A COMPUTATIONALLY EFFICIENT APPROXIMATION OF DEMPSTER-SHAFER THEORY

Frans Voorbraak

Department of Philosophy
University of Utrecht

Logic Group
Preprint Series
No. 35

ILLVSTRA NOS SOL IVSTITIAE

Department of Philosophy
University of Utrecht

A computationally efficient approximation of Dempster-Shafer theory

Frans Voorbraak

Department of Philosophy
University of Utrecht
Heidelberglaan 2
3584 CS Utrecht

abstract

An often mentioned obstacle for the use of Dempster-Shafer theory for the handling of uncertainty in expert systems is the computational complexity of the theory. One cause of this complexity is the fact that in Dempster-Shafer theory the evidence is represented by a belief function which is induced by a basic probability assignment, i.e. a probability measure on the powerset of possible answers to a question, and not by a probability measure on the set of possible answers to a question, like in a Bayesian approach. In this paper, we define a Bayesian approximation of a belief function and show that combining the Bayesian approximations of belief functions is computationally less involving than combining the belief functions themselves, while in many practical applications replacing the belief functions by their Bayesian approximations will not essentially affect the result.

Key words and phrases: expert systems, reasoning with uncertainty, Dempster- Shafer theory, computational efficiency.

Introduction

Recently, Dempster-Shafer theory (also known as evidence theory or theory of belief functions) has received much attention as a promising theory for the handling of uncertain information in expert systems. Its main attractions are the ease with which uncertainty deriving from ignorance is represented and the possibility of combining evidence by Dempster's rule of combination. This latter feature seems to give Dempster-Shafer theory an advantage over the Bayesian approach to the handling of uncertainty, where in general the combination of several bodies of evidence requires very strong independence assumptions. However, although the way uncertainty is represented in Dempster-Shafer theory may be intuitively sound and attractive, it does cause the reasoning with uncertainty to be computationally very expensive:

Let Θ be a set of possible answers to a question. Then evidence may point to a proper subset of Θ without pointing to a particular element. E.g. some scratches on a forced door may tell an expert that the burglar was left-handed, without giving any clue to the question which left-handed person was the burglar. The Bayesian representation of this evidence consists of a uniform distribution of the weight attributed to the evidence over all left-handed persons, while in Dempster-Shafer theory this weight is attributed to the set of all left-handed persons without attributing any weight to a particular left-hander. In Dempster- Shafer theory evidence is represented by a belief function which is induced by a probability measure on the powerset of Θ instead of by a probability measure on Θ itself, like in the Bayesian approach. A consequence of this is that the amount of computation required for the combination of evidence by Dempster's rule increases exponentially with the cardinality of Θ, which is generally found to be a serious obstacle for the use of Dempster-Shafer theory.

Barnett (1981) has described an algorithm for Dempster's rule of combination which achieves computational savings in case the only proper subsets of Θ supported by the evidence are a singleton and its complement. Gordon and Shortliffe (1985) were not satisfied with this result, since they were attracted to Dempster-Shafer theory by its "potential for handling evidence bearing on categories of diseases as well as on specific disease entities". They suggest an efficient algorithm for combining evidence which can be applied in case the hypotheses of interest form a hierarchy. In general, their algorithm only yields an approximation of the result that would be obtained by using the full hypothesis space, but it was improved on in this respect by Shafer and Logan (1987).

In this paper we define a Bayesian approximation of a belief function and show that in general the combination of Bayesian approximations of belief functions is computationally less involving than the combination of the belief functions themselves. If combining the belief functions would yield a Bayesian belief function, i.e. a probability measure on Θ, then the substitution of belief functions by their Bayesian approximation will not affect the result of Dempster's rule. In general, the combination of Bayesian approximations of belief functions yields the Bayesian approximation of the combination of those belief functions. This property makes the results obtained by employing Bayesian approximations useful in at least those cases where one is interested in final conclusions about the elements of Θ rather than subsets of Θ.

Section 1 reviews the basics of Dempster-Shafer theory. In section 2 the Bayesian approximation of a belief function is defined and some properties of this approximation are given. The remaining sections are devoted to the relation between a belief function and its Bayesian approximation, a digression on the (at least formally) interesting space of generalized belief functions and a description of the algorithm for the combination of evidence by applying Dempster's rule to Bayesian approximations.

1 Dempster-Shafer theory

In this section we briefly explain some notions and terminology of Dempster-Shafer theory. For a more detailed exposition and some background information see e.g. Shafer (1976) or Gordon and Shortliffe (1985).

Let Θ be a set of mutually exclusive and exhaustive hypotheses about some problem domain. (Θ may be regarded to be a set of possible answers to a question.) Relevant propositions are represented as subsets of this set Θ which is called the *frame of discernment*. A *basic probability assignment* (bpa) is a function m from 2^{Θ}, the powerset of Θ, to [0,1] such that

$$m(\emptyset) = 0 \quad \text{and} \quad \sum_{A \subseteq \Theta} m(A) = 1$$

The quantity m(A), called A's *basic probability number*, corresponds to the measure of belief that is committed exactly to the proposition (represented by the set) A and in general not to the total belief committed to A, since this also includes the measures of belief committed to subsets of A. Hence we define the *belief function* Bel induced by a bpa m by:

$$\mathrm{Bel}(A) = \sum_{B \subseteq A} m(B) \qquad (A,B \subseteq \Theta)$$

Bel(A) measures the total belief committed to A. Each belief function Bel is induced by a unique bpa m which can be recovered from Bel as follows:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} Bel(B)$$

(Here A-B denotes $A \cap B^c$, the intersection of A and the complement of B, and $|A-B|$ denotes the cardinality of this set.)

The *plausibility* of A, Pl(A), is defined by $Pl(A) = 1 - Bel(A^c)$. It is easy to see that we have:

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

Notice that each function from {m,Bel,Pl} uniquely determines the other two.

Some additional terminology: Let m be the bpa of Bel. If m(A)>0, then A is called a *focal element* of Bel. The union of all focal elements of Bel is called the *core* of Bel. A belief function is called *vacuous* if Θ is its only focal element. If all focal elements of Bel are singletons, then Bel is called *Bayesian*. Notice that if Bel is Bayesian, then Bel and Pl coincide and are equivalent to a probability measure on Θ.

Let m and m' be the bpa's of the belief functions Bel and Bel' with cores $\{A_1, \ldots A_p\}$ and $\{B_1, \ldots, B_q\}$ respectively. Then $m \oplus m'$, the orthogonal combination of m and m' is given by the following formula, which is called *Dempster's rule of combination*:

$$m \oplus m'(A) = \frac{\sum_{A_i \cap B_j = A} m(A_i) \cdot m'(B_j)}{\sum_{A_i \cap B_j \neq \emptyset} m(A_i) \cdot m'(B_j)} \qquad \text{if } A \neq \emptyset;\ m \oplus m'(\emptyset) = 0.$$

The factor $[\sum_{A_i \cap B_j \neq \emptyset} m(A_i) \cdot m'(B_j)]^{-1}$ is called the *renormalizing constant* of Bel and Bel'.

We write $Bel \oplus Bel'$ for the belief function induced by $m \oplus m'$. The intuition behind the rule is that the combined effect of the assignment of $m(A_i)$ probability mass to A_i and the assignment of $m(B_j)$ probability mass to B_j is the assignment of $m(A_i) \cdot m(B_j)$ probability mass to $A_i \cap B_j$. A given subset A of Θ may of course be the intersection of A_i and B_j for more than one pair (i,j). Hence to obtain the total probability mass exactly committed to A by the combination m⊕m' of m and m' we have to take the sum of all $m(A_i)\cdot m'(B_j)$ such that $A = A_i \cap B_j$. The case $A_i \cap B_j = \varnothing$ forms an exception, since by definition no probability mass is assigned to $\varnothing$. Therefore the measures of the remaining intersections of focal elements are rescaled by dividing through the sum of all $m(A_i)\cdot m'(B_j)$ such that $A_i \cap B_j \neq \varnothing$, provided this sum does not equal 0; otherwise one says that Bel⊕Bel' does not exists or that Bel and Bel' are not combinable.

We list some useful properties of ⊕:

1. $Bel_1 \oplus Bel_2 = Bel_2 \oplus Bel_1$
2. $Bel_1 \oplus (Bel_2 \oplus Bel_3) = (Bel_1 \oplus Bel_2) \oplus Bel_3$
3. If Bel_1 is vacuous, then $Bel_1 \oplus Bel_2 = Bel_2$
4. If Bel_1 is Bayesian, then $Bel_1 \oplus Bel_2$ is Bayesian.

Notation: Let I = {1,2,...,n}, then $\bigoplus_{i \in I} Bel_i$ denotes $Bel_1 \oplus Bel_2 \oplus \ldots \oplus Bel_n$.

2 Bayesian approximation

Definition Let m be a bpa and Bel the belief function induced by m. The *Bayesian approximation* $\underline{Bel}$ of Bel is induced by the bpa $\underline{m}$ defined by:

$$\underline{m}(A) = \frac{\sum_{A \subseteq B} m(B)}{\sum_{C \subseteq \Theta} m(C)\cdot|C|} \quad \text{if A is a singleton; otherwise, } \underline{m}(A) = 0.$$

The factor $[\sum_{C \subseteq \Theta} m(C)\cdot|C|]^{-1}$ will be called the *Bayesian constant* of Bel.

It is clear that $\underline{Bel}$ is Bayesian. Notice that in general the Bayesian approximation of Bel essentially differs from the Bayesian belief function obtained from Bel by distributing uniformly all probability mass assigned by m to subsets of Θ over their elements.

Example

Let Θ = {a,b,c}, m({a}) = 0.4 and m({b,c}) = 0.6. Then $\underline{m}$({a}) = 0.4/(0.4·1 + 0.6·2) = 0.25 and $\underline{m}$({b}) = $\underline{m}$({c}) = 0.6/(0.4·1 + 0.6·2) = 0.375.

The Bayesian constant c of Bel may be considered to be a measure of precision or specificity of the information given by Bel, where $c = |\Theta|^{-1}$ corresponds with the most imprecise, i.e. vacuous, belief function and $c = 1$ with Bayesian belief functions, which are maximally precise. (In Dubois and Prade (1987a) $\sum m(C)\cdot|C|$ is mentioned as an example of a measure of imprecision, although it does not possess all appropriate properties; $\sum m(C)\cdot\log_2|C|$ is better suited in this respect. See Dubois and Prade (1987b) and Ramer (1987).)

The following proposition summarizes some trivial facts:

Proposition 1

(i) Bel = <u>Bel</u> iff Bel is Bayesian iff the Bayesian constant of Bel = 1.

(ii) Bel and Bel' are not combinable iff <u>Bel</u> and <u>Bel'</u> are not combinable.

Proof trivial.

Proposition 2 below states that the combination of the Bayesian approximations of two combinable belief functions is identical to the Bayesian approximation of the combination of the belief functions themselves.

Proposition 2 If Bel and Bel' are combinable, then $\underline{\mathrm{Bel}}\oplus\underline{\mathrm{Bel'}} = \underline{\mathrm{Bel}\oplus\mathrm{Bel'}}$

Proof It is clear that if A is not a singleton, then $\underline{m}\oplus\underline{m'}(A) = 0 = \underline{m\oplus m'}(A)$. Let c (c') denote the Bayesian constant of Bel (Bel') and let k be the renormalizing constant of Bel and Bel'. Then we have:

$$\underline{m}\oplus\underline{m'}(\{a\}) = \frac{\sum_{\{a\}=B\cap C} \underline{m}(B)\cdot\underline{m'}(C)}{\sum_{B\cap C\neq\varnothing} \underline{m}(B)\cdot\underline{m'}(C)} = \frac{\underline{m}(\{a\})\cdot\underline{m'}(\{a\})}{\sum_{b\in\Theta} \underline{m}(\{b\})\cdot\underline{m'}(\{b\})}$$

$$= \frac{c\cdot(\sum_{a\in C} m(C))\cdot c'\cdot(\sum_{a\in D} m'(D))}{\sum_{b\in\Theta}(c\cdot(\sum_{b\in E} m(E))\cdot c'\cdot(\sum_{b\in F} m'(F)))} = \frac{(\sum_{a\in C} m(C))\cdot(\sum_{a\in D} m'(D))}{\sum_{b\in\Theta}((\sum_{b\in E} m(E))\cdot(\sum_{b\in F} m'(F)))}$$

$$= \frac{\sum_{a\in C\cap D} m(C)\cdot m'(D)}{\sum_{E\cap F\neq\emptyset} m(E)\cdot m'(F)\cdot|E\cap F|} = \frac{\sum_{a\in B}\ (\sum_{B=C\cap D} m(C)\cdot m'(D))}{\sum_{C\subseteq\Theta}\ (\sum_{E\cap F=C} m(E)\cdot m'(F)\cdot|E\cap F|)}$$

$$= \frac{k\cdot\sum_{a\in B}\ (\sum_{B=C\cap D} m(C)\cdot m'(D))}{k\cdot\sum_{C\subseteq\Theta}\ (\sum_{E\cap F=C} m(E)\cdot m'(F)\cdot|E\cap F|)} = \frac{\sum_{a\in B}\ (k\cdot\sum_{B=C\cap D} m(C)\cdot m'(D))}{\sum_{C\subseteq\Theta}\ (|C|\cdot k\cdot\sum_{E\cap F=C} m(E)\cdot m'(F))}$$

$$= \frac{\sum_{a\in B} m\oplus m'(B)}{\sum_{C\subseteq\Theta} m\oplus m'(C)\cdot|C|} = \underline{m\oplus m'}(\{a\})\,. \qquad \square$$

Since the Bayesian approximation of a belief function is Bayesian, the combination of Bayesian approximations will also be Bayesian. Hence, a necessary condition for the combination of Bayesian approximations to agree with the combination of the belief functions themselves is that this latter combination is Bayesian. Proposition 3 shows that this condition is also sufficient.

Proposition 3 Let I be some non-empty set and assume that the belief functions from $\{Bel_i \mid i\in I\}$ are combinable. Then

$$\oplus_{i\in I}Bel_i = \oplus_{i\in I}\underline{Bel_i} \text{ iff } \prod\{m_i(A_i) \mid i\in I\}>0 \text{ implies } |\cap_{i\in I}A_i| \leq 1.$$

Proof $\prod\{m_i(A_i) \mid i\in I\} > 0 \rightarrow |\cap_{i\in I}A_i| \leq 1 \;\Leftrightarrow\; \oplus_{i\in I}Bel_i$ is Bayesian

$\Leftrightarrow \oplus_{i\in I}Bel_i = \underline{\oplus_{i\in I}Bel_i}$ (prop. 1)

$\Leftrightarrow \oplus_{i\in I}Bel_i = \oplus_{i\in I}\underline{Bel_i}$ (prop. 2). $\square$

Corollary

(i) If $A\cap B$ is a singleton or empty whenever $m(A)m'(B) > 0$, then $Bel\oplus Bel' = Bel\oplus\underline{Bel'} = \underline{Bel}\oplus\underline{Bel'}$.

(ii) If Bel is Bayesian, then $Bel\oplus Bel' = Bel\oplus\underline{Bel'}$.

3 The relation between Bel and Bel'

The term "Bayesian approximation" may be somewhat misleading, since there is for example probably no natural concept of distance between two belief functions with respect to which $\underline{\text{Bel}}$ is the Bayesian belief function closest to Bel. Our justification of the term is that in many cases one can draw conclusions from the Bayesian approximation of a belief function which are similar to those that can be drawn from the belief functions themselves. If this would not be the case, then Bayesian approximations would not be very useful, since proposition 3 shows that the conditions under which the combination of Bayesian approximations give exactly the same result as the combination of the belief functions themselves are rather strict.

Although evidence may point to subsets of the frame of discernment Θ without pointing to any particular element, one is often just interested in final conclusions about the elements of Θ. (E.g. one is primarily interested in diagnosing the *particular* diseases of patients, although symptoms are usually most naturally interpreted as evidence for *categories* of diseases.) In Dempster-Shafer theory the information about the degree of certainty of an element a is represented by the belief interval $[\text{Bel}(\{a\}),\text{Pl}(\{a\})]$. It is easy to see that $\underline{m}(\{a\}) = c\cdot\text{Pl}(\{a\})$, where c is the Bayesian constant of Bel. Hence one can extract from $\underline{m}$ and c at least one of the two functions which carry the information about the degree of certainty of elements of Θ. Unfortunately, in general $\text{Bel}(\{a\})$ cannot be reconstructed from $\underline{m}$. However, the additional knowledge of $\text{Bel}(\{a\})$ does not yield too much new information:

Even when the belief intervals of elements of Θ are given there is no unique way to order them with respect to their degree of certainty. Two more or less natural orderings are those induced by the following orderings on intervals:

Definition

(i) The *minimal ordering* $\leq_{min}$ is defined by $[x,y] \leq_{min} [x',y']$ iff $y \leq x'$.

(ii) The *ordering by average* $\leq_{av}$ is defined by $[x,y] \leq_{av} [x',y']$ iff $(x+y)/2 \leq (x'+y')/2$.

Let $a,b\in \Theta$. We write $a \leq_{min} b$ for $[\text{Bel}(\{a\}),\text{Pl}(\{a\})] \leq_{min} [\text{Bel}(\{b\}),\text{Pl}(\{b\})]$ and $a =_{min} b$ for $a \leq_{min} b \wedge b \leq_{min} a$. $a \leq_{av} b$ and $a =_{av} b$ are defined similarly.

The choice for the minimal ordering corresponds with a rather cautious approach to the ordering of elements with respect to their certainty (resulting in general only in a partial ordering), whereas the choice for the ordering by average requires a rather audacious approach. In table 1 the orderings $\leq_{min}$ and $\leq_{av}$ are compared with the *plausibility ordering*

$\leq_{pl}$ defined by a $\leq_{pl}$ b iff Pl({a}) ≤ Pl({b}) (or equivalently $\underline{m}$({a}) ≤ $\underline{m}$({b})).

1	Bel({a}) ≤ Pl({a}) < Bel({b}) ≤ Pl({b}):	a $<_{min}$ b	a $<_{av}$ b	a $<_{pl}$ b
2	Bel({a}) ≤ Pl({a}) = Bel({b}) < Pl({b}):	a $\leq_{min}$ b	a $<_{av}$ b	a $<_{pl}$ b
3	Bel({a}) < Pl({a}) = Bel({b}) = Pl({b}):	a $\leq_{min}$ b	a $<_{av}$ b	a $=_{pl}$ b
4	Bel({a}) = Pl({a}) = Bel({b}) = Pl({b}):	a $=_{min}$ b	a $=_{av}$ b	a $=_{pl}$ b
5	Bel({a}) ≤ Bel({b}) < Pl({a}) < Pl({b}):	*	a $<_{av}$ b	a $<_{pl}$ b
6	Bel({a}) < Bel({b}) < Pl({a}) = Pl({b}):	*	a $<_{av}$ b	a $=_{pl}$ b
7	Bel({a}) = Bel({b}) < Pl({a}) = Pl({b}):	*	a $=_{av}$ b	a $=_{pl}$ b
8	Bel({a}) < Bel({b}) ≤ Pl({b}) < Pl({a}):	*	?	b $<_{pl}$ a

TABLE 1. "*" stands for "a and b are incomparable". "?" means that the ordering of a and b is not determined by the given description of the situation.

Notice that whenever $\leq_{min}$ and $\leq_{av}$ agree (i.e. in case 1 and 4 of table 1), $\leq_{pl}$ gives the same result. Hence if the belief intervals give rise to definite conclusions about the relative degree of certainty of elements of Θ, then these conclusions can already be drawn from $\underline{m}$ or the plausibility functions. A drawback of having available only $\underline{m}$ or {Pl({a}) | a∈ Θ} is that in that case one does not always know whether a conclusion based on $\leq_{pl}$ agrees with any reasonable conclusion or constitutes just a particular choice from several possible conclusions. Even additional information about the Bayesian constant c does not always enables one to decide whether one of the situations 1 and 4 of table 1 holds, although sometimes it does: e.g. Pl({b}) > Pl({a}) ∧ Pl({a}) < 0.5 is a necessary and Pl({b}) > Pl({a}) ∧ c > $[1 + \mathrm{Pl}(\{b\}) - \mathrm{Pl}(\{a\}]^{-1}$ is a sufficient condition for situation 1 and Pl({a}) = Pl({b}) ∧ c = 1 is a sufficient condition for situation 4.

However, to conclude that a particular element a from Θ is (likely to be) the case it often does not suffice to know that a is at least as certain as any b∈ Θ for any reasonable ordering of the elements of Θ with respect to their certainty. E.g. if the belief interval of a is [0.5,0.55] and that of b is [0.4,0.45], then for any reasonable ordering a ≥ b, but for many applications it would be unwise to disregard the possibility that b is the case. In fact, to conclude a it is often necessary and sufficient that for all b ≠ a Pl({a}) >> Pl({b}). This justifies to some extent the dominant role of plausibility in the process of deciding between elements on the basis of $\leq_{pl}$.

Another possible (partial) justification of the bias of $\leq_{pl}$ towards the plausibility of elements may be extracted from the fact that Dempster's rule is also somewhat biassed towards plausibility. This fact is illustrated by the following: let Bel' be a belief function such that all elements of Θ have the same belief intervals. If $Pl(\{a\}) \geq Pl(\{b\})$, then $Pl\oplus Pl'(\{a\}) \geq Pl\oplus Pl'(\{b\})$. Hence $\leq_{pl}$ is not affected by the combination with a "neutral" belief function, whereas $\leq_{min}$, $\leq_{av}$ and $\leq_{bel}$, defined by a $\leq_{bel}$ b iff $Bel(\{a\}) \leq Bel(\{b\})$, are not necessarily unaffected by the combination with a belief function like Bel'. One may conclude that if one is interested in conclusions about the elements of Θ, then $\leq_{pl}$, and therefore the Bayesian approximation, may often yield sufficient information.

However, one cannot always choose the frame of discernment to consist of just the propositions one is interested in, since the application of Dempster's rule requires the frame of discernment to discern all relevant interaction of the evidence to be combined. (See Shafer (1976), chapter 8.) Therefore the frame may contain some propositions which refer to details in which one is not primarily interested.

In general we have the following inequalities:

$$\max\{\underline{m}(\{a\}) \mid a\in A\} \leq Pl(A) \leq \underline{Pl}(A)/c$$

$$1 - (1 - \underline{Bel}(A))/c \leq Bel(A) \leq 1 - \max\{\underline{m}(\{a\}) \mid a\notin A\}$$

These inequalities are only likely to give some information if the Bayesian constant is close to 1. One might be inclined to think that the combination by Dempster's rule increases precision in the sense that the Bayesian constant of the combination is (weakly) larger than the Bayesian constants of the belief functions which are combined. This would imply that as the number of combinations increases, so would the likelihood of the above inequalities being informative. Unfortunately, this is not the case since, as is shown in the following section, the application of Dempster's rule involves a normalizing step which may cause the Bayesian constant to decrease.

4 Generalized belief functions

Definition Let Θ be a frame of discernment. The *space of generalized bpa's or belief functions* on Θ is a pair $(M,\otimes)$, where $M = \{m : 2^{\Theta} \rightarrow [0,1] \mid \sum_{A\subseteq\Theta} m(A) = 1\}$ and $\otimes$ is a binary operation on M such that for all $A\subseteq\Theta$: $m_1\otimes m_2(A) = \sum_{B\cap C=A} m_1(B)\cdot m_2(C)$.

The space $(M,\otimes)$ is introduced in Hummel and Landy (1988), where it is called "space of unnormalized belief states". An element m of M will be called a *generalized bpa*. A generalized bpa m induces a *generalized belief function Bel* by $Bel(A) = m(\varnothing)$, if $A=\varnothing$ and $Bel(A) = \sum_{\varnothing\neq B\subseteq A} m(B)$, otherwise.

$(M,\otimes)$ is an abelian monoid which is closely related to the space of bpa's: let m_0 be the function $2^{\Theta} \rightarrow [0,1]$ defined by $m_0(\emptyset) = 1$ and $m_0(A) = 0$ for all $A \neq \emptyset$ and let g be the map defined by $g(m_0) = m_0$ and for all $m \neq m_0$ $g(m) = m$, where m is the bpa such that for all $A \neq \emptyset$ $m(A) = m(A)/(1 - m(\emptyset))$. Then g maps $(M,\otimes)$ homomorphically onto $(M,\oplus)$, where M is the set of bpa's with as extra element m_0 and $\oplus$ is defined as usual, except that $m_1 \oplus m_2$ is defined to be m_0 in case m_1 and m_2 are not combinable according to the definition given in section 1. (See Hummel and Landy (1988).)

Although $(M,\otimes)$ is primarily introduced for technical reasons, it might be an intuitively appealing space for those who object against the normalizing step in Dempster's rule and its effect of suppressing the conflict of evidence (cf. Zadeh (1984]). The quantity $m(\emptyset)$ may be interpreted as a measure of this conflict of evidence. Further, the space of generalized belief functions may be used to make precise the claim that it is the normalizing factor which is responsible for the fact that the application of Dempster's rule may cause the Bayesian constant to decrease:

Proposition 4 Let $m_1, m_2 \in M$ and let $m_1 \otimes m_2(\emptyset) \neq 1$. Then

$$[\sum_{A \subseteq \Theta} m_1 \otimes m_2(A) \cdot |A|]^{-1} \geq [\sum_{B \subseteq \Theta} m_i(B) \cdot |B|]^{-1} \quad (i \in \{1,2\})$$

Proof

$$\sum_{A \subseteq \Theta} m_1 \otimes m_2(A) \cdot |A| = \sum_{B \cap C \neq \emptyset} m_1(B) \cdot m_2(C) \cdot |B \cap C|$$

$$= \sum_{B \subseteq \Theta} m_1(B) \cdot (\sum_{B \cap C \neq \emptyset} m_2(C) \cdot |B \cap C|)$$

$$\leq \sum_{B \subseteq \Theta} m_1(B) \cdot |B| \cdot (\sum_{B \cap C \neq \emptyset} m_2(C))$$

$$\leq \sum_{B \subseteq \Theta} m_1(B) \cdot |B|$$

Hence $[\sum_{A \subseteq \Theta} m_1 \otimes m_2(A) \cdot |A|]^{-1} \geq [\sum_{B \subseteq \Theta} m_1(B) \cdot |B|]^{-1}$. The case i = 2 is similar. □

Corollary Suppose that Bel_1 and Bel_2 are combinable and let k be the Bayesian constant of Bel_1 and Bel_2. Then

$$[\sum_{A \subseteq \Theta} m_1 \oplus m_2(A).|A|]^{-1} \geq [k \cdot \sum_{B \subseteq \Theta} m_i(B).|B|]^{-1} \quad (i \in \{1,2\})$$

The following example shows that the renormalizing constant may not be omitted from the statement above:

Example Let $\Theta = \{a,b,c,d\}$ and let m_1 be given by $m_1(\{a,b\}) = m_1(\{c\}) = 0.5$ and $m_2(\{a,b\}) = m_2(\{d\}) = 0.5$. Then the Bayesian constant of m_1 and m_2 is 2/3, while the Bayesian constant of $m_1 \oplus m_2$ is 0.5.

A further generalization of the space of bpa's is obtained by dropping in addition to $m(\varnothing) = 0$ also the requirement that $\sum m(A) = 1$:

Definition Let Θ be a frame of discernment. The *space of unnormalized generalized bpa's or belief functions* on Θ is a pair $(\mathbf{M},\otimes)$, where $\mathbf{M} = \{\mathbf{m} : 2^\Theta \to [0.1]\}$ and and $\otimes$ is a binary operation on $\mathbf{M}$ such that for all $A \subseteq \Theta$: $\mathbf{m_1} \otimes \mathbf{m_2}(A) = \sum_{B \cap C = A} \mathbf{m_1}(B) \cdot \mathbf{m_2}(C)$.

$(\mathbf{M},\otimes)$ is again an abelian monoid, which is mapped homomorphically onto $(M,\oplus)$ and $(\mathcal{M},\otimes)$ by h and h' respectively, where $h(m) = h'(m) = m_0$, if for all $A \neq \varnothing$ $m(A) = 0$ and otherwise $h'(m)(A) = m(A)/(\sum \{m(B) \mid B \subseteq \Theta\})$, $h(m)(\varnothing) = 0$ and for $A \neq \varnothing$ $h(m)(A) = m(A)/(\sum \{m(B) \mid B \neq \varnothing\})$. To formulate a generalized notion of Bayesian approximation in this formal framework, we need the following definition:

Definition

(i) the *space of Bayesian bpa's or belief functions* $(\underline{M},\oplus)$ is the submonoid of $(M,\oplus)$ such that $\underline{M} = \{m \in M \mid m(A) > 0 \text{ implies } |A| \leq 1\}$

(ii) the *Bayesian approximation map* $f : \mathbf{M} \to \underline{M}$ is the map given by $f(\mathbf{m}) = m_0$, if for all $A \neq \varnothing$ $\mathbf{m}(A) = 0$ and otherwise,

$$f(\mathbf{m})(A) = \frac{\sum_{A \subseteq B} \mathbf{m}(B)}{\sum_{C \subseteq \Theta} \mathbf{m}(C) \cdot |C|}$$, if A is a singleton and $f(\mathbf{m})(A) = 0$, if $|A| \neq 1$.

Since $M \subseteq \mathcal{M} \subseteq \mathbf{M}$, f can be restricted to M and $\mathcal{M}$. The functions thus obtained will also be called Bayesian approximation maps. Notice that if m is a bpa, then $f(m) = \underline{m}$. Hence f yields a possible generalization of the notion of Bayesian approximation.

Proposition 5 f maps $(\mathbf{M},\otimes)$, $(M,\oplus)$ and $(\mathrm{M},\oplus)$ homomorphically onto $(\underline{\mathrm{M}},\oplus)$ and the following diagram commutes:

$$\begin{tikzcd}
(\mathbf{M},\otimes) \arrow[rr, two heads, "h"] \arrow[dr, two heads, "h'"] \arrow[ddr, two heads, "f"'] & & (\mathrm{M},\oplus) \arrow[ddl, two heads, "f"] \\
& (M,\otimes) \arrow[ur, two heads, "g"] \arrow[d, two heads, "f"] & \\
& (\underline{\mathrm{M}},\oplus) &
\end{tikzcd}$$

Proof

Since f is the identity function on $\underline{\mathrm{M}}$ and $\underline{\mathrm{M}} \subseteq \mathrm{M} \subseteq M \subseteq \mathbf{M}$, f is clearly onto. The fact that f is a homomorphism follows essentially from prop. 1(ii) and (the proof of) prop. 2. The proof that the diagram commutes is straightforward. □

Determining the combination of the Bayesian approximation of n belief functions involves several normalizing steps: n applications of f and n-1 applications of $\oplus$ yield a total of 2n-1 normalizing steps. The following proposition shows that all these normalizing steps can be merged into one step. First we need some definitions.

Definition

(i) The *space of unnormalized generalized Bayesian bpa's or belief functions* $(\underline{\mathbf{M}},\otimes)$ is the submonoid of $(\mathbf{M},\otimes)$ such that $\underline{\mathbf{M}} = (\mathbf{m} \in \mathbf{M} \mid \mathbf{m}(A) > 0 \text{ implies } |A| \leq 1\}$.

(ii) The *unnormalized Bayesian approximation map* $f' : \mathbf{M} \rightarrow \underline{\mathbf{M}}$ is defined by $f'(\mathbf{m}) = m_0$, if for all $A \neq \varnothing$ $\mathbf{m}(A) = 0$, otherwise $f'(\mathbf{m})(A) = \sum \{\mathbf{m}(B) \mid A \subseteq B\}$ if A is a singleton and for all A with $|A| \neq 1$ $f'(\mathbf{m})(A) = 0$.

(iii) The *generalized normalization map* $g' : \underline{\mathbf{M}} \rightarrow \underline{\mathrm{M}}$ is defined by $g'(\underline{\mathbf{m}}) = m_0$, if for all non-empty A $\underline{\mathbf{m}}(A) = 0$, otherwise $g'(\underline{\mathbf{m}})(A) = \underline{\mathbf{m}}(A)/(\sum \{\underline{\mathbf{m}}(B) \mid B \neq \varnothing)\}$ for $A \neq \varnothing$ and $g'(\underline{\mathbf{m}})(\varnothing) = 0$.

Proposition 6 f' maps $(\mathrm{M},\otimes)$ homomorphically onto $(\underline{\mathbf{M}},\otimes)$, g' maps $(\underline{\mathbf{M}},\otimes)$ homomorphically onto $(\underline{\mathrm{M}},\oplus)$ and the following diagram commutes:

$$\begin{tikzcd}
(\mathbf{M},\otimes) \arrow[d, two heads, "f'"'] \arrow[dr, two heads, "f"] & \\
(\underline{\mathbf{M}},\otimes) \arrow[r, two heads, "g'"'] & (\underline{\mathbf{M}},\oplus)
\end{tikzcd}$$

Proof straightforward. □

Corollary Let m_1 and m_2 be bpa's. Then $f(m_1 \oplus m_2) = g'(f'(m_1 \otimes f'(m_2))$.

This fact will be used in the following section to avoid unnecessary and cumbersome normalizing steps.

5 An algorithm for combining Bayesian approximations

Let $|\Theta| = n$ and let $\{A_j \mid 0 \leq j \leq 2^n-1\}$ be a enumeration of the subsets of Θ.

(i) Compute for each belief function Bel_i its Bayesian approximation (up to a multiplicative constant) by adding for all j the quantity $m_i(A_j)$ to $f'(m_i)(\{a\})$, for all $a \in A_j$. (We suppose that the initial value of $f'(m_i)(\{a\})$ is zero.) Normalizing $f'(m_i)$ would yield the Bayesian approximation $\underline{m}_i$, but we will postpone the normalization until the functions have been combined. (Here we apply (the corollary of) proposition 6.)

(ii) $\bigoplus_{i \in I} \underline{Bel}_i$ is now obtained by first computing, for all $a \in \Theta$, $\prod_{i \in I} f'(m_i)(\{a\})$,

then taking the sum of these products over all elements of Θ and finally setting

$$\bigoplus_{i \in I} \underline{Bel}_i(\{a\}) = \frac{\prod_{i \in I} f'(m_i)(\{a\})}{\sum_{b \in \Theta} \left(\prod_{i \in I} f'(m_i)(\{b\})\right)}.$$

It is easy to see that the computation of $\bigoplus_{i \in I} \underline{Bel}_i$, as described in step (ii), only requires time polynomial in $|\Theta|$, while, in general, the computation of the combination of Bel_i would require time exponential in the cardinality of Θ. However, this does not imply that the combination of belief functions is always computationally more involving than the combination of the associated Bayesian approximations since if step (i) is implemented in a straightforward way, then, in general, the computation of $f'(m_i)$ requires time exponential in $|\Theta|$.

Still, there are some situations in which it is, from a computational point of view, clearly advantageous to combine the Bayesian approximations rather than the belief functions themselves. E.g. if the number of combinations, i.e. $|I|$, is not too small relative to $|\Theta|$ (in that case the computational savings obtained by combining the Bayesian approximations in stead of the belief functions themselves weighs up against the additional work needed to compute the Bayesian approximations) or if the evidence is not given in terms of basic probability assignments, but in terms of the plausibility of the elements of Θ (this would make step (i) redundant, since replacing $f'(m_i)$ by Pl_i in step (ii) will not affect the result). In short, for many applications it might be worth while to consider the possibility of using Bayesian approximations in stead of the belief functions themselves.

References

BARNETT J.A. (1981), Computational methods for a mathematical theory of evidence, in: *Proceedings IJCAI-81*, Morgan Kaufmann, Los Altos CA, 868-875.

DUBOIS D. and PRADE H. (1987a), The principle of minimum specifity as a basis for evidential reasoning, in: *Knowledge-Based Systems*, B. Bouchon & R.R. Yager (eds.), Lecture Notes in Computer Science, Berlin, 75-84.

DUBOIS D. and PRADE H. (1987b), Properties of measures of information in evidence and possibility theory, *Fuzzy Sets and Systems* 24, 161-182.

GORDON J. and H. SHORTLIFFE E.H. (1985), A method for managing evidential reasoning in a hierarchical hypothesis space, *Artificial Intelligence* 26, 323-357.

HUMMEL R.A. and LANDY M.S. (1988), A statistical viewpoint on the theory of evidence, *Technical Report 194*, New York University (1985; revised: 1987). To appear in *IEEE Transactions on Pattern Analysis and Machine Intelligence* (1988).

RAMER A. (1987), Uniqueness of information measure in the theory of evidence, *Fuzzy Sets and Systems* 24, 183-196.

SHAFER G. (1976), *A Mathematical Theory of Evidence*, Princeton NJ, Princeton University Press.

SHAFER G. and LOGAN R. (1987), Implementing Dempster's rule for hierarchical evidence, *Artificial Intelligence* 33, 271-298.

ZADEH, L.A. (1984), Review of Shafer's a mathematical theory of evidence, *The AI Magazine* 5(3), 81-83.

Logic Group Preprint Series
Department of Philosophy
University of Utrecht
Heidelberglaan 2
3584 CS Utrecht
The Netherlands

nr. 1 C.P.J. Koymans, J.L.M. Vrancken, *Extending Process Algebra with the empty process*, September 1985.
nr. 2 J.A. Bergstra, *A process creation mechanism in Process Algebra,* September 1985.
nr. 3 J.A. Bergstra, *Put and get, primitives for synchronous unreliable message passing,* October 1985.
nr. 4 A.Visser, *Evaluation, provably deductive equivalence in Heyting's arithmetic of substitution instances of propositional formulas*, November 1985.
nr. 5 G.R. Renardel de Lavalette, *Interpolation in a fragment of intuitionistic propositional logic,* January 1986.
nr. 6 C.P.J. Koymans, J.C. Mulder, *A modular approach to protocol verification using Process Algebra* , April 1986.
nr. 7 D. van Dalen, F.J. de Vries, *Intuitionistic free abelian groups,* April 1986.
nr.8 F. Voorbraak, *A simplification of the completeness proofs for Guaspari and Solovay's R*, May 1986.
nr.9 H.B.M. Jonkers, C.P.J. Koymans & G.R. Renardel de Lavalette, *A semantic framework for the COLD-family of languages,* May 1986.
nr.10 G.R. Renardel de Lavalette, *Strictheidsanalyse,* May 1986.
nr.11 A. Visser, *Kunnen wij elke machine verslaan? Beschouwingen rondom Lucas' argument*, July 1986.
nr.12 E.C.W. Krabbe, *Naess's dichotomy of tenability and relevance,* June 1986.
nr.13 Hans van Ditmarsch, *Abstractie in wiskunde, expertsystemen en argumentatie,* Augustus 1986
nr.14 A.Visser, *Peano's Smart Children, a provability logical study of systems with built-in consistency* , October 1986.
nr.15 G.R.Renardel de Lavalette, *Interpolation in natural fragments of intuitionistic propositional logic,* October 1986.
nr.16 J.A. Bergstra, *Module Algebra for relational specifications,* November 1986.
nr.17 F.P.J.M. Voorbraak, *Tensed Intuitionistic Logic,* January 1987.
nr.18 J.A. Bergstra, J. Tiuryn, *Process Algebra semantics for queues,* January 1987.
nr.19 F.J. de Vries, *A functional program for the fast Fourier transform,* March 1987.
nr.20 A. Visser, *A course in bimodal provability logic*, May 1987.
nr.21 F.P.J.M. Voorbraak, *The logic of actual obligation, an alternative approach to deontic logic,* May 1987.
nr.22 E.C.W. Krabbe, *Creative reasoning in formal discussion,* June 1987.
nr.23 F.J. de Vries, *A functional program for Gaussian elimination*, September 1987.
nr.24 G.R. Renardel de Lavalette, *Interpolation in fragments of intuitionistic propositional logic*, October 1987.(revised version of no. 15)
nr.25 F.J. de Vries, *Applications of constructive logic to sheaf constructions in toposes*, October 1987.
nr.26 F.P.J.M. Voorbraak, *Redeneren met onzekerheid in expertsystemen*, November 1987.
nr.27 P.H. Rodenburg, D.J. Hoekzema, *Specification of the fast Fourier transform algorithm as a term rewriting system,* December 1987.

nr.28 D. van Dalen, *The war of the frogs and the mice, or the crisis of the Mathematische Annalen,* December 1987.

nr.29 A. Visser, *Preliminary Notes on Interpretability Logic,* January 1988.

nr.30 D.J. Hoekzema, P.H. Rodenburg, *Gauß elimination as a term rewriting system*, January 1988.

nr. 31 C. Smorynski, *Hilbert's Programme,* January 1988.

nr. 32 G.R. Renardel de Lavalette, *Modularisation, Parameterisation, Interpolation,* January 1988.

nr. 33 G.R. Renardel de Lavalette, *Strictness analysis for POLYREC, a language with polymorphic and recursive types*, March 1988.

nr. 34 A. Visser, *A Descending Hierarchy of Reflection Principles,* April 1988.

nr. 35 F.P.J.M. Voorbraak, *A computationally efficient approximation of Dempster-Shafer theory,* April 1988.