. Mathematical properties of uncertainty measures

Although the focus of this paper is more on their numerical aspects, we consider it useful to present a brief summary of formal aspects of uncertainty measures. Over the years, a wide range of studies have analyzed the formal properties—in the form of desirable properties or axioms—of uncertainty measures. Recounting all of them would be beyond the scope of this paper, but it is safe to say that the following five have gained prominence:

1. *Probabilistic consistency*: If m takes the form of a probability distribution function, meaning that all focal elements are singletons, then the measure must be equal to the Shannon entropy.
2. *Set consistency*: When m focuses on a single set A (meaning that $m(A) = 1$ for one particular set A), then the measure of uncertainty must collapse into the Hartley entropy.
3. *Subadditivity*: When m is a joint BPA on the Cartesian product $X \times Y$ and m_X and m_Y are the respective associated marginal BPAs, then an uncertainty measure H is subadditive if and only if

   $$\mathsf{H}(m) \leq \mathsf{H}(m_X) + \mathsf{H}(m_Y).$$

4. *Additivity*: When m is a joint BPA on $X \times Y$ and m_X and m_Y are the respective associated marginal BPAs, then if these marginal BPAs are *non-interactive*, an uncertainty measure H is additive if and only if

   $$\mathsf{H}(m) = \mathsf{H}(m_X) + \mathsf{H}(m_Y).$$

5. *Monotonicity*: Consider any two BPAs m_1 and m_2 defined on the FOD X such that $\mathrm{Bel}_2(A) \leq \mathrm{Bel}_1(A)\ \forall A \subseteq X$, then H is monotone if and only if

   $$\mathsf{H}(m_1) \leq \mathsf{H}(m_2).$$

Other properties, for instance, consistency with Dempster–Shafer semantics, range, non-negativity, and maximum entropy, were discussed by Jiroušek and Shenoy [6]. Table 1 presents a summary of the uncertainty measures reported in this section and whether they satisfy the above five properties. We also attempt a classification according to the type of uncertainty that is captured by each measure and the paradigm on which the measure is based.

As recently confirmed by Moral-García and Abellán [46], the uncertainty measure of Harmanec and Klir is the only one proposed in the literature that satisfies all the above-mentioned properties. However, it is quite complex to implement in practice. It can be seen from Table 1 that a full analysis has not been carried out for the most recent measures of uncertainty. In some cases, the satisfaction of some properties has been verified using examples and therefore more work is needed toward a formal proof. The

Table 1
Measures of uncertainty and their properties: a review of the literature. Y: the property has been formally established, N: the property is not satisfied, ?: the property was not studied or evidence seems inconclusive. The column "Source" lists the publications in which the satisfaction of the properties is discussed. In the column "Content," measures are labeled according to the category they belong to: total uncertainty (TU), discord (D), or non-specificity (NS). The last column indicates whether the measure is based on the entropy formulation (EB) and/or uncertainty intervals (IB).

Eq.	Proponent(s)	P. cons.	S. cons.	Add.	Subadd.	Monot.	Source	Type	Content
(7)	Höhle	Y	N	Y	N	N	[8,37]	D	EB
(8)	Yager	Y	N	Y	N	N	[8,37]	D	EB
(9)	Smets	N	N	Y	N	N	[8]	D	EB
(10)	Nguyen	Y	N	Y	N	N	[8,37]	D	EB
(11)	Dubois and Prade	N	Y	Y	Y	Y	[8,37]	NS	EB
(12)	Klir and Ramer	Y	Y	Y	N	Y	[8,37]	TU	EB
(13)	Klir and Parviz	Y	Y	Y	N	Y	[8,37]	TU	EB
(14)	Lamata and Moral	Y	Y	Y	N	Y	[6]	TU	EB
(15)	Pal et al.	Y	Y	Y	N	Y	[8]	TU	EB
(16)	Harmanec and Klir	Y	Y	Y	Y	Y	[8,31]	TU	EB
(17)	George and Pal	Y	Y	Y	N	N	[42,8]	TU	
(18)	Jousselme et al.	Y	Y	Y	N	Y	[8,43]	TU	EB
(19)	Yang and Han	N	N	N	N	N	[44,45]	TU	IB
(20)	Deng and Wang	N	N	?	?	Y	[26]	TU	IB
(21)	Li et al.	N	N	N	N	?	[27]	TU	IB
(22)	Wang and Song	Y	Y	?	?	N	[29,44]	TU	EB, IB
(23)	Deng	Y	N	N	N	N	[31]	TU	EB
(24)	Pan and Deng	Y	?	N	N	Y	[33,8]	TU	EB
(25)	Jiroušek and Shenoy	Y	N	Y	N	Y	[6,37]	TU	EB
(26)	Zhou et al.	N	N	N	N	N	[32]	TU	EB
(27)	Cui et al.	N	N	N	N	N	[32]	TU	EB
(28)	Yan and Deng	Y	Y	N	N	?	[36]	TU	EB
(29)	Qin et al.	Y	N	N	?	?	[37]	TU	EB
(30)	Li and Pan	Y	?	N	N	?	[38]	TU	EB
(31)	Zhou and Deng	Y	N	Y	Y	?	[39]	TU	EB

relative importance of these properties is, however, still subject to debate. For example, Deng [44] stated that "monotonicity is one of the most important properties." Meanwhile, Yang and Han [25] have argued that properties like additivity and subadditivity are "meaningless" for their uncertainty measure, since it is not a direct extension of probability theory but an *ad hoc* proposal for evidence theory. Deng and Wang [26] stated that the "evidence theory framework is different from the probability theory framework and the classical set theory," meaning that substantially different properties should be required. They also argued that additivity and subadditivity are relevant only if the uncertainty measure is applied to the Cartesian product $X \times Y$ and irrelevant otherwise. For this reason, some alternative axiomatic frameworks have been promoted, for example in [44].

In addition to the above-mentioned formal properties, some desiderata have been proposed. In contrast to formal properties, they are often subjective and their fulfillment is a matter of degree. Moral-García and Abellán [46] proposed the following four desiderata:

- The calculation of the uncertainty measure should not be excessively complex.
- It should be possible to decompose the uncertainty measure into its conflict and non-specificity parts.
- The measure should be sensitive to changes in m.
- The measure should be applicable to frameworks more general than evidence theory.

For an updated inquiry into mathematical properties and desiderata of uncertainty measures, one can refer to Moral-García and Abellán [46], in which a deeper analysis is presented that makes a distinction between BPA- and interval-based uncertainty measures.

3.2. A brief historical perspective

Fig. 1 offers a snapshot of the development of uncertainty measures over time and, quite remarkably, seems to divide the last 40 years into four distinct periods. The first period (1982–1987) saw the inception of the most basic measures, capable of capturing only some facets of the total uncertainty. In the second period (1988–1995), starting with the measure proposed by Lamata and Moral [19], efforts were redirected toward (i) the formulation of more holistic measures that can capture the total uncertainty of mass assignment and (ii) their formal analysis. The third period (1996–2015) is the longest and saw few new proposals. In the fourth and current period (2016–today) we are witnessing a strong resurgence of uncertainty measures. In particular, new measures are often inspired by existing ones and justified by showing better performance with respect to some specific examples. Moreover, in recent years, as noted with some concern by Dezert and Tchamova [47], a weaker emphasis has been placed on the analysis of formal properties of these measures, possibly in light of the impossibility of finding a measure that can satisfy all properties and desiderata.

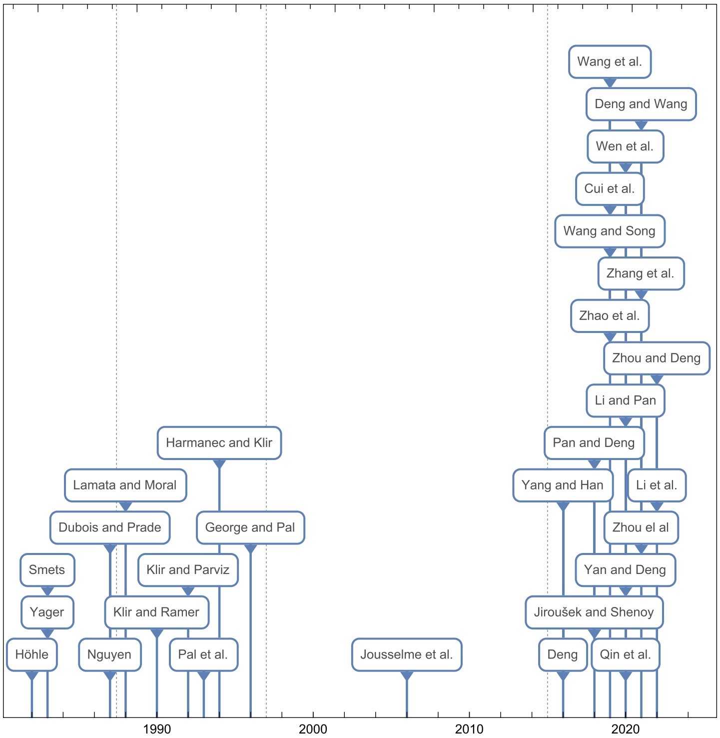


Fig. 1. Timeline of the seminal papers for various uncertainty measures.

4. Methodology and results

As already specified, we are interested in studying the relations between different uncertainty measures by means of simulations. The approach proposed in this study is based on an a priori definition of the cardinality n of X. Then, the set of focal elements $\mathcal{F}$ is constructed by randomly sampling subsets of the FOD X until the union of the sampled subsets is a *cover* of X, that is, until each element of X appears in at least one of the selected subsets. A BPA m over the elements of $\mathcal{F}$ is assigned using the Dirichlet function with parameter vector $\mathbf{1}_{|\mathcal{F}|}$. It is important to note that the choice of the Dirichlet function parameterized with $\mathbf{1}_{|\mathcal{F}|}$ determines that the sampling of the values of m over $\mathcal{F}$ occurs uniformly and that the sum of their values is one. In other words, once a set of focal elements $\mathcal{F}$ covering X is sampled, then the positive values assigned to the $|\mathcal{F}|$ focal elements are uniformly sampled from the $|\mathcal{F}|$-dimensional unit simplex.

This procedure for sampling BOEs is the same as that proposed by Burger and Destercke [48], in which the stopping condition is that the union of the elements in the FOD covers the set X. The simulation consists of fixing n and then repeating the procedure $s = 10,000$ times so that, for each repetition, the uncertainty of the BPA is measured by means of the uncertainty measures presented in Section 3. The simulation procedure is summarized in Algorithm 1 and the simulation code is available online.[2]

Let us note that some of the methods used here to analyze uncertainty measures have already been adopted within evidence theory to study some other aspects of belief functions. See, for example, the contribution by Jousselme and Maupin [49], in which scatter plots and dendrograms were used to compare distances between belief functions.

4.1. Similarity analysis

First, we analyze the similarities between uncertainty measures by checking their comonotonicity. We assume that the similarity between two uncertainty measures depends on their tendency to order BPAs in a similar way from the most to the least uncertain. That is, given s BPAs, two uncertainty measures are similar if they agree on how to order the s BPAs from the most to the least uncertain. Given the possible existence of non-linear relations between values obtained from different uncertainty measures, we prefer to use a comonotonicity measure that is agnostic to the nature of the comonotonicity. For this reason, for example, we excluded Pearson's linear correlation, which assumes any comonotonicity to be linear. For a similar reason, we ruled out polynomial and parametric regressions. Considering these requirements, we used the Spearman rank correlation coefficient ρ, which, given two numerical lists $\mathbf{a}, \mathbf{b} \in \mathbb{R}^s$, returns a value in $[-1, 1]$.

[2] https://gitlab.com/or-group-dii-unitn-public/evidence-theory.

Algorithm 1 The simulation procedure to compare the selected uncertainty measures.

```
X ← {x_1, ..., x_n}                                   ▷ Define the frame of discernment
N ← s                                                  ▷ Initialize N
Q ← {q_1, ..., q_m}                                    ▷ Define a list of entropy measures
D ← ∅                                                  ▷ Create an empty dataset
i ← 1
while i ≤ N do
    R ← ∅
    while ⋃_{A∈R} A ⊊ X do
        A ← sample(2^X \ R)                            ▷ Sample a set A ⊆ 2^X \ R
        R ← R ∪ A
    end while
    F ← ∅
    B ← Dirichlet(|R|)                                 ▷ Sample |R| values using Dirichlet so that ∑_{b∈B} b = 1
    for b ∈ B do
        F ← F ∪ {⟨A, b⟩}
    end for
    D ← D ∪ {⟨q_1(F), ..., q_m(F)⟩}                    ▷ Compute a new array of entropy measures
    i ← i + 1
end while
```

The Spearman coefficient assumes a preprocessing phase in which raw values a_i, b_i are transformed into ranks $R(a_i), R(b_i)$ so that $R(\mathbf{a})$ and $R(\mathbf{b})$ are the vectors containing the ordinal rankings of the components of **a** and **b**, respectively. Then,

$$\rho(\mathbf{a},\mathbf{b}) = \frac{\mathrm{cov}(R(\mathbf{a}), R(\mathbf{b}))}{\sigma_{R(\mathbf{a})}\sigma_{R(\mathbf{b})}},$$

where cov gives the covariance of the two ranking vectors and σ is the standard deviation. More precisely, the value $\rho(\mathbf{a},\mathbf{b}) = 1$ denotes perfect positive comonotonicity between the two lists, $\rho(\mathbf{a},\mathbf{b}) = -1$ denotes perfect negative comonotonicity, and $\rho(\mathbf{a},\mathbf{b}) = 0$ no comonotonicity.

Fixing $n = 4$, Fig. 2 shows scatter plots for all pairs of the considered uncertainty measures, together with their Spearman coefficients. Note that we exclude the measures defined by Eqs. (7)–(13) and (17) from the analysis to save space and because these measures can only partially capture uncertainty or have been superseded by more refined ones. Consequently, all the measures in Fig. 2 are measures of total uncertainty. A closer analysis of a selection of these scatter plots can shed more light on the similarities and dissimilarities between uncertainty measures and help quantify the extent of their possible drawbacks.

Fig. 3a compares the uncertainty measure proposed by Jiroušek and Shenoy with that of Wang and Song. With $\rho \approx 0.99$, this is a representative example of two extremely similar uncertainty measures that, however, have significantly different formulations, which may not have led one to suspect their similarity. This may suggest that, for practical purposes, the two measures are almost interchangeable. It is also worth noting that, as shown in Table 1, in spite of their similarity, the two measures have remarkably different mathematical properties.

Fig. 3b shows an example of two uncertainty measures that are negatively correlated ($\rho \approx -0.204$), even if their polarity should be the same. The scatter plot shows a cloud of points with a barely perceivable (negative) comonotonicity. In this case, it is not only difficult to accept the conclusion that the two measures capture the same phenomenon for which they were proposed, but according to the Spearman coefficients, they appear more likely to produce opposite results. Indeed, if applied to real-world problems, these two measures are likely to give contrasting assessments. However, given the complex formulation of both uncertainty measures it is, once again, difficult to find an explanation for this behavior.

The uncertainty measure AU proposed by Harmanec and Klir satisfies a number of desirable properties, but has been criticized in the literature for its insensitivity to changes in the mass assignment. In particular, changes in m often do not lead to changes in the estimated uncertainty. In fact, anytime the mass assignment m is compatible with a uniform probability distribution on the singletons, this latter is considered and the value of the uncertainty measure saturates and reaches its maximum. However, this insensitivity of the measure by Harmanec and Klir has not been sufficiently studied and its extent and consequences are presently unknown. Fig. 3c compares the uncertainty measure proposed by Harmanec and Klir with that of Li et al. One can see that the uncertainty measure of Harmanec and Klir seems to have a maximum value that attracts closer values: this corresponds to the long vertical alignment of points on the right-hand side of the plot. In fact, there is a significant gap between its maximum value and the second largest value. In this way, different mass assignments which, according to the uncertainty measure of Li et al., have greatly different levels of uncertainty may be associated with the maximum value, even if they are very much distinct in terms of total uncertainty. The comparative behavior of the uncertainty measure by Harmanec and Klir with other measures suggests that this undesirable behavior, already discussed in the literature, is a non-negligible phenomenon impairing the discriminating power of the measure AU.

At present, it may be hard to interpret values returned by different uncertainty measures. For instance, if we consider the measure AM proposed by Jousselme et al., which, for $n = 4$, lies in the interval $[0, 2]$, should we consider a value of, say, 1.25 representative of a high or low total uncertainty? While one may be tempted to lean toward the former, it could be more appropriate to prefer the latter, in light of the left-skewed distribution of values on the corresponding diagonal entry in Fig. 2 and detailed in Fig. 4a.

Note that the value 1.25 belongs to the decile containing the least uncertain mass assignments and for this reason, it cannot be considered representative of a high level of uncertainty. Conversely, a value of 1.25 for the uncertainty measure proposed by Lamata and Moral [19] represents a higher level of uncertainty, as can be seen in Fig. 4b. Note that both measures are expressed

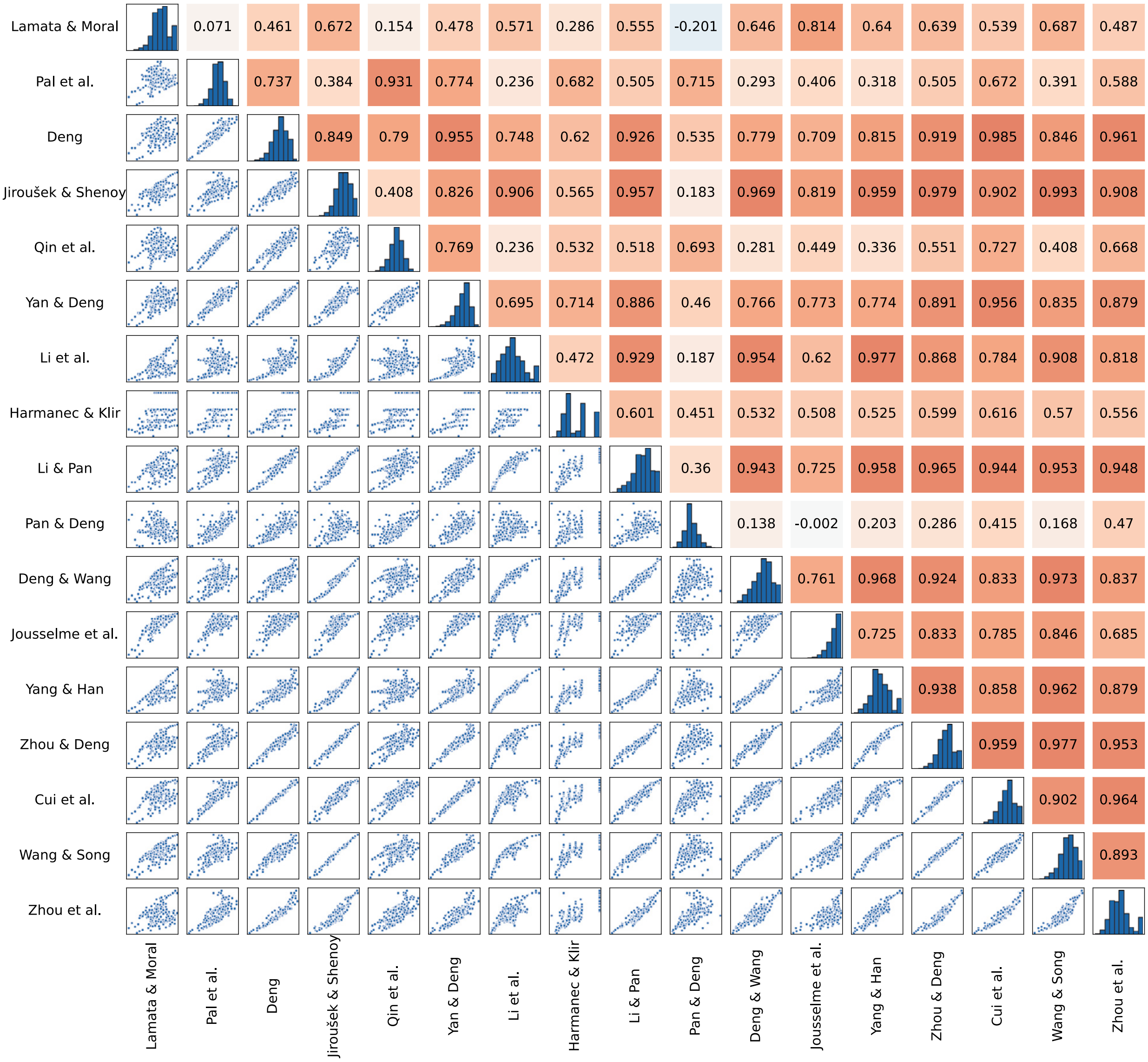
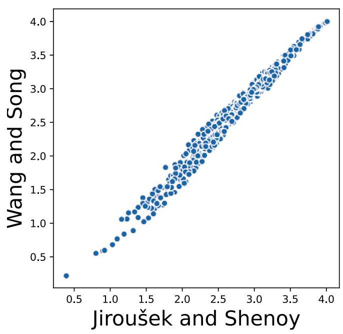
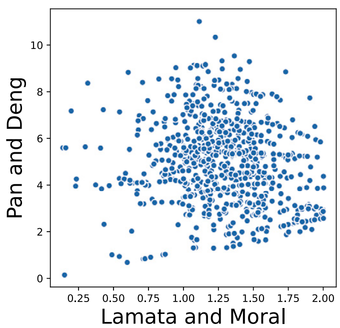


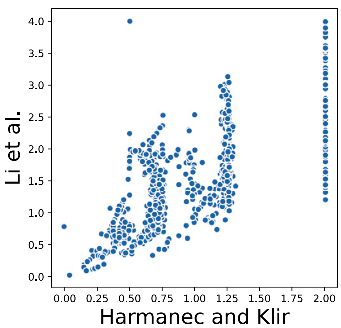
Fig. 2. Pairwise scatter plots for pairs of uncertainty measures and their Spearman rank correlation coefficients for $n = 4$. To enhance readability, the scatter plots are based on $s = 300$ simulations whereas, for greater stability, the values in the upper triangular part were obtained with $s = 10,000$.



(a) Jiroušek and Shenoy (H_j) vs. Wang and Song (SU).

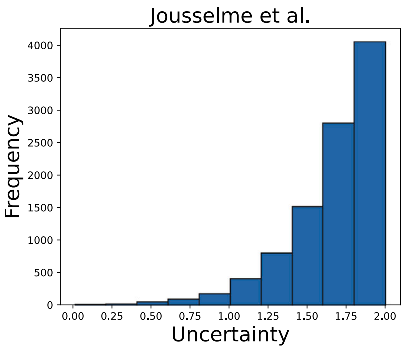


(b) Pan and Deng (H_{bel}) vs. Lamata and Moral (H_l).

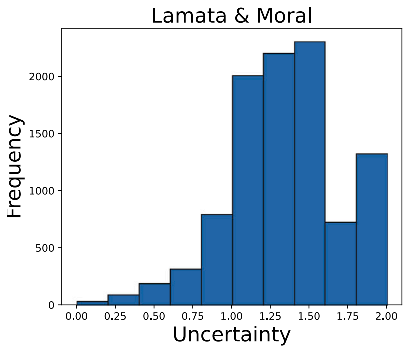


(c) Harmanec and Klir (AU) vs. Li et al. (TU).

Fig. 3. Three representative scatter plots with $s = 1000$.



(a) Bar chart of the values obtained using the uncertainty measure of Jousselme et al. [24].



(b) Bar chart of the values obtained using the uncertainty measure of Lamata and Moral [19].

Fig. 4. Two representative bar charts with $n = 4$ and $s = 10,000$.

on the same scale $[0, \log_2 n]$ and are therefore comparable. Analyzing the bar charts on the diagonal of Fig. 2, one can observe different patterns. Such patterns are not a mere matter of curiosity, but cast serious doubt on the possibility of interpreting values of uncertainty measures without knowing their distributions. A formal way to compare two uncertainty measures is to approximate their distributions to the normal distribution—using, for example, a power transform such as the Box–Cox transformation—or any other distribution, and then to use the transformed values for comparison. We conclude that knowing the range of uncertainty measures is not sufficient to interpret the levels of uncertainty of different mass assignments.

The results of this numerical study might also be useful when looking for support for a choice of uncertainty measure. In fact, one could argue against choosing a single measure and instead relying on the conjoint use of a range of measures to increase the robustness of the uncertainty analysis. Two measures with low similarity—here quantified by the Spearman rank correlation—are more likely to provide separate evidence and, possibly, stronger confirmation of whether a BOE is too uncertain. Conversely, considering a second extremely similar measure would add redundancy to the analysis. The more two measures are regarded as dissimilar but in fact produce similar uncertainty values, the more we may think that there is agreement on the measured phenomenon, and distinct evidence backing up the first measurement.

4.2. Hierarchical clustering

We consider the Spearman coefficient as the basis for estimating the dissimilarities between uncertainty measures. In particular, if we denote by $\rho_{\mathsf{H}_i,\mathsf{H}_j}$ the value of the Spearman coefficient calculated for two measures of uncertainty H_i and H_j, then we can consider $d_{ij} = |1 - \rho_{\mathsf{H}_i,\mathsf{H}_j}| < 2$ a measure of their dissimilarity. We can use hierarchical clustering to analyze the values of d_{ij}. Hierarchical clustering is a non-parametric technique that, however, depends on the choice of the heuristic that determines when two clusters, say X and Y, are merged. Initially, each element is a cluster of its own, and then they are progressively merged according to one of the following heuristics:

- *Single linkage*: The value at which the two clusters X and Y are merged corresponds to the minimum distance between an element of X and one of Y, that is, $\min\{d_{ij} | i \in X, j \in Y\}$.
- *Complete linkage*: The value at which the two clusters X and Y are merged corresponds to the maximum distance between an element of X and one of Y, that is, $\max\{d_{ij} | i \in X, j \in Y\}$.
- *Average linkage*: The value at which the two clusters X and Y are merged corresponds to the average distance between elements belonging to the two clusters, that is,

$$\frac{1}{|X||Y|} \sum_{i \in X} \sum_{j \in Y} d_{ij}.$$

- *Ward linkage*: The value at which two clusters X and Y are merged corresponds to the cluster distance. The distance between two clusters X and Y is

$$d(X,Y) = \sqrt{\frac{|Y|+|Z|}{T} d(Y,Z)^2 + \frac{|Y|+|W|}{T} d(Y,W)^2 - \frac{|Y|}{T} d(Z,W)^2},$$

where X is the newly joined cluster consisting of Z and W, Y is the cluster to be merged with, and $T = |Y| + |Z| + |W|$. When two clusters each contain a single element, $d(X,Y) = d_{XY}$. The Ward variance minimization algorithm [50] is used to perform clustering[3]

[3] https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.linkage.html.

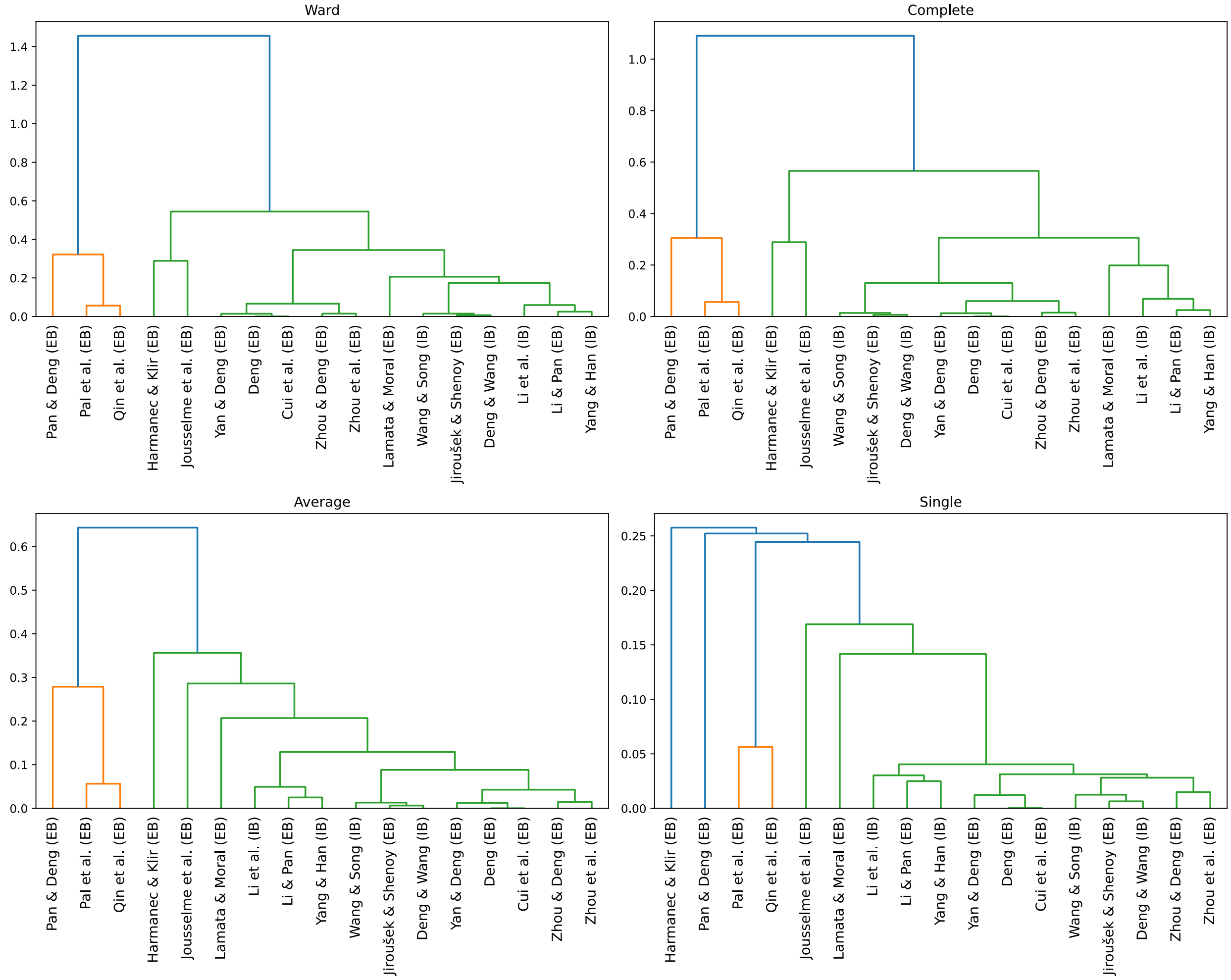


Fig. 5. A comparison of dendrograms obtained using different clustering heuristics. The values on the y-axis correspond to the distance between clusters according to the chosen heuristic. Along with the name of each measure, the EB/IB classification is reported as in Table 1; all measures express total uncertainty.

The application of these four heuristics to the values in Fig. 2 is reported in the dendrograms in Fig. 5, from which some conclusions can be drawn. Approaches based on distances between intervals do not belong to a separate cluster and are instead mixed with the others in a cluster of extremely similar measures—on the left-hand side in the dendrogram for the single linkage heuristic. In general, there is not a sharp separation of measures according to their formulations. Thus, it may be deceiving to judge the differences between methods based on their mathematical formulations. For instance, the three measures proposed by Pal et al. (Eq. (15)), Qin et al. (Eq. (29)), and Li and Pan (Eq. (30)) are based on similar formulations but only the former two are numerically similar. Interestingly, the separation into clusters seems to follow the years of inception of different measures: with few exceptions, all the most recent measures are part of a large cluster of similar measures. This issue will be further investigated in the next subsection.

4.3. Centrality analysis

The Spearman coefficients in Fig. 2, which are always in the range $[-1, 1]$, can be interpreted as degrees of closeness between methods. By doing so, we can consider the matrix of Spearman coefficients underlying Fig. 2 as an adjacency matrix.

Adjacency matrices are well-known mathematical structures used in network analysis and therefore it is natural to employ tools developed in network analysis, for example in the study of social networks. In particular, centrality measures, as the name suggests, are used in network analysis to measure the extent to which nodes can be considered "central" in the context of the graph of which they are elements.

If we consider weighted graphs in which the adjacency degree of two nodes can be interpreted as the degree of mutual support, a centrality measure shows the degree of support that each node receives from the others. The Spearman coefficient can be seen as a measure of the mutual support between two uncertainty measures when it comes to ranking BOEs from the least to the most uncertain: the higher the value, the more the results of one uncertainty measure support those of the other. Among the various measures of centrality, the eigenvector centrality, which corresponds to the eigenvector associated with the Perron–Frobenius eigenvalue of a

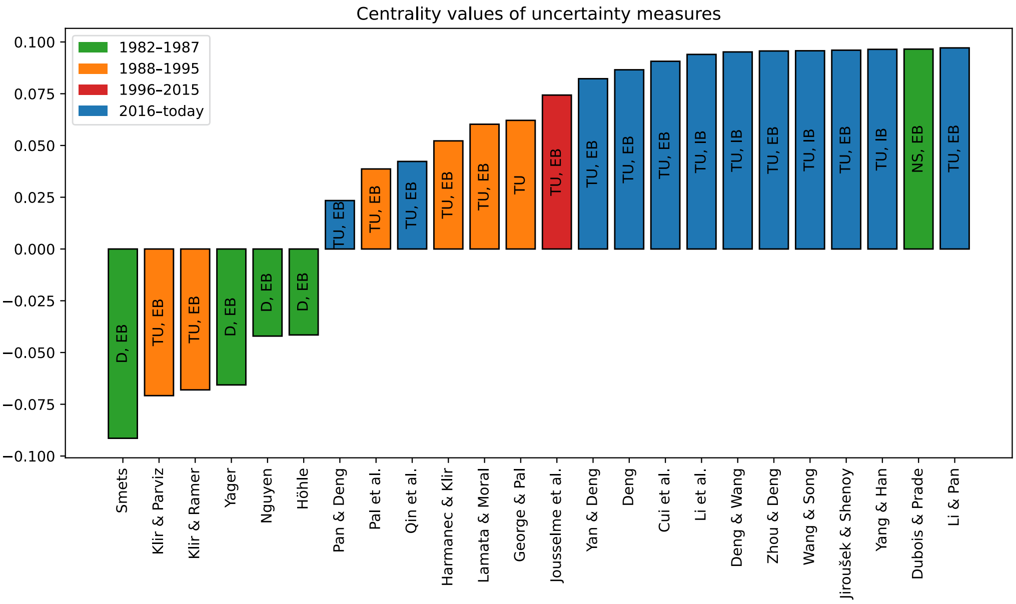


Fig. 6. A bar chart representing the values of normalized eigenvector centrality of each uncertainty measure.

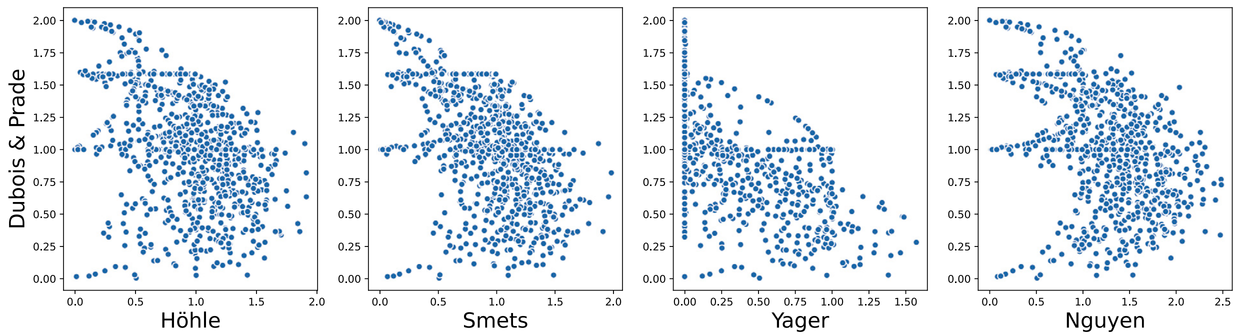


Fig. 7. The measure of non-specificity proposed by Dubois and Prade [17] compared with four measures of conflict. The respective Spearman coefficients are −0.488, −0.933, −0.705, and −0.525.

weighted adjacency matrix, has assumed a prominent role thanks to its capacity to account for indirect relations, its good axiomatic properties, and the fact that it is flexible enough to be used with signed graphs too. In practice, given a weighted adjacency matrix $\mathbf{A}$, in our case stemming from Fig. 2, the centrality values of the uncertainty measures correspond to the components of the vector $\mathbf{w}$ solving the eigensystem $\mathbf{A}\mathbf{w} = \lambda_{\max}\mathbf{w}$, where $\lambda_{\max}$ is the Perron–Frobenius eigenvalue of $\mathbf{A}$. Note that $\mathbf{w}$ is unique up to multiplication by a positive scalar.

Fig. 6 reports the normalized eigenvector centrality for the uncertainty measures considered in this study, including the older ones. The more central an uncertainty measure is, the more support it receives from the other measures. Therefore, a high value of centrality can be seen as an indication that a measure is holistic, as it can, to some extent, also represent the points of view offered by alternative measures. In this sense, in absence of decisive results, high centrality values should be seen favorably. On the other hand, it is also true that low centrality values are symptomatic of original approaches providing separate evidence. The results in Fig. 6 show a wide range of behaviors. A number of measures have negative centrality values, which is indicative of a negative comonotonicity with other measures. Remarkably, all the uncertainty measures with negative centrality are measures of conflict. Hence, it seems that when a term is added to a measure of conflict to account for the non-specificity, this does not only alter the measurement, but it does it in a very specific (and opposite) direction: the lower the non-specificity, the higher the conflict, and vice versa. This is illustrated in Fig. 7 and seems to further corroborate the need to account for both facets of uncertainty to avoid under- or overestimations. To understand the relevance of this observation, one may, for instance, consider evidence-based fusion systems in which one is interested in assessing the uncertainty level before choosing whether to fuse or not.

Furthermore, it is possible to see that, with a few exceptions, the most recent measures have high centrality values and the old ones have negative values. This could be symptomatic of a recent convergence of research toward a shared definition of uncertainty measure. Among the exceptions to this phenomenon, the measure of Dubois and Prade is remarkable: in spite of its extreme simplicity

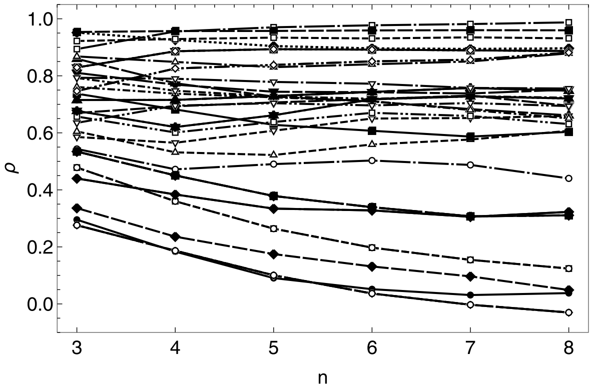


Fig. 8. Sensitivity analysis for randomly selected similarity values between uncertainty measures. Each line represents the similarity of a different pair of uncertainty measures and how it changes with respect to the cardinality of the FOD.

and the fact that it can capture only some aspects of uncertainty, it seems to be supported by other measures. One reason for this could be the fact that the formulation of this uncertainty measure is embedded in many of the most recent ones.

A cautionary note may be necessary. As the centrality values depend on the measures considered in the analysis, the values presented in Fig. 6 can vary according to the included measures. On the other hand, it is also sensible to assume that such variations would not be so significant as to invalidate the conclusions drawn from our results.

So far, the results have been presented for $n = 4$, which is a recurring size of the FOD in most of the illustrative examples presented in the literature. To explore the generalizability of the results, we explored the cases with $n = 3, \dots, 8$. Fig. 8 shows a graphical sensitivity analysis for 30 Spearman rank correlation coefficients for randomly chosen pairs of uncertainty measures.

Considering Fig. 8, the values for $n = 4$ correspond to 30 random values from Fig. 2. Then, for the same pairs of measures, the simulations were repeated for other values of n and the new similarities were collected and compared. While the results in Fig. 8 are qualitative, they certainly show some changes in the similarities but, with a few exceptions (low levels tend to become lower when n increases), it is difficult to spot clear patterns. In particular, high levels of similarity remain high regardless of the value of n. For reasons of space, a full comparative analysis with respect to the parameter n is not presented here, but the interested reader can refer to some supplementary material online.[4]

5. Conclusions

First, we outlined the most prominent uncertainty measures in the evidence theory, hoping that this could help other researchers navigate this *mare magnum*. Next, the results of Monte Carlo simulations were presented, in which the behavior of different uncertainty measures were compared pairwise, allowing us to analyze their similarities and differences. The formal tools employed in the analysis include rank correlation, hierarchical clustering, and eigenvector centrality.

Besides the results, already analyzed in the previous section, we want to stress the *necessity* of numerical studies to interpret values returned by different uncertainty measures, which could otherwise be interpreted only in an ordinal sense. That is, given an uncertainty measure H and two mass assignments m_1 and m_2 on the same FOD, $\mathsf{H}(m_1) \geq \mathsf{H}(m_2)$ implies that, according to H, m_1 is more uncertain than m_2, but it is presently hard to draw more conclusions on the uncertainty levels of both m_1 and m_2. Maybe m_1 and m_2 are both too uncertain to be sufficiently specific. Or maybe they are not. In fact, in some applications, it may be reasonable to disregard mass assignments when they are too uncertain to yield sufficient information. This problem is similar to the quantification of the inconsistency of preferences in decision analysis, as happens in the analytic hierarchy process, in which preferences that are too inconsistent may be rejected and sent for re-evaluation, seeking a greater level of consistency. This may also trigger some research on some acceptability thresholds, which could be estimated as quantiles of the empirical distributions or with methods such as AUC (area under the curve) of the ROC (receiver operating characteristic).

Given the existing literature on uncertainty measures and the results presented in this manuscript, it is reasonable to expect that future proposals of new measures should compare the newly introduced measure with the existing ones (i) numerically, to detect possible similarities/dissimilarities, and (ii) formally, to show a *clear advantage* with respect to the existing literature.

CRediT authorship contribution statement

Michele Urbani: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Gaia Gasparini:** Data curation, Investigation, Visualization, Writing – original draft. **Matteo Brunelli:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

[4] https://gitlab.com/or-group-dii-unitn-public/evidence-theory.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We appreciated the suggestions of three anonymous reviewers and the Editor, which helped us improve the quality of the manuscript. The research of Gaia Gasparini is supported by Autostrada del Brennero S.p.A., Italy.

References

[1] B.M. Ayyub, G.J. Klir, Uncertainty Modeling and Analysis in Engineering and the Sciences, Chapman and Hall/CRC, 2006.
[2] D. Dubois, H. Prade, Possibility Theory: An Approach to Computerized Processing of Uncertainty, Springer Science & Business Media, 1988.
[3] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat. 38 (1967) 325–339.
[4] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.
[5] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423.
[6] R. Jiroušek, P.P. Shenoy, A new definition of entropy of belief functions in the Dempster–Shafer theory, Int. J. Approx. Reason. 92 (2018) 49–65.
[7] K. Wen, Y. Song, C. Wu, T. Li, A novel measure of uncertainty in the Dempster-Shafer theory, IEEE Access 8 (2020) 51550–51559.
[8] Q. Pan, D. Zhou, Y. Tang, X. Li, J. Huang, A novel belief entropy for measuring uncertainty in Dempster-Shafer evidence theory framework based on plausibility transformation and weighted Hartley entropy, Entropy 21 (2019) 163.
[9] D. Wang, J. Gao, D. Wei, A new belief entropy based on Deng entropy, Entropy 21 (2019) 987.
[10] Q. Zhou, É. Bossé, Y. Deng, Modeling belief propensity degree: measures of evenness and diversity of belief functions, IEEE Trans. Syst. Man Cybern. Syst. (2022).
[11] A.G. Bronevich, A.E. Lepskiy, Measures of conflict, basic axioms and their application to the clusterization of a body of evidence, Fuzzy Sets Syst. 446 (2022) 277–300.
[12] G.J. Klir, A. Ramer, Uncertainty in the Dempster-Shafer theory: a critical re-examination, Int. J. Gen. Syst. 18 (1990) 155–166.
[13] U. Höhle, Entropy with respect to plausibility measures, in: Proc. of 12th IEEE Int. Symp. on Multiple Valued Logic, Paris, 1982, 1982, pp. 167–169.
[14] R.R. Yager, Entropy and specificity in a mathematical theory of evidence, Int. J. Gen. Syst. 9 (1983) 249–260.
[15] P. Smets, Information content of an evidence, Int. J. Man-Mach. Stud. 19 (1983) 33–43.
[16] H.T. Nguyen, On entropy of random sets and possibility distributions, in: The Analysis of Fuzzy Information, vol. 1, 1987, pp. 145–156.
[17] D. Dubois, H. Prade, Properties of measures of information in evidence and possibility theories, Fuzzy Sets Syst. 24 (1987) 161–182.
[18] G.J. Klir, B. Parviz, A note on the measure of discord, in: D. Dubois, M.P. Wellman, B. D'Ambrosio, P. Smets (Eds.), Uncertainty in Artificial Intelligence, Morgan Kaufmann, 1992, pp. 138–141.
[19] M.T. Lamata, S. Moral, Measures of entropy in theory of evidence, Int. J. Gen. Syst. 14 (1988) 297–305.
[20] N.R. Pal, J.C. Bezdek, R. Hemasinha, Uncertainty measures for evidential reasoning II: a new measure of total uncertainty, Int. J. Approx. Reason. 8 (1993) 1–16.
[21] D. Harmanec, G. Klir, Measuring total uncertainty in Dempster-Shafer theory: a novel approach, Int. J. Gen. Syst. 22 (1994) 405–419.
[22] V.-N. Huynh, Y. Nakamori, Notes on "reducing algorithm complexity for computing an aggregate uncertainty measure", IEEE Trans. Syst. Man Cybern., Part A, Syst. Hum. 40 (2009) 205–209.
[23] T. George, N.R. Pal, Quantification of conflict in Dempster-Shafer framework: a new approach, Int. J. Gen. Syst. 24 (1996) 407–423.
[24] A.-L. Jousselme, C. Liu, D. Grenier, E. Bossé, Measuring ambiguity in the evidence theory, IEEE Trans. Syst. Man Cybern., Part A, Syst. Hum. 36 (2006) 890–903.
[25] Y. Yang, D. Han, A new distance-based total uncertainty measure in the theory of belief functions, Knowl.-Based Syst. 94 (2016) 114–123.
[26] Z. Deng, J. Wang, Measuring total uncertainty in evidence theory, Int. J. Intell. Syst. 36 (2021) 1721–1745.
[27] R. Li, Z. Chen, H. Li, et al., A new distance-based total uncertainty measure in Dempster-Shafer evidence theory, Appl. Intell. 52 (2022) 1209–1237.
[28] X. Deng, F. Xiao, Y. Deng, An improved distance-based total uncertainty measure in belief function theory, Appl. Intell. 46 (2017) 898–915.
[29] X. Wang, Y. Song, Uncertainty measure in evidence theory with its applications, Appl. Intell. 48 (2017) 1672–1688.
[30] Y. Deng, Deng entropy, Chaos Solitons Fractals 91 (2016) 549–553.
[31] J. Abellán, Analyzing properties of Deng entropy in the theory of evidence, Chaos Solitons Fractals 95 (2017) 195–199.
[32] S. Moral-García, J. Abellán, Critique of modified Deng entropies under the evidence theory, Chaos Solitons Fractals 140 (2020) 110112.
[33] L. Pan, Y. Deng, A new belief entropy to measure uncertainty of basic probability assignments based on belief function and plausibility function, Entropy 20 (2018).
[34] D. Zhou, Y. Tang, W. Jiang, A modified belief entropy in Dempster-Shafer framework, PLoS ONE 12 (2017).
[35] H. Cui, Q. Liu, J. Zhang, B. Kang, An improved Deng entropy and its application in pattern recognition, IEEE Access 7 (2019) 18284–18292.
[36] H. Yan, Y. Deng, An improved belief entropy in evidence theory, IEEE Access 8 (2020) 57505–57516.
[37] M. Qin, Y. Tang, J. Wen, An improved total uncertainty measure in the evidence theory and its application in decision making, Entropy 22 (2020) 487.
[38] J. Li, Q. Pan, A new belief entropy in Dempster–Shafer theory based on basic probability assignment and the frame of discernment, Entropy 22 (2020) 691.
[39] Q. Zhou, Y. Deng, Fractal-based belief entropy, Inf. Sci. 587 (2022) 265–282.
[40] Y. Zhao, D. Ji, X. Yang, L. Fei, C. Zhai, An improved belief entropy to measure uncertainty of basic probability assignments based on Deng entropy and belief interval, Entropy 21 (2019).
[41] Y. Zhang, F. Huang, X. Deng, W. Jiang, A new total uncertainty measure from a perspective of maximum entropy requirement, Entropy 23 (2021).
[42] Y. Deng, Uncertainty measure in evidence theory, Sci. China Inf. Sci. 63 (2020) 1–19.
[43] G.J. Klir, H.W. Lewis, Remarks on "measuring ambiguity in the evidence theory", IEEE Trans. Syst. Man Cybern., Part A, Syst. Hum. 38 (2008) 995–999.
[44] X. Deng, Analyzing the monotonicity of belief interval based uncertainty measures in belief function theory, Int. J. Intell. Syst. 33 (2018) 1869–1879.
[45] J. Abellán, É. Bossé, Critique of recent uncertainty measures developed under the evidence theory and belief intervals, IEEE Trans. Syst. Man Cybern. Syst. 50 (2017) 1186–1192.
[46] S. Moral-García, J. Abellán, Required mathematical properties and behaviors of uncertainty measures on belief intervals, Int. J. Intell. Syst. 36 (2021).
[47] J. Dezert, A. Tchamova, On the effectiveness of measures of uncertainty of basic belief assignments, Inf. Secur. 52 (2022) 9–36.
[48] T. Burger, S. Destercke, How to randomly generate mass functions, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 21 (2013) 645–673.
[49] A.-L. Jousselme, P. Maupin, Distances in evidence theory: comprehensive survey and generalizations, Int. J. Approx. Reason. 53 (2012) 118–145.
[50] J.H. Ward Jr, Hierarchical grouping to optimize an objective function, J. Am. Stat. Assoc. 58 (1963) 236–244.