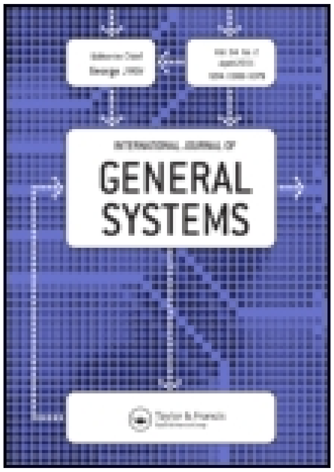

International Journal of General Systems
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ggen20

Requirements for total uncertainty measures in Dempster-Shafer theory of evidence

Joaquín Abellán [a] & Andrés Masegosa [a]

[a] Department of Computer Science and Artificial Intelligence , University of Granada , Granada, Spain

Published online: 03 Nov 2008.

To cite this article: Joaquín Abellán & Andrés Masegosa (2008) Requirements for total uncertainty measures in Dempster-Shafer theory of evidence, International Journal of General Systems, 37:6, 733-747, DOI: 10.1080/03081070802082486

To link to this article: http://dx.doi.org/10.1080/03081070802082486

International Journal of General Systems
Vol. 37, No. 6, December 2008, 733–747

Requirements for total uncertainty measures in Dempster–Shafer theory of evidence

Joaquín Abellán* and Andrés Masegosa

Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain

(*Received 9 October 2007; final version received 6 March 2008*)

Recently, an alternative measure of total uncertainty in Dempster–Shafer theory of evidence (DST) has been proposed in place of the maximum entropy measure. It is based on the pignistic probability of a basic probability assignment and it is proved that this measure verifies a set of needed properties for such a type of measure. The proposed measure is motivated by the problems that maximum (upper) entropy has. In this paper, we analyse the requirements, presented in the literature, for total uncertainty measures in DST and the shortcomings found on them. We extend the set of requirements, which we consider as a set of requirements of properties, and we use the set of shortcomings found on them to define a set of requirements of the behaviour for total uncertainty measures in DST. We present the differences of the principal total uncertainty measures presented in DST taking into account their properties and behaviour.

Also, an experimental comparative study of the performance of total uncertainty measures in DST on a special type of belief decision trees is presented.

Keywords: imprecise probabilities; theory of evidence; uncertainty based information; total uncertainty; conflict; non-specificity

1. Introduction

In the classical theory of probability, Shannon's entropy (Shannon 1948) is the tool used for quantifying uncertainty. Its main virtue is that it verifies a set of desirable properties for probability distributions. In situations where the probabilistic representation is inadequate, an imprecise probability theory can be used as seen in Walley (1991), such as Dempster–Shafer's theory (DST) (Dempster 1967, Shafer 1976), interval-valued probabilities (Campos *et al.* 1994), order-two capacities (Choquet 1953/54), upper-lower probabilities (Suppes 1974, Fine 1983) or general convex sets of probability distributions (Good 1962, Levi 1980, Walley 1991), also called credal sets. In order to quantify the uncertainty represented by these situations, Shannon's entropy has been used as the starting point. It can be justified in different ways, but the most common one is the axiomatic approach, i.e. by assuming a set of necessary basic properties that a measure must verify (Klir and Wierman 1998).

In Dempster–Shafer's theory (DST), Yager (1983) distinguishes between two types of uncertainty: conflict (or randomness or discord) and non-specificity. A total uncertainty measure is also justified in this theory by an axiomatic approach considering the one used in probability theory as a reference. We believe that some aspects of DST that do not

*Corresponding author. Email: {jabellan,andrew}@decsai.ugr.es

ISSN 0308-1079 print/ISSN 1563-5104 online

DOI: 10.1080/03081070802082486
http://www.informaworld.com

appear in classical probability theory, such as monotonicity, should be taken into account when studying the axiomatic approach.

Maeda and Ichihashi (1993) proposed a total uncertainty measure on DST adding up the generalised Hartley measure and upper entropy. They proved that this total uncertainty measure verifies all the necessary basic properties except for the required range. This property could, however, be discussed since there are more types of uncertainty in DST than in the probability theory.

DST is considered as a particular theory of credal sets. By applying the uncertainty invariance principle, a total uncertainty measure on general credal sets will be a generalisation of a total uncertainty measure on DST. With this aim, various studies of the quantification of uncertainty on credal sets have been published (Abellán and Moral 2003a, 2005b; Abellán *et al.* 2006).

In Abellán and Moral (2003a), it is proved that the maximum of Shannon's entropy (upper entropy) verifies on general credal sets all the basics properties that it verifies on DST. In Abellán and Moral (2005a) and Klir and Smith (2001), the use of maximum entropy on credal sets as a good measure of total uncertainty is justified. The problem lies in separating these functions into others, which really do measure the conflict and non-specificity parts by using a credal set to represent the information.

In DST, we have two total uncertainty measures that verify a set of basic required properties: Maeda and Ichihashi's total uncertainty measure and upper entropy. More recently, however, Jousselme *et al.* (2006) presented a new total uncertainty measure in DST based on the pignistic distribution. The authors proved that this measure verifies the necessary properties and it resolves other shortcomings of upper entropy. We therefore have three total uncertainty measures in DST verifying all the required properties, and we will study these measures in this paper.

In this paper, we justify an extension of the set of required properties for a total uncertainty measure on DST and we refer to this set as the requirements of properties. We present a comparative study of the properties verified for each total uncertainty measure, and in doing so, we will see that Jousselme *et al.*'s total uncertainty measure has some undesirable defects.

We will also analyse the shortcomings reported for the upper entropy by certain authors (Jousselme *et al.* 2006). In order to do so, we will revise the set of requirements of behaviour that a total uncertainty measure in DST must verify and study these requirements on the most significant total uncertainty measures defined in DST. By considering the new results on upper entropy, we will see that the upper entropy behaves correctly, with the criticisms being not totally justified.

One important aspect of total uncertainty measures in DST is their applicability, and this involves a not too complicated calculation. In Appendix A of this paper, we present an application of these measures, having conducted an experimental study of these total uncertainty measures on a special type of belief decision trees (Abellán and Moral 2003b, 2005a), i.e. decision trees where the DST is used to represent the information expressed by a database on a query variable. In this procedure, the way to quantify the information plays an important role in the success obtained. In our experimentation, we will use the different total uncertainty measures analysed in this paper as tools to quantify the information.

In Section 2, we will introduce some necessary basic concepts and notation. Section 3 presents the extended set of basic properties verified by each total uncertainty measure. In Section 4, we analyse the shortcomings that total uncertainty measures present. Starting with this set of shortcomings, we will define a set of requirements of behaviour for this type of measure. Section 5 discusses our conclusions.

2. Previous concepts

2.1 Dempster–Shafer theory of evidence

Let X be a finite set considered as a set of possible situations, $|X| = n$, $\wp(X)$ the power set of X and x any element in X. The Dempster–Shafer theory is based on the concept of basic probability assignment. A basic probability assignment (b.p.a.), also called a mass assignment, is a mapping $m : \wp(X) \rightarrow [0, 1]$, such that $m(\emptyset) = 0$ and $\sum_{A \subseteq X} m(A) = 1$. A set A where $m(A) > 0$ is called a focal element of m.

Let X, Y be finite sets. Considering the product space of the possible situation $X \times Y$ and m a b.p.a. on $X \times Y$, the marginal b.p.a. on X, m_x and similarly on Y, m_Y is defined in the following way:

$$m_X(A) = \sum_{R|A=R_X} m(R), \quad \forall A \subseteq X;$$

where R_x is the set projection of R on X.

There are two functions associated with each basic probability assignment: a belief function, Bel, and a plausibility function, Pl: $Bel(A) = \sum_{B \subseteq A} m(B)$, $Pl(A) = \sum_{A \cap B \neq \emptyset} m(B)$. These can be seen as the lower and upper probability of A, respectively.

We may note that belief and plausibility functions are inter-related for all $A \in \wp(X)$, by $Pl(A) = 1 - Bel(A^c)$, where A^c denotes the complement of A. Furthermore, $Bel(A) \leq Pl(A)$.

2.2 Uncertainty in DST

The classical measure of entropy (Shannon 1948) on probability theory is defined by the following continuous function: $S(p) = -\sum_{x \in X} p(x) \log_2(p(x))$, where $p = (p(x))_{x \in X}$ is a probability distribution on X, $p(x)$ is the probability of value x and $\log_2$ is normally used to quantify the value in bits[1]. The value[2] $S(p)$ quantifies the only type of uncertainty presented on probability theory and it verifies a large set of desirable properties (Shannon 1948, Klir and Wierman 1998).

In DST, Yager (1983) distinguishes between two types of uncertainty: the first is associated with cases where the information focuses on sets with empty intersections, and the second is associated with cases where the information focuses on sets with greater-than-one cardinality. These are called conflict (or randomness or discord) and non-specificity, respectively.

The following function, introduced by Dubois and Prade (1984), has its origin in classical Hartley measure (Hartley 1928) on classical set theory and in the extended Hartley measure on possibility theory (Higashi and Klir 1983). It represents a measure of non-specificity associated with a b.p.a. It is expressed as follows: $I(m) = \sum_{A \subseteq X} m(A) \log(|A|)$.

$I(m)$ attains its minimum, zero, when m is a probability distribution. The maximum, $\log(|X|)$, is obtained for a b.p.a., m, with $m(X) = 1$ and $m(A) = 0$, $\forall A \subset X$.

Many measures were introduced to quantify the conflict degree that a b.p.a. represents (Klir and Wierman 1998). One of the most representative conflict functions was introduced by Yager (1983):

$$E(m) = -\sum_{A \subseteq X} m(A) \log Pl(A).$$

This function, however, does not verify all the required properties on DST.

Harmanec and Klir (1996) proposed the measure $S^*(m)$ which is equal to the maximum of the entropy (upper entropy) of the probability distributions verifying $Bel(A) \leq \sum_{x \in A} p(x) \leq Pl(A)$, $\forall A \subseteq X$. This set of probability distributions is the credal set associated with a b.p.a. m, and will be denoted as K_m.

Maeda and Ichihashi (1993) proposed a total uncertainty measure using the above measures which quantifies the conflict and non-specificity contained in a b.p.a. on X in the following way:

$$MI(m) = I(m) + S^*(m),$$

where $I(m)$ is used as a non-specificity function and $S^*(m)$ is used as a measure of conflict. This measure was analysed in Abellán and Moral (1999).

Harmanec and Klir (1996) proposed S^* as a total uncertainty measure in DST, i.e. as a measure that quantifies conflict and non-specificity, but they do not separate this into parts that quantify these two types of uncertainty on DST. More recently, Abellán *et al.* (2006) proposed upper entropy as an aggregate measure on more general theories than DST, coherently separating conflict and non-specificity. These parts can also be obtained in DST in a similar way. In DST, we can consider

$$S^*(m) = S_*(m) + (S^* - S_*)(m),$$

where $S^*(m)$ represents maximum entropy and $S_*(m)$ represents minimum entropy on the credal K_m associated to a b.p.a. m, with $S_*(m)$ coherently quantifying the conflict part and $(S^* - S_*)(m)$ its non-specificity part.

Quite recently, Jousselme *et al.* (2006) presented a measure to quantify ambiguity (discord or conflict and non-specificity) in DST, i.e. a total uncertainty measure on DST. This measure is based on the pignistic distribution on DST: let m be a b.p.a. on a finite set X, then the pignistic probability distribution $BetP_m$, on all the subsets A in X is defined by

$$BetP_m(A) = \sum_{B \subseteq X} m(B) \frac{|A \cap B|}{|B|}.$$

For a singleton set $A = \{x\}$, we have $BetP_m(\{x\}) = \sum_{x \in B} [m(B)/|B|]$. Therefore, the ambiguity measure for a b.p.a. m on a finite set X is defined as

$$AM(m) = -\sum_{x \in X} BetP_m(x) \log (BetP_m(x)).$$

3. Basic properties of total uncertainty measures in DST

In Klir and Wierman (1998), we can find five requirements for a total uncertainty measure (TU) in DST, i.e. for a measure which captures both conflict and non-specificity. Using the above notation, these requirements can be expressed in the following way:

(P1) *Probabilistic consistency:* when all the focal elements of a b.p.a. m are singletons, then a total uncertainty measure must be equal to the Shannon entropy:

$$TU(m) = \sum_{x \in X} m(x) \log m(x).$$

(P2) *Set consistency:* when a set A exists such that $m(A) = 1$, then a TU must collapse to the Hartley measure:

$$TU(m) = \log |A|.$$

(P3) *Range:* the range of $TU(m)$ is $[0, \log|X|]$.
(P4) *Subadditivity:* let m be a b.p.a. on the space $X \times Y$, m_X and m_Y its marginal b.p.a.s on X and Y, respectively, then a TU must satisfy the following inequality:

$$TU(m) \leq TU(m_X) + TU(m_Y).$$

(P5) *Additivity:* let m be a b.p.a. on the space $X \times Y$, m_X and m_Y its marginal b.p.a.s on X and Y, respectively, such that these marginals are not interactive ($m(A \times B) = m_X(A)m_Y(B)$, with $A \subseteq X$, $B \subseteq Y$ and $m(C) = 0$ if $C \neq A \times B$), then a TU must satisfy the equality

$$TU(m) = TU(m_X) + TU(m_Y).$$

With these requirements, we hope to extend those of Shannon entropy in probability theory, although there are situations in DST that can never appear in probability theory. For instance, a probability distribution can never contain another probability distribution. In DST, however, the information of a b.p.a. can be contained by the information of another b.p.a. Let us consider the following example.

Example 1. In a first situation, we have three pieces of evidence (e_1 e_2 and e_3) about the type of disease (d_1, d_2 or d_3), which a patient has. Hence, an expert quantifies the information available using a basic probability assignment and considers the following b.p.a. on the universal $X = \{d_1, d_2, d_3\}$:

$$e_1 \longrightarrow m_1(\{d_1, d_2\}) = \frac{1}{3},$$

$$e_2 \longrightarrow m_1(\{d_1, d_3\}) = \frac{1}{2},$$

$$e_3 \longrightarrow m_1(\{d_2, d_3\}) = \frac{1}{6}.$$

Let us now assume that the expert finds that the reasons for discarding d_3 in e_1 are false and that it is necessary to change his b.p.a. to the following:

$$e_1 \longrightarrow m_2(\{d_1, d_2, d_3\}) = \frac{1}{3},$$

$$e_2 \longrightarrow m_2(\{d_1, d_3\}) = \frac{1}{2},$$

$$e_3 \longrightarrow m_2(\{d_2, d_3\}) = \frac{1}{6}.$$

In the above example, we go from a first situation with one amount of information to another more confused situation. It is logical to consider that the second situation involves a greater level of uncertainty (minor information). Here, we have $Bel_2(A) \leq Bel_1(A)$ and

$Pl_1(A) \leq Pl_2(A)$, $\forall A \subseteq X$; implying a larger level of uncertainty for m_2. This also implies that $K_{m_1} \subseteq K_{m_2}$, where K_{m_1} and K_{m_2} are the credal sets associated to m_1 and m_2, respectively.

We consider that the situation expressed by Example 1 should be taken into account for a total uncertainty measure in DST. This situation allows us to consider the following property:

> *(P6) Monotonicity:* a total uncertainty measure in DST must not decrease the total quantity of uncertainty in a situation where a clear decrease in information (increment of uncertainty) is produced.

Formally, let two b.p.a.s be on a finite set X, m_1 and m_2, verifying that $K_{m_1} \subseteq K_{m_2}$, then $TU(m_1) \leq TU(m_2)$.

Here, we must remark that the monotone dispensability definition of Harmanec (1995) could be used, but we prefer a more general one, which can be extended in a direct way to general credal sets. Monotone dispensability always implies monotonicity axiom but not the contrary, as can be easily checked[3].

If we use the results of the works of Klir and Wierman (1998), Maeda and Ichihashi (1993) and Jousselme *et al.* (2006), it can checked that the MI, S^* and AM functions verify the following sets of requirements in DST:

MI: P1, P4, P5 and P6.
S^*: P1, P2, P3, P4, P5 and P6.
AM: P1, P2, P3 and P5.

Considering the list above, we should mention the following:

1. Function MI does not satisfy the P2 and P3 requirements. Its range is $[0, 2\log|X|]$ because it uses a clear split between the quantification of the two types of uncertainty[4], each with range $[0, \log|X|]$.
2. Jousselme *et al.* proved that function AM satisfies the P4 requirement, but recently, Klir and Lewis (2007) found an error in this proof and gave a counter example that proves that AM does not satisfy the P4 requirement.
3. Function AM does not satisfy the P6 requirement. If we consider Example 1, it can be proved that $K_{m_1} \subseteq K_{m_2}$, and we have

$$BetP_{m_1}(\{d_1\}) = \frac{5}{12}, \quad BetP_{m_1}(\{d_2\}) = \frac{3}{12}, \quad BetP_{m_1}(\{d_3\}) = \frac{4}{12};$$

$$BetP_{m_2}(\{d_1\}) = \frac{13}{36}, \quad BetP_{m_2}(\{d_2\}) = \frac{7}{36}, \quad BetP_{m_2}(\{d_3\}) = \frac{16}{36}.$$

Hence,

$$AM(m_1) = 1.078 > AM(m_2) = 1.047.$$

We can therefore see that only S^* satisfies all the proposed requirements.

4. Requirements of behaviour for total uncertainty measures in DST

The paper by Jousselme *et al.* analyses certain shortcomings of the S^* function (upper entropy) in DST in order to compare this function with the AM function. These shortcomings have been presented in publications by Klir *et al.* and can be expressed in the following way:

1. Computing complexity.
2. Concealment of the two types of uncertainty coexisting in the evidence theory: conflict and non-specificity.
3. Insensitivity to changes in evidence.

We consider Klir *et al.*'s considerations about the behaviour of a total uncertainty measure (TU) in DST to be very important, because a TU in DST makes no sense if it verifies all the basic properties (P1–P6) but its calculation is unfeasible. A TU in DST should also give us information about the quantification of the two types of uncertainty coexisting in DST. Finally, a TU should be sensitive to changes in evidence directly or via its parts of conflict or non-specificity since it is possible for an increase in conflict to cause a decrease in non-specificity, and vice versa, and we could have two situations with similar total uncertainty values but with different conflict and non-specificity parts of the uncertainty.

This set of shortcomings found in certain total uncertainty measures could therefore be used to present a set of requirements of behaviour of a TU in DST. P1–P6 can be considered as requirements of properties that a total uncertainty measure in DST must satisfy. The set of requirements of behaviour (RB) of a TU in DST could be expressed in the following way:

(RB1) The calculation of a TU should not be too complex.
(RB2) A TU must not conceal the two types of uncertainty (conflict and non-specificity) co-existing in the evidence theory.
(RB3) A TU must be sensitive to changes in evidence either directly or via its parts of conflict and non-specificity.

There are certain situations where the information available is more suitable for being mathematically quantified with more general models than the DST. In such cases, we are talking about a 'Generalised Information Theory' [see Klir (2006)] and take into account Klir's following principle of Klir:

> *Principle of uncertainty invariance*: the amount of uncertainty (and information) must be preserved when a representation of uncertainty in one mathematical theory is transformed into its counterpart in another theory. That is, the principle guarantees that no information is unwittingly added or eliminated solely by changing the mathematical framework by which a particular phenomenon is formalised.

By this principle, a TU in DST should allow us to extend it on more general theories than DST. This one could be considered as a requirement of behaviour for a TU in DST, that we can call extensibility:

(RB4) The extension of a TU in DST on more general theories must be possible.

We will now review the above requirements of behaviour for functions *MI*, S^* and *AM*:

RB1: As we can see in Jousselme *et al.* (2006), the *AM* function has a simpler calculation than the other functions (*MI* includes S^* in its definition), and it is only necessary to obtain the pignistic probability distribution of a b.p.a. The calculation of S^* in DST has a high computational complexity. Meyerowitz *et al.*'s algorithm (1994) was the first to obtain this value. More recently, the computation of this algorithm was reduced by

Liu *et al.* (2007). Although the computational cost of every TU in DST is clearly different, the calculation of every TU in DST is simple.

RB2*: MI* can be separated coherently in conflict and non-specificity by definition. Here, S^* is used as a conflict measure and function I as a non-specificity measure. Recently, Abellán *et al.* (2006) separated S^* into two parts that coherently quantify conflict and non-specificity for more general theories than DST. Here, S_* (minimum of entropy) is used as a conflict measure and $S^* - S_*$ is used as a non-specificity function. In order to obtain these parts, Abellán and Moral (2005b) present a branch and bound algorithm to obtain S_* on more general theories than DST. Only the *AM* function has no clear separation between conflict and non-specificity. In Jousselme *et al.* (2006), *AM* is presented as a special case of the function $\delta S^* + (1 - \delta)I$, for an unknown $\delta \in (0, 1)$. Therefore, when value $AM(m)$ is used, it is impossible to know what quantity corresponds to conflict and what to non-specificity.

RB3*:* In this point, we want to review the analysis presented in Jousselme *et al.* (2006) about the sensitivity of S^* in DST using an example by Klir and Smith (1999):

Example 2. Suppose there are two elements in the given frame of discernment $X = \{1, 2\}$, and we know $m(\{1\}) = m_1$, $m = (\{2\}) = \mathrm{m}_2$, so $m(\{1, 2\}) = m_{12} = 1 - m_1 - m_2$. At this point, we should mention that by definition (Yager 1983), the non-specificity part of m depends only on the m_{12} value and the conflict part of m depends on the interaction between m_1 and m_2 values.

In Klir and Smith's original example (1999), S^* was identified as 'highly insensitive to changes in evidence', an 'unsatisfactory situation'. S^* gives the same value for all bodies of evidence for which both m_1 and m_2 fall into the range [0, 0.5]. When $m_2 \subseteq [0.5, 1]$, the S^* measure is entirely independent of the value of m_1 and vice versa.

Jousselme *et al.* (2006) proved that the *AM* measure does not behave in the same way and *AM* is neither independent of m_1 or m_2 by considering the above example. We will use the example and will apply it to every TU considered in this paper. Without loss of generality, it is supposed that m_1 is known. They then consider that three cases appear: $m_1 > 0.5$, $m_1 = 0.5$, and $m_1 < 0.5$.

(1) $m_1 > 0.5$, e.g. $m_1 = 0.6$. Here, $m_{12} = 0.4 - m_2$. We have

$$S^*(m) = S(0.6,\ 0.4); \quad S_*(m) = S(1 - m_2, m_2);$$

$$I(m) = (0.4 - m_2)\log 2;$$

$$AM(m) = S\Big(0.8 - \frac{m_2}{2}, \quad 0.2 + \frac{m_2}{2}\Big).$$

MI: The conflict part of this function (S^*) is constant, and does not vary when m_2 changes. No distinction is made between the values of m_2 and this does not make any sense because of the definition of the conflict part of a b.p.a.

S^*: The variations of the conflict part of this function (S_*) make sense; if m_2 increases then so does the conflict part. Similarly, the non-specificity part behaves in a similar way: a decrease in m_2 leads to an increase in m_{12} and as we can see via its non-specificity part

$$S^* - S_* = S(0.6,\ 0.4) - S(1 - m_2, m_2),$$

it results in an increase in non-specificity. We therefore consider that the constant value S^* as a total uncertainty value makes sense because in this example with $X = \{1, 2\}$, an increase in conflict can result in a decrease in non-specificity, and vice versa, which can be compensated mutually.

AM: This function is not constant and depends on the m_2 values. This situation makes sense but we do not know what happens to the variations of conflict or non-specificity when we modify m_2 value.

(2) $m_1 = 0.5$ and $m_{12} = 0.5\,m_2$. We have:

$$S^*(m) = S(0.5,\ 0.5); \quad S_*(m) = S(1 - m_2, m_2);$$

$$I(m) = (0.5 - m_2)\log 2;$$

$$AM(m) = S\Big(0.75 - \frac{m_2}{2}, \quad 0.25 + \frac{m_2}{2}\Big).$$

In this case, we can observe a similar situation to that in Case (1).

(3) $m_1 < 0.5$, e.g. $m_1 = 0.2$. Here, $m_{12} = 0.8 - m_2$. We have:

$$S^*(m) = \begin{cases} S(0.5, 0.5) \text{ if } m_2 < 0.5, \\ \\ S(1 - m_2, m_2) \text{ if } M_2 \geq 0.5. \end{cases}$$

$$S_*(m) = \begin{cases} S(\alpha, 1 - \alpha) \text{ if } m_2 < 0.5 \ (\alpha = \min\{m_1, m_2\}), \\ \\ S(0.2, 0.8) \text{ if } M_2 \geq 0.5. \end{cases}$$

$$I(m) = (0.8 - m_2)\log 2;$$

$$AM(m) = S\Big(0.8 - \frac{m_2}{2}, \quad 0.2 + \frac{m_2}{2}\Big).$$

MI: If $m_2 > 0.5$, the conflict part of this function (S^*) is constant and does not vary when there is a change in m_2. It does distinguish between the values of m_2 and this makes more sense because of the definition of the conflict part of a b.p.a. than in Situation (1). However, if $m_2 \geq 0.5$, the conflict part depends on m_2, and here *AM* makes more sense than in Case (1).

S^*: Here, S^* is constant when $m_2 < 0.5$ and it depends on m_2 (or m_{12}) when $m_2 \geq 0.5$. Analysing the parts of S^*, the variations of the conflict part of this function (S_*) make sense. It depends on the minimum value of m_1 and m_2, called α. Studying the non-specificity part is more complex. We must now take into account the following situations: (3.1) $m_2 \geq 0.5$; (3.2) $m_1 \leq m_2 < 0.5$ and (3.3) $m_2 \leq m_1 < 0.5$. In (3.1) and (3.3), we have similar results as in Case (1). In (3.2),

$$S^* - S_* = S(0.5, 0.5) - S(0.8, 0.2),$$

which is a constant value. If m_2 decreases *by a quantity* ϵ, with $0 < \epsilon < m_2 - m_1$, we obtain the same value of *non-specificity*. This could be considered to be an 'unsatisfactory situation'. However, we think that here it makes sense to obtain a maximum total uncertainty value (log 2) because we always have that $1 - m_i > 0.5$ for $i = 1,2$; i.e. the plausibility value for each element is always greater than 0.5 in this example[5].

AM: There is a similar situation to the one in Case (1).
***RB4**:* Using Abellán and Moral (2000, 2003a) results and those of Abellán *et al.* (2006), *MI* and S^* can be extended on more general theories than DST, verifying similar sets of properties[6]. The *I* function verifies the subadditivity property for all the theories with the exception of general credal sets [see Abellán and Moral (2005a)]. The extension of the *AM* function on more general theories is an open question. One possibility for this extension could be to use the Möbius transform as for the *I* function [see Abellán and Moral (2000)], although its calculation would be more complex. As we can see in Abellán *et al.* (2006) with the *E* conflict measure in DST, the generalisation of certain uncertainty measures defined in DST could have many problems when we want to extend them on more general theories. However, this is still an open problem for the *AM* function.

5. Conclusions

We have analysed the set of requirements of properties for total uncertainty measures in DST and consider it necessary to extend them with a monotonicity property. Considering this new set, we show that upper entropy is the only one that verifies all the property requirements. The new total uncertainty measure *AM* does not verify certain important properties. It is not subadditive and it does not verify the monotonicity property and this produces incoherent situations.

Taking into account the set of defects found in the behaviour of upper entropy in DST, we have defined a set of desirable behaviours that a total uncertainty measure in DST must verify and we have called this the set of behaviour requirements. In future work, it would be interesting to analyse the possible extension of this set of behaviour requirements in greater depth.

In view of the results obtained for upper entropy, where there is a coherent separation between conflict and non-specificity (Abellán *et al.* 2006), we present a new way of verifying this set of behaviour requirements. We have shown the performance of *MI*, S^* and *AM* on this set and have observed that Maeda and Ichihashi's *MI* function uses a separation of conflict and non-specificity which is not entirely satisfactory, while Jousselme *et al.*'s *AM* function conceals the two types of uncertainty in DST, has a better computing complexity than the other functions and has not yet been defined on more general theories. We have shown that by splitting upper entropy into conflict and non-specificity, this function is sensitive to evidence changes. In the examples normally used to query upper entropy, we have shown that an increase in conflict results in a decrease in non-specificity, and vice versa, and this implies a constant value of upper entropy in a coherent way.

Since applicability is an important aspect of the measures studied here, in Appendix A we present an application of TU in DST on a special type of belief decision trees. From the experiments conducted, we can see that upper entropy is a clear winner with respect to the other TUs in DST.

Acknowledgement

This work has been supported by the Spanish Ministry of Science and Technology under the Projects TIN2007-67418-C03-03 and TIN2005-02516. A shortened version of this paper was presented at NAFIPS'08.

Notes

1. log and $\log_2$ are used indifferently in the literature for this aim. In this paper, we will use log.
2. We will use the notation $S(p)$ or $S(p)(x_1), p(x_2), \ldots)$ indifferently.
3. It can be proved that if a b.p.a. m is contained by monotonicity axiom into another b.p.a. m', we can define a set of b.p.a.s $\{m_i | i = 1, \ldots, k\}$ such that $m_1 = m$ and $m_k = \mathrm{m}'$ and m_i is contained by monotone dispensability into m_{i+1}.
4. If we want to quantify two types of uncertainty (one more than in the probability theory), perhaps the range requirement should be extended as suggested in existing literature on total uncertainty measures in DST. This is a question that requires further reflection. Function $N = \frac{1}{2}MI$ satisfies P1 $-$ P6 requirements.
5. This reasoning is compatible with the principle of maximum uncertainty [see Klir (2006)].
6. Within these theories algorithms do exist for obtaining S^* and S_*, as we can see in Abellán and Moral 2005b, 2005c).
7. We have repeated the experiments with this discretisation method and similar results have been obtained; only *NB* obtains slightly better results.

Notes on contributors

Joaquín Abellán received his PhD degree in January of 2003 from the University of Granada. He is an associate professor of Computer Science and Artificial Intelligence and member of the research group on Uncertainty Treatment on Artificial Intelligence (UTAI) at the University of Granada. His current research interests are representation of information via imprecise probabilities, definition of information/uncertainty measures on imprecise probabilities and their applications, principally in data mining.

Andrés Masegosa received his Computer Engineering degree in December of 2003 at the University of Granada. He is a PhD student in the Computer Science and Artificial Intelligence department and member of the research group on Uncertainty Treatment on Artificial Intelligence (UTAI) at the University of Granada. His PhD interests include data mining, probabilistic graphical models, imprecise probabilities and their applications, especially in information retrieval problems.

References

Abellán, J., 2006. Uncertainty measures on probability intervals from Imprecise Dirichlet model. *International journal of general systems*, 35 (5), 509–528.

Abellán, J., Klir, G.J. and Moral, S., 2006. Disaggregated total uncertainty measure for credal sets. *International journal of general systems*, 35 (1), 29–44.

Abellán, J. and Moral, S., 1999. Completing a total uncertainty measure in Dempster–Shafer theory. *International journal of general systems*, 28, 299–314.

Abellán, J. and Moral, S., 2000. A non-specificity measure for convex sets of probability distributions. *International journal of uncertainty, fuzziness and knowledge-based systems*, 8, 357–367.

Abellán, J. and Moral, S., 2003a. Maximum of entropy for credal sets. *International journal of uncertainty, fuzziness and knowledge-based systems*, 18, 1215–1225.

Abellán, J. and Moral, S., 2003b. Using the total uncertainty criterion for building classification trees. *International journal of intelligent systems*, 18 (12), 1215–1225.

Abellán, J. and Moral, S., 2005a. Upper entropy of credal sets. Applications to credal classification. *International journal of approximate reasoning*, 39, 235–255.

Abellán, J. and Moral, S., 2005b. Difference of entropies as a non-specificty measure for credal sets. *International journal of general systems*, 34 (3), 201–214.

Abellán, J. and Moral, S., 2005c. An algorithm that computes the upper entropy for order-2 capacities. *International journal of uncertainty, fuzziness and knowledge-based systems*, 14 (2), 141–154.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J., 1984. *Classification and regression trees*. Belmont: Wadsworth Statistics, Probability Series.

Campos, L.M., de Huete, J.F. and Moral, S., 1994. Probability intervals: A tool for uncertainty reasoning. *International journal of uncertainty, fuzziness and knowledge-based system*, 2, 167–196.

Choquet, G., 1953/54. Théorie des Capacités. *Annal institute Fourier*, 5, 131–292.

Dempster, A.P., 1967. Upper and lower probabilities induced by a multivaluated mapping. *Annal mathematics statistic*, 38, 325–339.

Demsar, J., 2006. Statistical comparison of classifiers over multiple data sets. *Journal of machine learning research*, 7, 1–30.

Dubois, D. and Prade, H., 1984. A note on measure of specificity for fuzzy sets. *BUSEFAL*, 19, 83–89.

Duda, R.O. and Hart, P.E., 1973. *Pattern classification and scene analysis*. New York: John Wiley and Sons.

Fayyad, U.M. and Irani, K.B., 1993. Multi-valued interval discretization of continuous-valued attributes for classification learning. In: *Proceedings of the 13th international joint conference on artificial intelligence*. San Mateo: Morgan Kaufmann, 1022–1027.

Fine, T.L., 1983. Foundations of probability. In: S. Kotz, C. Read and D. Banks, eds. *Encyclopedia of Statistical Sciences*, Vol. 3. New York: Wiley, 246–254.

Good, I.J., 1962. Subjetive probability as the measure of a non-measurable set. In: E. Nagel, P. Suppes, and A. Tarski, eds. *Login, methodology and philosophy of science*. California: Stanford University Press, 319–329.

Harmanec, D., 1995. Toward a characterization of uncertainty measure for the Dempster–Shafer theory. In: P. Besnard and S. Hanks, eds. *Proceedings of the 11th annual conference on uncertainty in artificial intelligence*. San Mateo, CA: Morgan Kaufmann, 255–261.

Harmanec, D. and Klir, G.J., 1994. Measuring total uncertainty in Dempster–Shafer theory, a novel approach. *International journal of general system*, 22, 405–419.

Hartley, R.V.L., 1928. Transmission of information. *The bell systems technical journal*, 7, 535–563.

Higashi, M. and Klir, G.J., 1983. Measures of uncertainty and information based on possibility distributions. *International journal of general system*, 9, 43–58.

Jousselme, A.L., Liu, C., Grenier, D. and Boss, E., 2006. Measuring ambiguity in the evidence theory. *IEEE Transactions on systems, man, and cybernetics. Part A: systems and humans*, 36 (5), 890–903.

Klir, G.J., 2006. *Uncertainty and information, foundations of generalized information theory*. Hoboken, NJ: John Wiley.

Klir, G.J. and Lewis, H.W., 2008. Remarks on 'measuring ambiguity in the evidence theory'. *IEEE transactions on systems, man and cybernetics. Part A: systems and humans*, 38, 995–999.

Klir, G.J. and Smith, R.M., 1999. Recent developments in generalized information theory. *International journal of fuzzy systems of mathematics and artificial intelligence*, 1 (1), 1–13.

Klir, G.J. and Smith, R.M., 2001. On measuring uncertainty and uncertainty-based information, recent developments. *Annals of mathematics and artificial intelligence*, 32, 5–33.

Klir, G.J. and Wierman, M.J., 1998. *Uncertainty-based information*. New York: Phisica-Verlag.

Levi, I., 1980. *The enterprise of knowledge*. London: NIT Press.

Liu, C., Grenier, D., Jousselme, A.-L. and Bosse, E., 2007. Reducing algorithm complexity for computing an aggregate uncertainty measure. *IEEE transactions on systems, man and cybernetics. Part A: systems and humans*, 37 (5), 669–679.

Maeda, Y. and Ichihashi, H., 1993. A uncertainty measure with monotonicity under the random set inclusion. *International journal of general systems*, 21, 379–392.

Meyerowitz, A., Richman, F. and Walker, E.A., 1994. Calculating maximum-entropy probabilities densities for belief functions. *International journal of uncertainty, fuzziness and knowledge-based systems*, 2, 377–389.

Nadeau, C. and Bengio, Y., 2001. Inference for the generalization error. *Machine learning*.

Quinlan, J.R., 1986. Induction of decision trees. *Machine learning*, 1, 81–106.

Shafer, G., 1976. *A mathematical theory of evidence*. Princeton: Princeton University Press.

Shannon, C.E., 1948. A mathematical theory of communication. *The bell system technical journal*, 27, 379–423, 623–656.

Suppes, P., 1974. The measurement of belief (with discussion). *Journal of the royal statistical society B*, 36, 160–191.

Walley, P., 1991. *Statistical reasoning with imprecise probabilities*. New York: Chapman and Hall.

Walley, P., 1996. Inferences from multinomial data: Learning about a bag of marbles. *Journal of the royal statistical society B*, 58, 3–57.

Wilcoxon, F., 1945. Individual comparison by ranking methods. *Biometrics*, 1, 80–83.

Witten, I.H. and Frank, E., 2005. *Data mining, practical machine learning tools and techniques*. 2nd ed. San Francisco: Morgan Kaufmann.

Yager, R.R., 1983. Entropy and specificity in a mathematical theory of evidence. *International journal of general systems*, 9, 249–260.

Appendix A: Application of total uncertainty measures on belief decision trees

It is possible to apply classification methods on a given database containing several samples where each sample also contains a set of values belonging to an attribute or predictive variable set and a variable labelled class variable. In the field of machine learning, the classification subject is based on the use of several techniques that infer rules from a given data set in order to predict new values of the class variable (discrete or discretised) using a new set of values for the remaining variables (known as attribute variables). The data set used to obtain these rules is labelled the training data set and the data set used to check the classifier is called the test data set. Classification has important and distinguished applications in the fields of medicine, bioinformatics, physics, pattern recognition, economics, etc., and is used for disease diagnosis, meteorological forecasts, insurance, text classification, to name but a few.

In our case, we use the imprecise probability theory (Walley 1991) to construct a classification tree. This structure is easy to understand and an efficient classifier and originates in Quinlan's ID3 algorithm (Quinlan 1986). As a basic reference, we should mention the book by Breiman *et al.* (1984). We shall use Walley's imprecise Dirichlet model (1996) (IDM) to estimate the probabilities of membership to the respective classes defined by the class variable and a split criterion based on total uncertainty measures. Abellán (2006) shows how intervals from the IDM can be represented by means of belief functions. We can therefore say that our procedure could be considered to be a belief decision tree.

Belief decision trees and split criteria

A decision tree is a simple structure that can be used as a classifier. Within a decision tree, each node represents an attribute variable and each branch represents one of the states of this variable. A tree leaf specifies the expected value of the class variable depending on the information contained in the training data set. When we obtain a new sample or instance of the test data set, we can make a decision or prediction about the state of the class variable following the path to the tree from the root until a leaf using the sample values and the tree structure. Associated to each node is the most informative variable, which has not already been selected in the path from the root to this node (as long as this variable provides more information than if it had not been included). In this last case, a leaf node is added with the most probable class value for the partition of the data set defined with the configuration given by the path until the tree root.

In the experimentation presented here, we will modify the imprecise info-gain criterion (IIG), i.e. the criterion used to split the sample (called the split criterion) to build decision trees using the procedure in Abellán and Moral's method (2003b). We will use the same procedure (changing the total uncertainty measure used) and we will use MI, S^* and AM measures in each case.

Abellán and Moral's procedure (2003b) uses Walley's IDM (1996), which depends on a hyperparameter s. Higher values of s give a more cautious inference. Walley does not give a definitive recommendation for the value of this parameter but he suggests values between $s = 1$ and $s = 2$.

Experimentation

In order to check the above procedure, we have used a wide and different set of 64 known databases, obtained from the UCI repository of machine learning databases which can be downloaded directly from ftp://ftp.ics.uci.edu/machine-learning-databases. The databases chosen are very different in terms of their sample size, number and type of attribute variables, number of states of the class variable, etc.

For our experimentation, we have used Weka software (Witten and Frank 2005) on Java 1.5, and we have added the necessary methods to build decision trees using the IIG split criterion with each TU measure. We should mention here that all the measures may easily be implemented in a programming language. So as not to extend this appendix excessively, we only present the results obtained with the value $s = 1$ for the IDM. We have repeated the experiments with $s = 1.5$ and $s = 2$ and similar results have been obtained. We have also used a well-known classification method, Duda and Hart's Naive Bayes (*NB*) (1973), to compare our results with each TU measure.

We have applied the following preprocessing: databases with missing values have been replaced with mean values (for continuous variables) and mode (for discrete variables) using Weka's own filters. In the same way, continuous variables have been discretised in five equal frequency intervals. If we use Fayyad and Irani's discretisation method (Fayyad and Irani 1993), a large number of variables have only one state which is the same as removing them[7]. Using equal frequency discretisation is therefore of no benefit to any of the classification methods presented. We note that the preprocessing has been applied using the training set and then translated to the test set. For each database, we have repeated 10 times a k-10-fold cross-validation procedure.

Due to the large number of databases used, in order to compare the results of one classifier with another, we do not present all the percentages of all the methods on each database so as not to extend this appendix unnecessarily. We have applied the following test procedures [see Demsar (2006)] on these percentages:

Number of wins, losses and ties: this counts the number of databases where a classifier method beats, loses to or ties with another classifier method.

Paired t-test: this checks whether the average difference in their performance over the database is significantly different from zero. This type of test can be prone to certain weaknesses as discussed in Demsar (2006). We will use the corrected two-tailed paired *t*-test (Nadeau and Bengio 2001) with a 5% significance level.

Wilcoxon signed-ranks test (Wilcoxon 1945): this is a non-parametric alternative to the paired *t*-test, which ranks the differences in performance of two classifiers for each database, ignoring the signs, and compares the ranks for the positive and negative differences. We will use this test with a 0.05 level of significance.

For the sake of simplicity, we will use the names *MI*, S^* and *AM* to denote the classification procedures which result from applying belief decision trees with the IIG procedure and each of the *MI*, S^* and *AM* functions as TU measures in their split criteria, respectively.

Table 1. Number of wins, ties and losses (W/T/L) of the percentage of correct classification of each classifier with each other one.

	NB	MI	S^*	AM
NB	–	(30/1/33)	(29/1/34)	(29/1/34)
MI	(33/1/30)	–	(37/2/25)	(31/2/31)
S^*	(34/1/29)	(25/2/37)	–	(21/3/40)
AM	(34/1/29)	(31/2/31)	(40/3/21)	–
$W-L$	13	−15	26	−24

Table 2. Number of wins, ties and losses (W/T/L) on the paired *t*-test carried out on the percentage of correct classification of each classifier with each other one with 0.05 level of significance.

	NB	MI	S*	AM
NB	–	(18/24/22)	(22/21/21)	(20/21/23)
MI	(22/24/18)	–	(13/50/1)	(13/45/6)
S^*	(21/21/22)	(1/50/13)	–	(2/58/4)
AM	(23/21/20)	(6/45/13)	(4/58/2)	–
$W-L$	6	-23	15	2

Table 3. Winer in the Wilcoxon signed-ranks test with 0.05 level of significance. 'Tie' expresses that this test presents a tie between the two classifiers. [.] indicates the winer with a level of significance of 0.075.

	NB	MI	S*	AM
NB	–	Tie	Tie	Tie
MI	Tie	–	[S*]	Tie
S^*	Tie	[S^*]	–	S^*
AM	Tie	Tie	S^*	–

In Table 1, therefore, we can see the number of wins, ties and losses (W/T/L) of each procedure with another. For example, the first value (30/1/33) implies that *NB* beats *MI* on 33 databases, ties on 1 database and loses (*MI* wins) on 30 databases. The last row expresses the difference between wins and losses considering one classifier against the rest, i.e. considering all the results of each column. We can appreciate differences in favour of S^* and *NB* with respect to the others. The last row indicates the difference between won and lost (W–L). Here, the value of S^* column is significant in relation to *MI* and *AM*: 26 against -15 and -24.

In Table 2 we can see similar differences with respect to the paired *t*-test carried out in favour of S^* and *NB*, but now the differences are lower. Considering the results of one method in relation to the others presented in Tables 1 and 2, S^* can be considered a clear winner. The last row indicates the difference between won and lost (W–L). Once again, S^* performs best.

In Table 3, we can see that now by using the Wilcoxon test, only S^* obtains a positive result with respect to *AM* with a 0.05 level of significance. It is also worth mentioning that S^* beats *MI* with this test if we increase the level of significance to 0.075.

Finally, in the experimentation conducted, we observed the following points:

- Considering all the tests presented, the belief decision tree which uses S^* in its split criterion is the clear winner, even taking into account the *NB* method.
- There is no a great difference between the results obtained with S^* and *MI* because with the IDM we obtain constant values for the *I* function (see Abellán 2006).
- *AM* can be considered the worst TU to use in our procedure.