COMBINATION AND PROPAGATION OF UNCERTAINTY WITH BELIEF FUNCTIONS - A REEXAMINATION -

Didier Dubois Henri Prade

Universite Paul Sabatier
Laboratoire Langages et Sysiemes Informatiques
118 route de Narbonne, 31062 TOULOUSE CEDEX - FRANCE

ABSTRACT

The behavior of Dempster's rule of combination in typical situations is examined. Particularly, it is shown that assessing a zero value or a very small value may lead to very different results. Moreover a comparison with a possibility theory-based approach in case of conflicting information is provided. The general problem of representing uncertainty with one or several numbers is addressed. Lastly, the propagation of uncertainty from a fact and "if...then..." rule is discussed in the framework of belief functions.

I - INTRODUCTION

The treatment of uncertain information in knowledge engineering has encountered an increasing interest among AI researchers in the recent years. Roughly speaking, there are at least two basic problems when reasoning with uncertain facts or rules ; namely the combination problem and the propagation problem. The combination problem refers to the aggregation of uncertain pieces of information issued from different sources dealing with the same matter. The propagation problem deals with the aggregation of the uncertainty concerning the satisfaction of the condition-part of a rule with the uncertainty of the rule itself in order to deduce the uncertainty pervading the conclusion of the rule. Two theoretical frameworks have recently emerged for discussing these problems : the Dempster-Shafer theory of evidence [4] and Zadeh's possibility theory ; see Prade [3] for a comparative overview. In the following we examine how Dempster's rule of combination behaves in typical situations and how the result given by this rule may depend on the assessment of the values of the basic assignment. Then the propagation problem is briefly discussed in Shafer's framework.Comparisons are made with the possibilistic approach as well as with the treatment of uncertainty in inference systems such as MYCIN. The key question of the representation of the uncertainty of a fact or of a rule by more than one number is discussed throughout the whole paper.

II - COMBINATION PROBLEM

In Shafer's approach an uncertain body of evidence is represented by a so-called basic probability assignment m which is a set-function from the set Ω of possible elementary events to the real interval [0,1] such that

$$m(\emptyset) = 0 \; ; \; \sum_{A \subseteq \Omega} m(A) = 1 \qquad (1)$$

The quantity m(A) is supposed to represent our part of belief in the occurrence of A exactly. See Garvey et al. (IJCAI-81, pp 319-325) for an application-oriented presentation. Dempster's rule combines two probability assignments m_1 and m_2 pertaining to two sources of information, and obtains a new basic probability assignment m satisfying (1), which is defined by $\forall C \neq \emptyset$, $C \subseteq \Omega$, $m(C) =$

$$\left(\sum_{A \cap B = C} m_1(A).m_2(B)\right) \Big/ \left(\sum_{A \cap B \neq \emptyset} m_1(A).m_2(B)\right) \qquad (2)$$

Moreover this rule is associative.

1) The case of two alternatives

In the following we completely investigate the case where Ω only contains two known alternatives, each one being regarded as opposite to the other one, i.e. $\Omega = \{a, \neg a\}$. This case is interesting in the scope of expert systems because it is found when general rules enable the same conclusion a to be derived with various levels of uncertainty, and as such, is often encountered in applications. In this case a basic probability assignment m is defined by the three numbers $m(a)$, $m(\neg a)$, $m(\Omega)$ which have to satisfy the constraint (1). Then the degree of belief in the alternative a is given by Bel(a) = m(a) while the plausibility of a is computed as $Pl(a)=m(a)+m(\Omega)=1-Bel(\neg a)$, from the general formulas [4]

$$Bel(A) = \sum_{B \subseteq A} m(B) \; ; \; Pl(A) = 1-Bel(\bar{A}) = \sum_{A \cap B \neq \emptyset} m(B) \qquad (3)$$

where the overbar denotes the set-complementation. Alternative a may be regarded as probable to a degree which belongs to the interval $[Bel(a),Pl(a)]=[m(a),m(a)+m(\Omega)]$ while the probability of the opposite alternativeia lies in the interval $[1-m(a)-m(\Omega),1-m(a)]$. The quantity $m(\Omega)$ corresponds to the amount of ignorance. Note that in general we need two numbers, m(a) and $m(\Omega)$, for expressing the uncertainty pertaining to alternative a. However if $m(\Omega)=0$, the uncertainty of a is expressed by a single number, its probability ; when $m(a)+m(\Omega)=1$ or when m(a)=0, the useful information about an alternative lies in its degree of belief or in the degree of belief attached to the opposite alternative, respectively. In that case, the plausibility function is a possibility measure in the sense of Zadeh [6].

Indeed, when the collection of subsets A of Ω such that m(A)>0 is nested with respect to set inclusion, the belief and plausibility functions defined by (3) are respectively nothing but necessity and possibility measures in Zadeh's sense, see [3]. In the case of two alternatives a and $\neg a$, we are in such a situation as soon as $m(a)=0$ or $m(\neg a)=0$ since obviously $\{a\} \subset \Omega$ and $\{\neg a\} \subset \Omega$.

With $Bel_i(a)=m_i(a)=s_i$ and $Pl_i(a)=m_i(a)+m_i(\Omega)=t_i$, where i=1,2, Dempster's rule of combination yields

$$\begin{aligned} m(a) &= (s_1 t_2+s_2 t_1-s_1 s_2)/d \\ m(\neg a) &= (1-s_1-s_2+s_2 t_1+s_1 t_2-t_1.t_2)/d \\ m(\Omega) &= (t_1-s_1)(t_2-s_2)/d \end{aligned} \qquad (4)$$

with $d = 1-s_1(1-t_2)-s_2(1-t_1)$

The formulas (4) encompass 4 noticeable particular cases :

i) $s_i=t_i$. In this case we have just to combine two ordinary probability distributions since $m_i(\Omega)=0$. We get

$$m(a)=(s_1.s_2)/(1-s_1-s_2+2s_1.s_2) \; ; \; m(\neg a)=1-m(a) \; , \qquad (5)$$

a formula used by Kayser [2] for combining uncertain pieces of information, also a symmetric sum (Silvert, IEEE Trans. SMC,9, pp 657-659, 1979). When $s_1=\varepsilon$ and $s_2=1-\varepsilon$ (i.e. there is a conflict which is all the more severe as ε is nearer 0 or 1), we get $\forall \; \varepsilon>0$, $m(a)=m(\neg a)=\frac{1}{2}$. $\varepsilon=0$ is forbidden as well as $\varepsilon = 1$.

ii) $t_1=t_2=1$: In this case, both sources of information agree that the plausibility of the alternative a is equal to 1 and thus $m_1(\neg a)=m_2(\neg a)=0$. Then (4) gives

$$m(a) = s_1+s_2-s_1.s_2 \ ; \ m(\neg a) = 0 \quad (6)$$

We recognize the formula used in MYCIN [1] for combining two non-conflicting pieces of evidence.

iii) $t_1=1, s_2=t_2$: This case is an hybrid of i and ii. (4) gives

$$m(a)=s_2/(1-s_1(1-s_2)) \ ; \ m(\neg a)=1-m(a) \quad (7)$$

iv) $t_1=1, s_2=0$: In this case the basic probability assignment m_1 expresses that we believe in the completely plausible alternative a with the degree s_1 while on the contrary m_2 expresses a belief in $\neg a$ (considered as completely plausible since $s_2=0$) with a degree equal to $1-t_2$. We have thus a conflict between the two bodies of evidence. Formulas (4) give

$$m(a) = \frac{s_1 t_2}{1-s_1(1-t_2)} \ ; \ m(\neg a) = \frac{(1-s_1)(1-t_2)}{1-s_1(1-t_2)} \quad (8)$$

and $Pl(a)=m(a)+m(\Omega)= t_2 \ /(1-s_1(1-t_2))$ (9)

In these 4 particular cases, the basic probability assignments m_1 and m_2 are such that either $m_i(\Omega)=0$ then m_i is an ordinary probability allocation or $m_i(a)=0$ or $m_i(\neg a)=0$ (then Pl_i is a possibility measure). In cases i) and iii), the result of the combination is an ordinary probability allocation, and the weight of alternative a is always reinforced in iii). In case ii), two consistent possibility measures are combined and yield a possibility measure. Contrastedly, case iv) combines two conflicting possibility measures, and what is obtained is neither a probability measure nor a possibility measure. In other words, two numbers, distinct from 0 or 1, are needed here for representing the uncertainty pervading the result of the combination in case of a conflict.

N.B. If we extend the combination formula (6) of two ordinary probability distributions (Indeed (6) is nothing but $Prob(a)=Prob_1(a).Prob_2(a)/(Prob_1(a).Prob_2(a)+Prob_1(\neg a).Prob_2(\neg a))$)from scalar values to interval values of the form $[Bel_i(a),Pl_i(a)]$, then this kind of sensitivity analysis yields results which are generally more imprecise those obtained from Dempster's rule of combination in cases ii, iii or iv. For instance in situation iv, we get the whole interval [0,1] instead of the interval $[m(a),Pl(a)]$ defined by (8)-(9).

Since in cases ii and iv we deal with possibility measures, we may then also use a possibilistic rule of combination. For sake of brevity we only consider such a rule when Ω has two elements a and $\neg a$ in the following. In this particular case a possibility distribution is a pair of plausibility values $(Pl(a),Pl(\neg a))$ such that

$$\max(Pl(a),Pl(\neg a)) = 1 \quad (10)$$

(since $m(a)=0$ or $m(\neg a)=0$). Then the intersection of two possibility distributions, viewed as fuzzy sets [6],[3], is given by $(Pl(a),Pl(\neg a))=(\min(Pl_1(a),Pl_2(a))/d, \min(Pl_1(\neg a),Pl_2(\neg a))/d)$ with

$$d=\max(\min(Pl_1(a),Pl_2(a)),\min(Pl_1(\neg a),Pl_2(\neg a))) \quad (11)$$

The normalization in (11) maintains constraint (10). Then applying (11) to cases ii and iv respectively gives

ii) $t_1=Pl_1(a)=t_2=Pl_2(a)=1$

$$m(a)=1-Pl(\neg a)=1-\min(1-s_1,1-s_2)=\max(s_1,s_2) \quad (12)$$

since $s_i=1-Pl_i(\neg a)$; moreover $m(\neg a)=0$.

iv) $t_1=Pl_1(a)=1 \ ; \ s_2=1-Pl_2(\neg a)=0$

$$m(a)=1-\frac{\min(1-s_1,1)}{\max(t_2,1-s_1)} = \frac{\max(0,s_1+t_2-1)}{1-\min(s_1,1-t_2)} \quad (13)$$

$$m(\neg a) = 0 \ ; \ Pl(a)=t_2/(1-\min(s_1,1-t_2)) \quad (14)$$

The expressions (12), (13) and (14) have a structure similar to the one of formulas (6), (8) and (9) respectively. Indeed both the probabilistic sum and the operation maximum extend the disjunction in multiple-valued logic, while the product, the operation min and max(0,x+y-1) are the main extensions of the conjunction in this logic. What it is important in (13)-(14) is that the obtained result is of the same nature as the combined items, which was no longer the case with (8)-(9). Particularly, $m(\neg a)\neq 0$ in (8). Moreover if we replace the min operations in (11) by the product, then we recover (6) exactly, while (13) and (14) remain unchanged. The apparent advantage of (6) is that in case of several pieces of positive evidences (in favor of a) there is a reinforcement of our belief in a, due to the asymptotic behavior of (6) towards 1. However this reinforcement can be only justified with an independence hypothesis which is difficult to check ; thus (12) may appear as a more cautious rule.

Lastly, let us examine the difference of behaviors of (8)-(9) and of (13)-(14) in case of a strong symmetric conflict such as :

$m_1(a)=s$, $m_1(\neg a)=0$, $m_1(\Omega)=1-s$

$m_2(a)=0$, $m_2(\neg a)=s$, $m_2(\Omega)=1-s$. Then (8)-(9) yields

$$[Bel(a),Pl(a)]=[\frac{s}{1+s},\frac{1}{1+s}] = [Bel(\neg a),Pl(\neg a)] \quad (15)$$

which is an interval centered around $\frac{1}{2}$ and which shrinks as s becomes nearer to 1 (strong conflict). Contrastedly (13)-(14) then gives $[Bel(a),Pl(a)]=[0,1]$. At first glance the results look very different, but both of them express that a and $\neg$ a are equally believable, plausible, possible. Note that when the conflict is nothing but an absolute contradiction (s=1), (15) reduces to the value $\frac{1}{2}$ which is the probabilistic view of total ignorance.

2 - Highly improbable is not impossible

Let us now consider the particular case where we have three mutually exclusive alternatives in Ω, i.e. $\Omega=\{a,b,c\}$, in order to study the sensitivity of (2) to slight modifications of m_1 and m_2. First let us consider the situation discussed by Zadeh (AI Magazine,5(3),pp81-83,1984)

$$m_1(a)=0, \quad m_1(b)=k, \quad m_1(c)=1-k$$
$$m_2(a)=1-k, \quad m_2(b)=k, \quad m_2(c)=0 \quad ; \ k>0$$

Note that here m_1 and m_2 reduce to ordinary probability distributions on Ω. When k is a small number , there is a strong conflict between the two assignments since the first one regards a as completely impossible and c as almost certain, while the second one holds the opposite position ; both assignments only agree on the point that b is quite improbable. However, whatever the value of k, (2) yields $m(a)=0$, $m(b)=1$, $m(c)=0$ due to the normalization effect. Suppose now we modify m_1 and m_2 in the following way, ε being a very small amount :

$$m_1(a)=\varepsilon, \quad m_1(b)=k, \quad m_1(c)=1-k-\varepsilon$$
$$m_2(a)=1-k-\varepsilon, \quad m_2(b)=k, \quad m_2(c)=\varepsilon$$

. Then, (2) gives

$$m(a)=m(c)=\frac{\varepsilon(1-k-\varepsilon)}{k^2+2\varepsilon(1-k-\varepsilon)} \ ; \ m(b) = \frac{k^2}{k^2+2\varepsilon(1-k-\varepsilon)}$$

For k=0.1 and ε = 0.01 , we obtain $m(a)=m(c)\approx 0.32$; $m(b)=0.36$. For k=0.1 and $\varepsilon=0.001$, we obtain $m(a)=m(c)\approx 0.008$; $m(b)\approx 0.84$. Clearly, $\lim_{\varepsilon\to 0} m(a)=\lim_{\varepsilon\to 0} m(c)=0$ and $\lim_{\varepsilon\to 0} m(b)=1$. Thus we observe that when $m_1(a)$ and $m_2(c)$ are not strictly zero but remain small in comparison with k, which is itself small in comparison with $1-k-\varepsilon$, we may obtain results by (2), which are extremely different from the case where $m_1(a)=m_2(c)=0$. In this latter case m_1 (resp. m_2) expresses the complete certainty that a (resp. c) is impossible ; then the alternative b, although quite improbable, remains the only possible alternative. Contrastedly when a non-zero, even very small, value is assigned to an alternative, this alternative is not definitely ruled out although highly improbable ; it tacitely expresses that this value may be revised in the light of new information.

Thus assigning a zero value or a very small one may lead to very different conclusions in some instances.

Note that the acceptance of some unforeseeable outcome may be better expressed by assessing a small value ε' to the whole set Ω rather than assessing a non-zero value to each alternative in Ω. Indeed it is sometimes difficult to identifying them. Let

$m_1(a)=0$, $m_1(b)=k$, $m_1(c)=1-k-\varepsilon$, $m_1(\Omega)=\varepsilon$

$m_2(a)=1-k-\varepsilon$, $m_2(b)=k$, $m_2(c)=0$, $m_2(\Omega)=\varepsilon$, we get

$$m(a)=m(c)=\frac{\varepsilon(1-k-\varepsilon)}{d},\; m(b)=\frac{k^2+2k\varepsilon}{d},\; m(\Omega)=\frac{\varepsilon^2}{d}$$

with $d=k^2+2\varepsilon(1-\varepsilon)$. Note that we get results which only slightly differ from those obtained with $m_1(a)=m_2(c)=\varepsilon$ and $m_1(\Omega)=m_2(\Omega)=0$, which is satisfactory.

III - PROPAGATION PROBLEM

In the following some proposals are made in order to deal with the propagation problem in Shafer's framework. There already exists an interesting attempt (Lu et al. AAAI-84, pp216-221) but which seems in need of justifications.

A belief function always satisfies the following inequalities $\forall\ A\subseteq\Omega$, $\mathrm{Bel}(A\cap B)\geq \mathrm{Bel}(A)+\mathrm{Bel}(B)-\mathrm{Bel}(A\cup B)\geq \max(0,\mathrm{Bel}(A)+\mathrm{Bel}(B)-1)$ (16)

Applying (16) to $B=(\bar{A}\cup B)\cap(A\cup B)$, we get $\mathrm{Bel}(B)\geq \max(0,\mathrm{Bel}(\bar{A}\cup B)+\mathrm{Bel}(A\cup B)-1)\geq \max(0,\mathrm{Bel}(\bar{A}\cup B)+\mathrm{Bel}(A)-1)$. Thus since $\bar{A}\cup B$ is the material implication $A\to B$, we can prove the validity of

$$\begin{array}{ll} \mathrm{Bel}(A\to B) \geq r & \quad (17)\\ \mathrm{Bel}(A) \geq s & \\ \hline \mathrm{Bel}(B) \geq \max(0,r+s-1) & \end{array}$$

. This pattern can be proved equivalent to some in (Garvey et al., IJCAI-81, pp 319-325) , but is expressed here in terms of the implication connective. The lower bound of Bel(B) we have in (17) is the same as can be obtained using probabilities instead of belief functions. An improved lower bound is obtained when the belief function is a necessity measure, see [3]. Contrastedly, with probability, which corresponds to the remarkable particular case where plausibility function and belief function are equal, an improved lower bound cannot be obtained and this in spite of the additivity axiom. To the modus ponens-like pattern of reasoning (17) corresponds a modus tollens-like pattern of reasoning

$$\begin{array}{ll} \mathrm{Bel}(A\to B) \geq r' & \\ \mathrm{Pl}(B) \leq t & \quad (18)\\ \hline \mathrm{Pl}(A) \leq \min(1,1-r'+t) & \end{array}$$

combining (17) and (18) we obtain

$$\begin{array}{ll} \mathrm{Bel}(A\to B) \geq r & \\ \mathrm{Bel}(B\to A) \geq r' & \\ s \leq \mathrm{Bel}(A),\ \mathrm{Pl}(A) \leq t & \quad (19)\\ \hline \max(0,r+s-1)\leq \mathrm{Bel}(B),\ \mathrm{Pl}(B)\leq \min(1,1-r'+t) & \end{array}$$

Thus, if we characterize the uncertainty of the rule "if A, then B" by the two numbers $\mathrm{Bel}(A\to B)$ and $\mathrm{Bel}(B\to A)$, it is possible from a lower bound of the belief in A and from an upper bound of the plausibility of A to derive lower and upper bounds respectively of the belief in B and the plausibility of B. $\mathrm{Bel}(A\to B)$ may be viewed as a degree of sufficiency of having "A true" for deriving "B true", while $\mathrm{Bel}(B\to A)$ is a degree of necessariness of having "A true" for deriving "B true". Thus the pattern (19) provides a tool for using belief functions in deductive schemes.

Another approach, partially used in (Ginsberg, AAAI-84, pp 126-129), is to view belief and plausibility degrees as lower and upper bounds of probabilities and to perform a sensitivity analysis on the formula

$$\mathrm{Prob}(B)=\mathrm{Prob}(B\mid A).\mathrm{Prob}(A)+\mathrm{Prob}(B\mid \bar{A}).\mathrm{Prob}(\bar{A}) \quad (20)$$

This method is in the spirit of C.A.B. Smith's view [5] of lower and upper probabilities rather than Dempster/Shafer's. With $\mathrm{Prob}(A)\in[\mathrm{Bel}(A),\mathrm{Pl}(A)]=[s,t]$; $\mathrm{Prob}(\bar{A})\in[1-t,1-s]$; $\mathrm{Prob}(B\mid A)\in[\underline{r},\bar{r}]$ and $\mathrm{Prob}(B\mid\bar{A})\in[\underline{r}',\bar{r}']$, we get

$$\mathrm{Prob}(B)\in[\min(\underline{r}s+\underline{r}'(1-s),\underline{r}t+\underline{r}'(1-t)),\ \max(\bar{r}s+\bar{r}'(1-s),\bar{r}t+\bar{r}'(1-t))] \quad (21)$$

When $\mathrm{Prob}(B\mid\bar{A})$ is completely unknown, i.e. $[\underline{r}',\bar{r}']=[0,1]$ (21) reduces to $\mathrm{Prob}(B)\in[\underline{r}s,1-s+\bar{r}s]$ (22)

Note that the information pertaining to the uncertainty of the rule "if A, then B" is represented here in terms of conditional probabilities by the 4 numbers $\underline{r}$, $\bar{r}$, $\underline{r}'$, $\bar{r}'$. This approach seemingly provides stronger results than the previous one since the lower bound obtained in (22) is greater than the one obtained by (19).

However, the upper bounds in (19) and (22) are of different nature, since in (19) the upper bound is obtained from lower bounds on $\mathrm{Bel}(B\to A)$ and Bel(A), while in (22) the upper bound is computed from lower bounds on $\mathrm{Prob}(\bar{B}\mid A)$ and Prob(A). However lower bounds in (19) and (22) look similar, since in both cases it is a conjunctive combination (in different multi-valued logics) of lower bounds on Bel(A) and $\mathrm{Bel}(A\to B)$ in (19), and of lower bounds on Prob(A) and $\mathrm{Prob}(B\mid A)$ in (22). Upper bounds correspond to multivalued implications. Thus the differences between (19) and (22) are due to different views of the rule (in terms of implication or in terms of conditionning) and to the absence of information in (22) regarding the uncertainty of the rule "if B, then A" or if we prefer "if not A, then not B". Besides from $\mathrm{Bel}(A\to B)\geq r$ and $\mathrm{Bel}(A)\geq s$ we can establish that $\mathrm{Bel}(A\to\bar{B})\leq 2-(r+s)$; but from this latter inequality and $\mathrm{Bel}(A)\geq s$ no non-trivial lower bound on $\mathrm{Bel}(\bar{B})$ (i.e. no upper bound on Pl(B)) can be obtained. Thus we need the complementary information on $\mathrm{Bel}(B\to A)$ for getting an upper bound on Pl(B) in (19).

Lastly we may think of introducing conditional belief and plausibility functions for dealing with the propagation problem. This might be done using a conditional basic probability assignment (Ishizuka et al., Inf.Sci., 28 (1982) 179-206).

IV - CONCLUDING REMARKS

The intended purpose of this paper was a careful examination of the behavior of Dempster's rule of combination in typical cases. It appears that assigning an extremely small value rather than a zero value may lead to very different results. Moreover Dempster's rule applies to situations which are probabilistic in nature ($m(\Omega)=0$ in case of two alternatives) as well as to situations which are possibilistic in nature ($m(a)+m(\Omega)=1$ or $m(\neg a)+m(\Omega)=1$ in case of two alternatives) where a possibilistic rule of combination might be more suitable. Dempster's rule of combination remains an appealing tool but it should not be used in a blind manner without caution. Besides we discussed two ways of dealing with the propagation problem using belief functions. These two approaches are not equivalent and we have to take care of the intended meaning of the numbers we use for encoding the uncertainty. Particularly we may think of an "if...then..." rule in terms of an implication or in terms of conditioning. Moreover viewing plausibility degrees as upper and lower bounds of probability degrees and then using a sensitivity analysis, or directly dealing with degrees of belief do not lead to the same result. Lastly, the two approaches to propagation which were briefly discussed are not learning processes as Bayesian inference is ; the extension of Bayesian inference in the Dempster/Shafer framework is still a topic for further research

[1] Buchanan B., Shortliffe E. : Ruled-Based Expert Systems, Addison-Wesley, 1984.

[2] Kayser, D. see ref. [117] in [3].

[3] Prade, H. IEEE Trans. PAMI, 7, 1985, 260-283.

[4] Shafer, G. A Mathematical Theory of Evidence, Princeton, 1976.

[5] Smith, C. J. Roy. Statist Soc., A-128, 1965, 469-499

[6] Zadeh, L. Fuzzy Sets & Systems, 1, 1978, 3-28.