



Possibility theory and data fusion in poorly informed environments

Didier Dubois, Henri Prade

► To cite this version:

Didier Dubois, Henri Prade. Possibility theory and data fusion in poorly informed environments. Control Engineering Practice, 1994, 2 (5), pp.811–823. 10.1016/0967-0661(94)90346-8 . hal-04054185

HAL Id: hal-04054185

<https://hal.science/hal-04054185>

Submitted on 3 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Possibility theory and data fusion in poorly informed environments

Didier DUBOIS and Henri PRADE

*Institut de Recherche en Informatique de Toulouse (I.R.I.T.) – C.N.R.S.
Université Paul Sabatier, 118 route de Narbonne
31062 Toulouse Cedex, France*

ABSTRACT

Some data fusion problems seem to be naturally handled in the framework of possibility theory. As an example, the problem of modelling expert knowledge about numerical parameters in the field of reliability is reconsidered in that framework. Usually expert opinions about quantities such as failure rates are modelled, assessed and pooled in the setting of probability theory. In this paper, we formulate a model of expert opinion by means of possibility distributions that are thought to better reflect the imprecision pervading expert judgments. They are weak substitutes to unreachable subjective probabilities. Assessment evaluation is carried out in terms of accurateness and level of precision, respectively measured by membership grades and fuzzy cardinality indices. Lastly, elaborating from previous works on data fusion using possibility theory, we present various pooling modes with their formal model under various assumptions concerning the sources of information. This framework is particularly suitable when sources are heterogeneous and statistical data are not available.

Key Words. Possibility theory; fuzzy set theory; data fusion; expert opinion pooling

1. INTRODUCTION

The problem of combining pieces of evidence issued from several sources of information can be encountered in various fields of application, particularly in i) sensor fusion, i.e., when pieces of information coming from different sensors are to be aggregated, ii) multiple source interrogation systems where each of the sources can provide precise, imprecise or uncertain information about values of interest, iii) expert opinion pooling, when different individual statements have to be synthesized. Our basic claim is that there cannot be a unique mode of combination, which would be satisfactory in any situations, even when the framework for representing information is chosen.

Various combination problems exist, especially, i) preference aggregation versus information aggregation and ii) the combination of information coming from parallel sources versus the revision of already available information. In the preference aggregation problem it makes sense to find the opinion of the "average man" in an homogeneous group of individuals, to look for trade-offs between preferences, while on the contrary, if the information aggregation is a matter of truth and reliability, logical combinations are natural candidates. In this latter case conjunctive combinations apply when all the sources are reliable, while disjunctive combinations deal with the case of unreliable sources hidden in a group of other reliable ones. Obviously weighted logical combinations may be considered in particular when the sources are not equally reliable. Averaging operations in information aggregation can be justified when the set of sources can be viewed as a single random source producing different inputs. In that case, indeed, the set of data to be fused can be interpreted as standard statistics. For instance several successive measurements from a single sensor can be viewed as the result of a random experiment. Then the discrepancies between the sources can be explained in terms of random variability. However in the case of unique measurements issued from distinct sensors, or in the case of expert opinions, it is not clear that averaging combination modes make sense. Besides we oppose the case of merging information from *parallel* sources to the problem of belief revision where sources do not play a symmetrical role. In the first situation, all the sources provide information simultaneously, while in the revision process there is a chronological ordering between the source which represents the present state of belief and the source which issues the new information. In each case the pooling obeys different requirements, for instance belief revision is generally not commutative.

In this paper we consider the information aggregation problem in the case of distinct parallel sources, and with no a priori knowledge. The typical example of such a problem is the pooling of expert judgments in the field of reliability. In the field of reliability and safety analysis of newly designed installations, statistical data are not always available, especially regarding rare or destructive events, or for devices whose novelty implies a scarcity of experimental data at the time when the safety analysis must be carried out. In such cases the knowledge of experts is very useful for evaluating unknown parameters, typically the number of times a given event occurs within a given period, or the number of hours needed to repair some equipment, etc.

In order to get useful information from the experts, several problems must be solved. The first one is a proper modeling of the pieces of data supplied by a single expert about a given parameter. This type of data is almost never precise and reliable because the expert only possesses a rough idea of the value of quantitative parameters, due to limited precision of human assessments and to the variability of such values (e.g., failure rates). In most studies the expert's response is represented by a probability distribution. The reason for that is often that probability theory is the only well-known framework for modelling uncertainty. In this paper we argue that a pure probabilistic model of expert knowledge is not so satisfactory, and that possibility theory is a more natural framework. The second task to be solved is to assess the quality of the expert, namely his calibration, and the precision of his response. In the case of probabilistic modeling, scoring rules have been devised for this purpose. See Cooke (1991) for a survey. Here we suggest a rating index which may act as a scoring rule for the case of possibilistic modeling. Lastly when several expert responses are available, they must be combined so as to yield a unique, hopefully better response. This problem is also addressed here.

The next section contains a short discussion of the probabilistic approach and its limitations when pooling evidence stemming from heterogeneous incompletely informed sources. Section 3 addresses the representation of expert imprecise data by means of possibility distributions. Section 4 describes numerical indices for expert response assessment and outlines an experiment that aims at assessing the quality of experts. Section 5 deals with the pooling of uncertain pieces of information. This paper borrows from a case study whose preliminary results appear in (Kalfsbeek, 1990; Dubois and Kalfsbeek, 1990) and in more details in (Sandri, 1990, 1991; Sandri et al., 1993). It also freely uses material from a previous survey paper (Dubois and Prade, 1992b) on possibilistic combination rules.

2. SOME LIMITATIONS OF THE PROBABILISTIC APPROACH

Typically, an expert is asked about the value of some parameter x by specifying fractiles of a probability distribution function (pdf), for instance the 5% and the 95% fractiles. In other words, the expert supplies values x_e and x_u such that $P(x \leq x_e) = 0.05$ and $P(x \leq x_u) = .95$ respectively. Moreover some information about the mode, or the mean, or else the median of the distribution is often asked. Based on these values, and on the choice of a parametrized family of distribution functions, (for instance a beta-distribution), a given distribution function is chosen that supposedly best represents the available information. The nature of the probabilities appearing during the elicitation is sometimes controversial. Some authors call them "subjective probabilities". This term is ambiguous insofar as it may mean the numerical estimate of a feeling of certainty, or a subjectively assessed objective frequency. In reliability applications the second interpretation sounds more instrumental. Note that in some approaches, experts are only asked for point values of x , as in the Bayesian method of Mosleh and Apostolakis (1984).

When several experts supply this kind of information, their responses are pooled so as to derive a single distribution that reflects the opinion of the group. However it is clear that the opinion of reliable experts should be more important than the one of unreliable ones. There are two main approaches to the pooling of probability distributions: the consensus method justified by Wagner and Lehrer (1981), used by Cooke (1988), and the Bayesian approach, exemplified by the works of Mosleh and Apostolakis (1984). In the consensus method each expert E_i supplies a pdf p_i , and the resulting distribution is a weighted average $p = \sum_i w_i p_i$ where the weights w_i reflect the reliability of experts. Cooke (1988) has developed a theory of weights that act as scoring rules. They tend to force experts to be calibrated and informative. In the Bayesian method, the a priori opinion of the analyst about the true value of x is revised on the basis of expert opinions, expressed as point-values. The credibility of experts, from the standpoint of the analyst, is modelled by conditional probabilities of what an expert will claim the true value of x is, given this true value. Once the expert point-values are known, the a priori probability distribution of x , as possessed by the analyst is revised through Bayes theorem. The model by Mosleh and Apostolakis (1984) tries to account for the dependence between experts via a correlation coefficient.

The probabilistic approaches can be criticized for several reasons:

- First, the identification of a probability distribution requires more information than what an expert is able to supply. There are many distribution functions corresponding to given .5 and .95 fractiles and prescribed mode. The choice of a parametrized family of distribution functions is basically a matter of making calculations simple. As a consequence the faithfulness of the modeling of the expert opinion can be questioned. The expert is not asked about the shape of his pdf, and anyway he might be at a loss to produce any one.
- Experts better supply intervals rather than point-values because their knowledge is not only of limited reliability but also tainted with imprecision. Probability theory is not concerned with the modeling of imprecision but rather captures the notion of random variability, or in the case of subjective probabilities, betting odds.
- The consensus method has a basic flaw in the context of reliability: it is a voting-like procedure. Indeed if two reliable experts have conflicting opinions about the value of x , such that one gives a small value to x , and the other gives a high value to x , the consensus method will supply a probability distribution whose mean value is medium, i.e., a value on which both expert agree as not being the true one. What is needed is a method which in the best case, guesses the true value and discards the wrong expert, or in the worst case, proposes a cautious response that fits the available data (e.g., x is either small or large, but certainly not medium). The weighted average method sounds more natural when expert opinions express preference but does not seem to be adapted when a true answer is to be determined instead of a preferred one.
- Another interpretation of the consensus method comes down to considering the expert data as issued from a single random source. The underlying homogeneity assumption is questionable in the case of experts some of which are wrong, or multiple sources, some of which are erratic or erroneous. Moreover the weighted average method may affect the variance, in the

sense that the variance of the result may become smaller than the one of any input distributions. This phenomenon is acceptable in the scope of independent experts. However experts often share a lot of technical background, and the expert independence assumption which underlies the single random source model is highly questionable.

- The main drawback of the Bayesian method seems to be, as usual, the need for a priori knowledge about the value of x . In other words the analyst that looks for expert advice must be an expert himself. However in many cases the analyst has no idea about the value of x and all that he may learn is about the reliability of experts, by a technique as the one by Cooke (1988), that inspired the one described below. The Bayesian method cannot update from the state of complete ignorance. Moreover, in the case of several conflicting experts, voting-like effects resulting in values that no expert supplies can be observed with the Bayesian method (see Dubois and Prade, 1992a).

The possibilistic approach that is proposed in this paper tries to cope with most of the difficulties faced by the probabilistic approach on the problem of pooling imprecise information coming from heterogeneous sources; its main features are: faithfulness of the representation of subjective data, no need for a priori knowledge, and a variety of pooling methods whose choice depends about the reliability of experts or sources and the level of conflict between their opinions.

3. IMPRECISE DATA AS POSSIBILITY DISTRIBUTIONS

A possibility distribution π_x (Zadeh, 1978) can be viewed as the membership function of the fuzzy set of possible values of a quantity x . These values are assumed to be mutually exclusive, since x takes on only one value (its true value), that belongs to a set S which is considered here to be a closed, bounded real interval $[s_0, s_m]$. The actual value of x is unknown. But, since one of the elements of S is the true value of x , $\pi_x(s^*) = 1$ for at least one value $s^* \in S$. This is the normalization condition that claims that one value is viewed as totally possible. When unique, s^* is the most plausible value of x . If s and s' are such that $\pi_x(s) > \pi_x(s')$, s is considered a more plausible value than s' . When $\pi_x(s) = 0$, s then x cannot take on value s .

Knowing a possibility distribution, the likelihood of events can be described by means of two set-functions: the possibility measure and the necessity measure (Dubois and Prade, 1988b), denoted Π and N respectively. When π is the membership function of a crisp set E , it expresses that the value of parameter x certainly belongs to E . Then, an event B is said to be possible if and only if $E \cap B \neq \emptyset$, and certain if and only if $E \subseteq B$; by definition we let $\Pi(B) = 1$ and $N(B) = 1$ in these respective situations. When $\Pi(B) = 1$, there is a value s of x in B , considered as totally possible (since B and E intersect). When $N(B) = 1$ no possible value of x lies outside B and it is thus certain that x lies in B . In the general case where π_x is the membership function of a fuzzy set, the possibility and necessity measures are defined as follows.

$$\Pi(B) = \sup_{s \in B} \pi_x(s) \quad (1)$$

$$\begin{aligned} N(B) &= \inf_{s \notin B} 1 - \pi_X(s) \\ &= 1 - \Pi(\bar{B}) \end{aligned} \quad (2)$$

where \bar{B} is the complement of B with respect to S . This duality expresses that B is all the more certain as \bar{B} is impossible. These evaluations only exploit the ordinal part of the information, not its quantitative contents since only the plausibility ordering is involved in (1) and (2).

The simplest form of a possibility distribution on S is a membership function of a sub-interval $[L, U]$ of S , i.e., $\pi_X(s) = 1$ if $x \in [L, U]$, 0 otherwise. This type of possibility distribution is naturally obtained from experts claiming that "x lies between L and U". This way of expressing knowledge is more natural than giving a point-value s^* for x right away, because it allows for some imprecision: the true value of x is more likely to lie between L and U than to be equal to s^* . Allowing for imprecision reduces uncertainty. However this representation is not entirely satisfactory for two reasons. First, claiming that $\pi_X(s) = 0$ for some s means that $x = s$ is impossible. This is too strong for the expert who is then tempted to give wide uninformative intervals (e.g., $L = s_0$, $U = s_m$). Moreover, it is more satisfactory to ask from the expert several intervals with various levels of confidence, and to admit that even the widest, safest intervals do not rule out some residual uncertainty, denoted e , that the value of x lies outside.

A possibility distribution π_X can represent a finite family of nested confidence ("focal") subsets $\{A_1, A_2, \dots, A_m\}$ where $A_i \subset A_{i+1}$, $i = 1, m-1$ as soon as the set of possibility values $\{\pi(s) \mid s \in S\}$ is finite. Each confidence subset A_i is attached a positive confidence level λ_i . The level of confidence λ_i can be conveniently interpreted as a lower bound on the probability that the true value of x hits A_i . The links between the confidence levels λ_i 's and the degrees of possibility are defined by postulating $\lambda_i = N(A_i)$ the degree of necessity of A_i (Dubois and Prade, 1988b, 1992c). This entails that $\lambda_1 \leq \dots \leq \lambda_m$ due to the monotonicity of N . The possibility distribution equivalent to the family $\{(A_1, \lambda_1), (A_2, \lambda_2), \dots, (A_m, \lambda_m)\}$ is defined as the least specific (see, e.g., Yager, 1992) possibility distribution π that obeys the constraints $\lambda_i = N(A_i)$, $i = 1, n$ as per equation (2). It comes down to maximizing the degrees of possibility $\pi(s)$ for all s in S , subject to these constraints. The solution is unique and is

$$\begin{aligned} \forall s, \pi_X(s) &= \min_i \max(1 - \lambda_i, \mu_{A_i}(s)) \\ &= \begin{cases} 1 & \text{if } s \in A_1 \\ \min_{i: s \notin A_i} (1 - \lambda_i) & \text{otherwise} \end{cases} \end{aligned} \quad (3)$$

where μ_{A_i} is the characteristic function of A_i . This solution is the least committed one with respect to the available data, since by allowing the greatest possibility degrees in agreement with the constraints, it defines the least restrictive possibility distribution. Conversely, the family $\{(A_1, \lambda_1), (A_2, \lambda_2), \dots, (A_m, \lambda_m)\}$ of confidence intervals can be reconstructed from the possibility distribution π_X . Namely it can be proved that if the set of possibility values $\pi_X(s)$ is $\{\alpha_1 = 1, \alpha_2 \geq \alpha_3 \dots \geq \alpha_m\}$, and letting $\alpha_{m+1} = 0$ we have

$$A_i = \{s \mid \pi_x(s) \geq \alpha_i\} ; \lambda_i = 1 - \alpha_{i+1}, \forall i = 1, m. \quad (4)$$

In particular $\lambda_m = 1$ and A_m is the subset which for sure contains x ; we may have $A_m = S$ if no strict subset of S can be ascertained as including x . Note that there is a set of weights p_1, p_2, \dots, p_m summing to one, such that (Dubois and Prade, 1988a)

$$\forall s, \pi_x(s) = \sum_{i: s \in A_i} p_i. \quad (5)$$

Namely $p_i = \alpha_i - \alpha_{i+1}$, $\forall i = 1, m$. Hence the possibility distribution can be cast in the setting of random sets. From a mathematical point of view, the information can be viewed as a nested random set $\{(A_i, p_i), i = 1, m\}$, which allows for imprecision (the size of the A_i 's) and uncertainty (the p_i 's). And p_i is the probability that the expert would supply exactly A_i as a faithful representation of his knowledge of x (it is *not* the probability that x belongs to A_i). In such a context $\Pi(B)$ and $N(B)$ are expected possibility and certainty degrees in the usual probabilistic sense since we have:

$$\Pi(B) = \sum_{i=1, m} p_i \Pi_i(B) \quad (6)$$

$$\begin{aligned} N(B) &= \sum_{i=1, m} p_i N_i(B) \\ &= 1 - \Pi(\bar{B}) \end{aligned} \quad (7)$$

where Π_i and N_i are the $\{0,1\}$ -valued possibility and necessity measures induced by the set A_i (i.e., $\Pi_i(B) = 1$ iff $A_i \cap B \neq \emptyset$, $N_i(B) = 1$ iff $A_i \subseteq B$).

The expert is supposed to be capable of supplying several nested intervals A_1, \dots, A_m directly, together with levels of confidence $\lambda_1, \dots, \lambda_m$ (e.g., from the point of view of the expert, the best lower bound of the proportion of cases where $x \in A_i$ from his experience). In practice (Sandri et al., 1993), three intervals only have been kept and the confidence levels have been predefined: A_1 with $\lambda_1 = 0.05$, A_2 with $\lambda_2 = 0.5$, and A_3 with $\lambda_3 = 0.95$. A_1 corresponds to the "usual values" of x , and there is a 0.05 probability ($= e$) that x misses A_3 , i.e., the residual uncertainty of the conservative evaluation. Finally, the focal subset $A_m = A_4$ (with $\lambda_4 = 1$) is always S itself, due to the residual uncertainty. The following table summarizes the data supplied by one expert.

The three first lines of Table 1 correspond to specific questions asked to experts (see Kalfsbeek, 1990). Although no exact counterparts of intervals $[C_e, C_u]$, $[M_e, M_u]$, $[L, U]$ are actually used in the probabilistic approaches, these intervals can be interpreted in terms of fractiles of a probability distribution, e.g., $[L, U]$ corresponds to the range between the 2.5% and the 97.5% fractiles. In terms of fuzzy sets, $[C_e, C_u]$ corresponds to the core of the fuzzy set with membership function π_x since $\forall s \in [C_e, C_u], \pi_x(s) = 1$. The obtained possibility distribution is pictured in Figure 1. The possibilistic representation is more faithful than the use of a single probability distribution since it exactly accounts for the supplied data and encompasses a family

of probability distributions (due to equating necessity degrees and lower probability bounds). Possibility distributions are here simple ways of encoding sets of probability distributions when one wishes to refrain from selecting one (Dubois and Prade, 1992c). See Coolen (1994) for the use of imprecise probabilities in the modeling of expert opinions.

A ₁	[C_e, C_u]	0.05	1	0.05
A ₂	[M_e, M_u]	0.5	0.95	0.45
A ₃	[L, U]	0.95	0.5	0.45
A ₄	S	1	0.05	0.05
	selected intervals	levels of confidence λ_i	degrees of possibility α_i	weights p_i

Table 1 Data supplied by experts
(L, U, M_e, M_u, C_e, C_u) (in the bold-face rectangle)

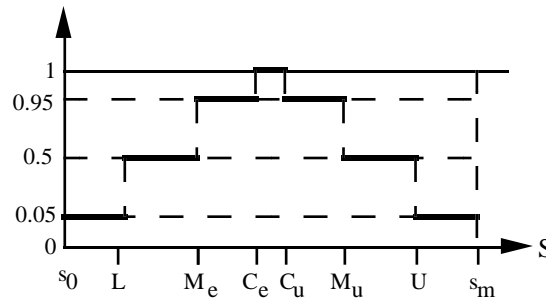


Fig.1. Expert-originated possibility distribution

4. ASSESSMENT OF EXPERT JUDGMENT

In order to assess the quality of experts, the latter are asked questions whose answers are known, and they are rated on the basis of these results. The questions pertain to the true values of a series x_1, x_2, \dots, x_n of "seed" variables; the values of these parameters are either known by the analyst and not known by the experts, or more often can be determined afterwards by means of physical experiments, or other means. In order to build a meaningful rating system, one must first identify the type of deficiencies experts may be prone to, and then define indices that enable the true answer and the expert answer to be compared and take these deficiencies into account. This program has been carried out in the probabilistic setting by Cooke (1991). Note that the true value of a seed-variable may be ill-known itself, sometimes because the state-of-the-art in the field does not allow for its precise evaluation, or because the available information consists of some histogram.

Experts can be deficient with regard to three aspects

- *inaccuracy*: values given by the expert are inconsistent with the actual information about the seed variables. For instance, he always underestimates the true value. The expert is then said to be inaccurate.
- *overcautiousness*: the expert is too cautious because the intervals he supplies are too large to be informative, although he is not inaccurate. The expert is said to be underconfident.
- *overconfidence*: the value of the seed variable is not precisely known to-date but the expert supplies intervals that are too narrow (or point-values). He is then said to be overconfident.

In order to build scoring indices that reflect these issues, let us first consider a seed variable x whose value s^* is precisely known, and let E be the fuzzy set supplied by the expert, in order to describe his knowledge about x . Let μ_E be the membership function of E (so that $\mu_E = \pi_x$). In this situation overconfidence does not appear. It is easy to see that

- the greater $\mu_E(s^*)$, the more accurate is the expert. Indeed if $\mu_E(s^*) = 0$, E totally misses s^* while if $\mu_E(s^*) = 1$, s^* is acknowledged as a usual value of x . Hence $\mu_E(s^*)$ is a natural measure of accurateness.
- if E is a crisp interval $[a, b]$ the wider E , the more imprecise (hence under-confident) the expert. The width of E is then $|E| = b - a$. When E is fuzzy the width of E is generalized by

$$|E| = \int_S \mu_E(s) \, ds. \quad (8)$$

This is a generalized fuzzy cardinality (where cardinality is changed into the Lebesgue measure). Other extended cardinalities exist to evaluate imprecision (see Klir and Folger, 1988). In our situation, where E is a finite nested random set, the equality $|E| = \sum_{i=1, m} |A_i| p_i$ can be established. This evaluation must be rescaled so as to account for the ever present residual uncertainty $e = p_m$ and so that it rates 1 when $L = U$ (precise response) and 0 when $L = s_0$, $U = s_m$ (empty response). A reasonable underconfidence index is then

$$f(|E|) = \frac{|S| - |E|}{(1 - e) \cdot |S|}. \quad (9)$$

On the whole, the overall rating of the expert regarding a single seed-variable can be defined as

$$Q(E, x) = \mu_E(s^*) \cdot f(|E|) \quad (10)$$

that requires him to be both accurate and informative in order to score high. One may admit that by convention, $Q(E, x) = 0$ if $s^* \notin [L, U]$ instead of $e \cdot f(|E|)$, if we do not want to account for the residual uncertainty in the calibration index. In the probabilistic approach (Cooke, 1991) the expert data is modelled by means of a probability distribution function, and the rating of the expert combines a degree of calibration and a degree of informativeness. The latter is based on entropy instead of cardinality. Probabilistic calibration is quite different from accurateness because it is a conjoint evaluation pertaining to several seed variables simultaneously while accurateness deals with single seed variables. Calibratedness can go along with significant inaccuracy (see Sandri et al., 1993)

When the seed variable is not precisely known, the index $Q(E, s^*)$ can be extended as follows:

- if the seed variable value is described by a histogram leading to a probability distribution P then

$$Q(E, x) = P(E) \cdot f(|E|) \quad (11)$$

where $P(E)$ is the probability of the fuzzy event E (Zadeh, 1968), i.e., $P(E) = \int_S \mu_E(s) dP(s)$.

- if the seed variable is described by a possibility distribution $\pi_x^* = \mu_F$ then

$$Q(E, x) = \Pi^*(E) \cdot f(|E \Delta F|) \quad (12)$$

where Π^* is the possibility measure attached to π_x^* and Δ is the symmetric difference of fuzzy sets. More specifically, $\Pi^*(E) = \sup_S \min(\pi_x^*(s), \mu_E(s))$ is the possibility of the fuzzy event E (Zadeh, 1978), and $\mu_{E \Delta F}(s) = |\mu_E(s) - \mu_F(s)|$ (see Dubois and Prade, 1988b). $\Pi^*(E)$ evaluates to what extent the expert's response is consistent with the available information about x , and $f(|E \Delta F|)$ penalizes both underconfidence and overconfidence of the expert. When the possibility (or the probability) distribution of x reduces to deterministic information ($x = s^*$) then the above indices collapse into the first definition when the support of E is a continuum.

Thus an expert j is rated by the set $\{Q(E_{ji}, x_i) \mid i = 1, m\}$ of evaluations. Ranking of experts can be obtained based on the average rating of each expert. The standard deviation is also useful to check the significance of the gaps between average ratings of experts. Based on these evaluations a set K of experts can be divided into groups of unequal reliability. Moreover the fuzzy set R of reliable experts can be defined by the membership function

$$\mu_R(j) = 1/m \cdot \sum_{i=1, m} Q(E_{ji}, x_i), j = 1, k$$

if there are k experts. The cardinality of R , say

$$|R| = \sum_{j=1, k} \mu_R(j)$$

gives a good idea of the number of reliable experts in the group. However it is not clear that the coefficients $\mu_R(j)$ have probabilistic significance such as the probability that the expert is right. Rather, these coefficients indicate a ranking of experts, suggesting who seem to be the most credible ones in the concerned domain.

5. THE POOLING OF POSSIBILITY DISTRIBUTIONS

5.1. A Logical View of the Combination of Uncertainty

The general ideas of the possibilistic approach to the aggregation of information issued from several distinct sources are first that there is no unique combination mode, and second that the choice of the combination mode depends on an assumption about the reliability of sources, as formulated by some analyst. No a priori knowledge about the inquired parameter is needed, and the sources, viewed in parallel, are to be combined in a symmetric way if all sources are equally reliable. There are basically two extreme modes of symmetric combination: the conjunctive mode when all sources agree and are considered as reliable; and the disjunctive mode when sources disagree so that at least one of them is wrong. These modes are implemented respectively in terms of fuzzy set intersections and fuzzy set unions. Indeed in order to clarify the situation it is useful to look at combination from a set-theoretic (or logical) point of view (see, e.g., Dubois and Prade, 1988a). Set-theoretic operations such as unions and intersections (expressing disjunction and conjunction) are the basic consensus rules in set theory. Consider the case of two sources of information giving evidence about the value of some variable x under the form of a set. Namely

Source 1 $x \in E_1 \subseteq S$

Source 2 $x \in E_2 \subseteq S$

What should be said about x ?

We claim that the choice of a type of pooling rule that solves this problem is a matter of context. There is no theory liable of prescribing a universal pooling method for these two pieces of information, that would apply to all situations. The chosen pooling scheme depends not only of formal properties the method should fulfil, but also on how much the sources agree, and what is known about their reliability. It means that a panoply of combination schemes together with underlying assumptions must be laid bare and used in an adaptive way. In the above case if $E_1 \cap E_2$ is not empty the assumption that both sources overlap is tenable and the combination rule can be the intersection, i.e., $x \in E_1 \cap E_2$. Otherwise at least one of the sources is necessarily wrong, and a more natural combination rule is the union, i.e., $x \in E_1 \cup E_2$, which assumes that the other source is not wrong. This rule does not require to identify what is the right source. This view, introduced by Dubois and Prade (1987) has been applied to multiple source interrogation systems (Sandri et al., 1989), to expert opinion pooling (Sandri et al., 1993) and to data fusion in computer vision (Deveughèle and Dubuisson, 1993). In the case of expert knowledge the pooling mode depends upon the results of the assessment step and the extent to which experts responses on the inquired parameter value agree with one another.

5.2. On Some Properties of Combination Rules

Several authors (e.g., Cheng and Kashyap, 1989; Hajek, 1985), not necessarily in the framework of fuzzy set and possibility theory, have discussed combination operations in terms of requested algebraic properties only. Indeed, a combination law looks all the better if it possesses such nice algebraic properties. Let us discuss them briefly.

Closure: The closure property is one that is often used without being explicitly stated. It says that if we use some representation framework then the result of the combination should also belong to that framework. For instance any probability theory tenant would assume that pooling two probability measures should produce a probability measure. Similarly, in proposing fuzzy set-theoretic operations, Zadeh (1965) took the natural requirement that the intersection or the union of two fuzzy sets is still a fuzzy set. This kind of closure assumption is natural once we want to stay within a given mathematical framework.

Note that the closure property can be expressed at two levels in the possibilistic framework, namely pointwisely on the possibility distributions (or if we prefer on the fuzzy sets) or eventwisely on the possibility (or equivalently on the necessity) measures. Both closure properties are not equivalent as shown by the following result (Dubois and Prade, 1990). The only way of combining possibility measures Π_1, \dots, Π_n into a possibility measure Π , in an eventwise manner (i.e., $\forall A, \Pi(A) = f(\Pi_1(A), \dots, \Pi_n(A))$), is a max-combination of the form

$$\forall A, \Pi(A) = \max(f_1(\Pi_1(A)), \dots, f_n(\Pi_n(A))) \quad (13)$$

where f_i is a monotonically increasing function such that $f_i(0) = 0, \forall i$ and $\exists j, f_j(1) = 1$ which modifies the shape of the possibility distribution π_i underlying Π_i . An example of eventwise possibility consensus function is the weighted maximum operation, i.e.

$$\forall A, \Pi(A) = \max_{j=1,n} \min(\lambda_j, \Pi_j(A)) \quad (14)$$

with $\max_{j=1,n} \lambda_j = 1$, where λ_j represents the relative importance of the source yielding Π_j ; see (Dubois and Prade, 1986) on weighted maximum and minimum operations. However, in (14), the minimum can be changed into a product, or into the linear operation $\max(0, a + b - 1)$, and more generally into any operation $*$ increasing in both places and such that $1 * 1 = 1, 0 * 1 = 0 = 1 * 0$. In fact, the weighted max-combination is the counterpart in possibility theory of the linear convex combination in probability theory; the weighted max-combination can be interpreted in the possibilistic framework as a median value, just as the convex combination can be interpreted in terms of probabilistic expectation. Note that the eventwise closure property may invalidate useful combination modes such as set intersection, in the above elementary example, since only the set union is coherent with (14). The eventwise combination rule (13) expresses that what can be "logically" inferred from a set of distinct knowledge bases, is only the common part which can be inferred from each of them, taking into account the relative reliability of each base (Dubois et al., 1992). Indeed $x \in E_1 \cup E_2$ is a valid conclusion which

can be both derived from a source asserting $x \in E_1$ and from a source asserting $x \in E_2$ ($E_1 \cup E_2$ is even the smallest set which contains both E_1 and E_2).

Commutativity and associativity: Commutativity is good when sources of information are exchangeable. Associativity is not absolutely required. For instance the averaging combination rule underlying the Kalman filtering technique is not right away associative since it accounts for an ordering of the sources. A weaker property such as quasi-associativity is often sufficient (Yager, 1987): a combination operation f is quasi-associative if and only if there is an associative operation $*$ and an invertible function ϕ such that $f(a_1, a_2, \dots, a_n) = \phi(a_1 * a_2 * \dots * a_n)$. Then the main advantage of associativity, i.e., modularity of the combination when integrating the information from a $(n + 1)^{\text{th}}$ source, remains. Namely if $\hat{a} = f(a_1, a_2, \dots, a_n)$ then $f(a_1, \dots, a_n, a_{n+1}) = \phi(\phi^{-1}(\hat{a}) * a_{n+1})$. This is what happens in Kalman filtering-based fusion where pieces of information arrive in sequence, and are absorbed one after the other.

Idempotence: That means that if two sources supply the same data the result of the combination should not alter these data. This property is not always to be accepted. Indeed if each source regards an alternative as surprising for independent reasons, it seems natural to conclude that the alternative should be *very* surprising since we have different reasons for considering it as such. When the information about independence is not available, an idempotent rule appears to be more cautious. Indeed an idempotent operation can even cope with redundant information. Concludingly, adopting idempotence is really a matter of context. It looks natural with expert opinion pooling where experts often have common background.

Auto-duality: Cheng and Kashyap (1989) have defended the auto-duality property, namely operations \square that satisfy:

$$E_1 \square E_2 = \overline{(\bar{E}_1) \square (\bar{E}_2)}. \quad (15)$$

The meaning of (15) is a neutrality property with respect to the membership scale that contains membership grades $\mu_{E_1}(s)$ and $\mu_{E_2}(s)$. As a consequence, it implicitly rejects conjunctive and disjunctive modes of combination, since intersection and union exchange under De Morgan law and (15) express an invariance property under this law. Hence auto-duality cannot be used as a universal property. It looks natural if the numbers $\mu_{E_1}(s)$ and $\mu_{E_2}(s)$ reflect *preference* intensities, in the scope of social choice (Dubois and Koning, 1991), but does not look very suitable to the problem of information aggregation.

Adaptiveness: The 2-source elementary example in the beginning of this section points out that the overlap between each piece of information affects the choice of the proper combination rule, since the absence of overlap prevents the use of conjunctive rules. Particularly, when $E_1 \cap E_2 \neq \emptyset$, the assumption that both sources are reliable becomes less and less plausible, as the overlap between E_1 and E_2 becomes small. Hence one might request that such an overlap be

evaluated and be used as a source of information of its own. The advantage is to obtain an adaptive combination rule that gradually turns from a conjunctive combination into a disjunctive one as the disagreement between the sources increases. This practical requirement may be incompatible with formal ones such as associativity, for instance. It seems more suitable for sources with common background, whose almost disagreement suggests that one of them is wrong. When sources refer to different backgrounds, the reasons why source 1 supplies E_1 and source 2 supplies E_2 might be distinct; computing $E_1 \cap E_2$ then legitimately produces a significant improvement of the precision of the information.

5.3. Basic Combination Modes in the Possibilistic Setting

Since a particular case of possibility distribution is the characteristic function of a set, the basic combination modes in the possibilistic setting are conjunctions and disjunctions. There exist several possible choices among pointwise fuzzy set-theoretic operations for defining conjunctive and disjunctive combinations of possibility distributions. Let π_i be the possibility distribution supplied by source i , $i = 1, n$. Define

$$\forall s \in S, \pi_{\wedge}(s) = *_{i=1,n} \pi_i(s) \quad (\text{fuzzy set intersection}) \quad (16)$$

and

$$\forall s \in S, \pi_{\vee}(s) = \perp_{i=1,n} \pi_i(s) \quad (\text{fuzzy set union}) \quad (17)$$

where $*$ and \perp are $[0,1]$ -valued operations defined on $[0,1] \times [0,1]$ which enjoy the duality relation $a \perp b = 1 - (1 - a) * (1 - b)$ in order to express De Morgan's law. Candidates for $*$ and \perp are so-called triangular norms and co-norms; i.e., $*$ is an operation which is associative, commutative, increasing (in the wide sense) in both places and such that $a * 1 = a$ and $0 * 0 = 0$ (Schweizer and Sklar, 1983). The main continuous solutions are $a * b = \min(a, b)$, $a * b = a \cdot b$ (product) and $a * b = \max(0, a + b - 1)$, which leads to $a \perp b = \max(a, b)$, $a \perp b = a + b - a \cdot b$, and $a \perp b = \min(1, a + b)$; see Dubois and Prade (1988b) for details.

As already said, the conjunctive aggregation makes sense if all the sources are considered as equally and fully reliable, while the disjunctive aggregation corresponds to a weaker reliability hypothesis, namely, in the group of sources there is at least one reliable source for sure, but we do not know which one. In the conjunctive case, the min operation corresponds to a purely logical view of the combination process: the source which assigns the least possibility degree to a given value is considered as the best-informed with respect to this value. Note that with min, when all sources perfectly agree ($\forall i \neq j, \pi_i = \pi_j$), there is no reinforcement effect. Generally, agreement between experts is due to common background, and the idempotence of min deals with such a kind of redundancy. On the contrary if $*$ = product, if all the sources agree that a value s is not fully possible, this value will receive a possibility degree strictly smaller than $\min_{i=1,n} \pi_i(s)$, i.e., the lack of complete possibility is reinforced; a necessary condition for

choosing such an operation is the independence of the sources. This assumption may be more adapted to sensor fusion problems. The min and the product operations can be justified in the setting of upper and lower probability as best approximations respecting the closure property (Dubois and Prade, 1992c).

An important issue with conjunctive combination as defined by (16) is the fact that the result may be subnormalized, i.e., it may happen that $\nexists s, \pi_{\wedge}(s) = 1$. In that case it expresses a conflict between the sources. Clearly the conjunctive mode makes sense if all the π_i significantly overlap, i.e., $\exists s, \forall i, \pi_i(s) = 1$, expressing that there is at least a value of s that all sources consider as completely possible. If $\forall s, \pi_{\wedge}(s)$ is significantly smaller than 1 this mode of combination is debatable since in that case at least one of the sources or experts is likely to be wrong, and a disjunctive combination might be more advisable.

When (16) provides subnormal results ($\sup \pi_{\wedge} < 1$), we may think of renormalizing π_{\wedge} , thus leading to the new rule (here written for $n = 2$ for simplicity)

$$\forall s \in S, \pi(s) = \frac{\pi_{\wedge}(s)}{h(\pi_1, \pi_2)} \quad (18)$$

where $\pi_{\wedge}(s) = \pi_1(s) * \pi_2(s)$ and $h(\pi_1, \pi_2)$ is the height of the intersection of π_1 and π_2 defined by

$$h(\pi_1, \pi_2) = \sup_{s \in S} \pi_1(s) * \pi_2(s) = \sup \pi_{\wedge} \quad (19)$$

$h(\pi_1, \pi_2)$ is a natural measure of overlap between two possibility distributions. It is called a consistency index. A similar behavior can be obtained if we use a Jacquard index $\frac{|F_1 \cap F_2|}{|F_1 \cup F_2|}$ instead of $h(\pi_1, \pi_2)$, where $\mu_{F_1} = \pi_1$, $\mu_{F_2} = \pi_2$, and $|F| = \sum_{s \in S} \mu_F(s)$.

The normalized conjunction enforces the assumption of full reliability of the sources since all values that are considered impossible by one source but possible by the other are rejected. This normalized rule is very analogous to Dempster's rule of combination of Shafer (1976)'s belief functions, which, from a random set point of view is nothing but a normalized random set intersection under a stochastic independence assumption. Moreover for $*$ = product, (18) coincides (up to the normalization factor) with Dempster's rule applied to consonant belief functions (which are mathematically the same as necessity measures), for the computation of the plausibility of singletons. It has been also pointed out (Dubois and Prade, 1988a) that on a binary frame, i.e., a 2-element set $S = \{s, \bar{s}\}$, if we interpret the MYCIN certainty factor $CF(s)$ as $\pi(s) - \pi(\bar{s})$, (18) yields MYCIN rule of combination (Buchanan and Shortliffe, 1984), for $*$ = product. Moreover for $*$ = product, the normalized rule of combination (18) is associative, otherwise it is only quasi-associative. These remarks indicate that most if not all pooling operations commonly used in the expert system literature are conjunctive. Note that the operations min and max, when applied to possibility distributions only exploit the ordinal part

of them, and assume that levels of possibility used by the various sources are commensurate. On the contrary, the use of product and its De Morgan dual assumes that possibility degrees are genuine numbers. In any case, the results of the pooling should be interpreted in an ordinal way, i.e., as a ranking of the values of the parameter x under study, in terms of their respective plausibility. It is more difficult to interpret the resulting degrees of possibility with a frequentist approach. However ordinal information is often sufficient for practical purposes.

5.4. Towards Adaptive Rules of Combination

Clearly the renormalization erases the conflict between the sources as expressed by the subnormalization. Even if it is good that the result of the combination focuses on the values on which all the sources partially agree (in the sense that none of them gave to these values a possibility equal to 0), it would be better that the result also keeps track of the conflict in some way. Also, as pointed out in (Dubois and Prade, 1988a), the normalized rule may be very sensitive to rather small variations of possibility degrees around 0; indeed the rule is not continuous in the vicinity of the total conflict expressed by $h(\pi_1, \pi_2) = 0$. A natural idea for keeping track of a partial conflict is to discount the result given by (18) by a weight corresponding to the lack of normalization, i.e., $1 - h(\pi_1, \pi_2)$. Namely $1 - h(\pi_1, \pi_2)$ is viewed as the degree of possibility that both sources are wrong since when $h(\pi_1, \pi_2) = 1$, the two sources are supposed to be right. As an example of such a discounted conjunctive combination rule we have

$$\begin{aligned} \forall s \in S, \pi'(s) &= \max(\pi(s), 1 - h(\pi_1, \pi_2)) \\ &= \max\left(\frac{\pi_1(s) * \pi_2(s)}{h(\pi_1, \pi_2)}, 1 - h(\pi_1, \pi_2)\right). \end{aligned} \quad (20)$$

The amount of conflict $1 - h(\pi_1, \pi_2)$ induces a uniform level of possibility for all values outside the ones emerging in the subnormalized intersection of π_1 and π_2 , i.e., the result of the combination (18) is not fully certain. Clearly, $\lim_{h(\pi_1, \pi_2) \rightarrow 0} \pi'(s) = 1$, so that the discontinuity effect is coped with, resulting in total ignorance.

When $h(\pi_1, \pi_2)$ is low, we could be more optimistic and assume that the discrepancy between the sources is due to one of them being wrong, not both. Then, instead of transferring the amount of conflict as a uniform degree of possibility on the whole referential S , we may limit ourselves to the union of F_1 and F_2 ($\pi_1 = \mu_{F_1}$, $\pi_2 = \mu_{F_2}$), considering that there is no reason to put a non-zero possibility degree on values that both sources agree to consider as impossible. Hence the more elaborated adaptive rule (Dubois and Prade, 1992b)

$$\forall s \in S, \pi_{AD}(s) = \max\left(\frac{\min(\pi_1(s), \pi_2(s))}{h(\pi_1, \pi_2)}, \min(\max(\pi_1(s), \pi_2(s)), 1 - h(\pi_1, \pi_2))\right). \quad (21)$$

$\pi_{AD}(s)$ evaluates at $x = s$ the fuzzy logic expression: "the two sources are reliable and claim that $x = s$, OR (they are not both reliable AND one of them is reliable and claims that $x = s$)", where $h(\pi_1, \pi_2)$ evaluates to what extent both sources are reliable, AND and OR being expressed by min and max.

The kind of result obtained by this adaptive rule is given on Figure 2. The pooling rule tries to keep the assumption that both sources are reliable as long as can be, i.e., the area where the two sources agree is always preferred. This rule has been tested in the setting of computerized image processing by Deveughèle and Dubuisson (1993).

The extension of this rule to more than two sources is not obvious because of its lack of associativity and even of quasi-associativity. This is the price paid for adaptiveness. An obvious extension to $k > 2$ sources would mean changing $\min(\pi_1, \pi_2)$ and $\max(\pi_1, \pi_2)$ into $\min(\pi_1, \dots, \pi_k)$, and $\max(\pi_1, \dots, \pi_k)$ respectively in (21). But this extension, if natural, will not be efficient because it only considers the two assumptions "all sources are right" and "one source is right", among which it builds a trade-off. Clearly the more sources, the more likely they will supply scattered information so that most of the time, the agreement index $h(\pi_1, \pi_2, \dots, \pi_k) = \sup_s \min_i \pi_i(s) = 0$, while $\max_i \pi_i$ will yield a very uninformative result, i.e., the rule will behave disjunctively. Other intermediary assumptions about the sources can be envisaged as considered in the sequel.

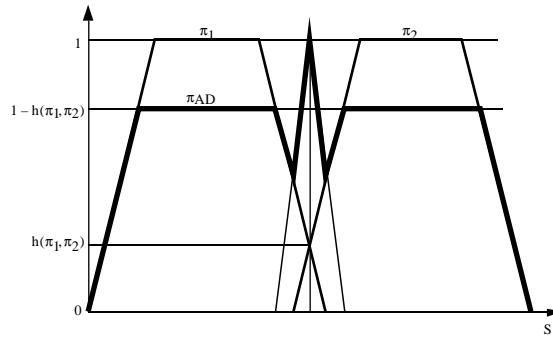


Fig. 2. Adaptive conjunctive/disjunctive pooling rule

5.5. Pooling Based on Numerical Quantifiers

An intermediary mode of pooling the π_i 's ranging between the conjunctive and the disjunctive modes consists in assuming that j experts out of $k = |K|$ are reliable. The pooling method then consists in selecting a subset $J \subseteq K$ of experts such that $|J| = j$, assume that they are reliable and combine their opinions conjunctively. Then, considering that it is not known which one of these subsets J contain reliable experts, combine the intermediary results disjunctively. The following formula is obtained (Dubois, Prade and Testemale, 1988).

$$\pi_{(j)}(s) = \max_{\substack{J \subseteq K \\ |J|=j}} \min_{i \in J} \pi_i(s). \quad (22)$$

Clearly, $\pi_{(k)} = \pi_{\wedge}$ (for $*$ = min) and $\pi_{(1)} = \pi_{\vee}$ (for \perp = max) i.e., this mode of aggregation subsumes the conjunctive and disjunctive ones. The above combination rule is equivalent to some also proposed by Yager (1985), and can be easily calculated, as follows for each value of s :

- i) rank-order the $\pi_i(s)$ such that $\pi_{i_1}(s) \geq \pi_{i_2}(s) \geq \dots \geq \pi_{i_k}(s)$
- ii) then $\pi_{(j)}(s) = \pi_{i_j}(s)$.

This scheme can be extended to fuzzy quantifiers, in order to model assumptions such as "most experts are reliable", "approximately j experts are reliable", etc. (Dubois et al., 1988; Yager, 1985).

The choice of the number j of supposedly reliable sources can be guided by the scalar cardinality $|R|$ obtained by the assessment step. But a more interesting idea is to extend the adaptive rule (21): one may derive two values m and n representing a pessimistic and an optimistic evaluation of the number of reliable experts. Let $T \subseteq K$ be a subset of experts and $h(T) = \sup_s \min_{i \in T} \pi_i(s)$ be the index of agreement among experts in T . Then define

$$m = \sup\{|T|, h(T) = 1\} ; n = \sup\{|T|, h(T) > 0\}.$$

Then it is plausible to assume that at least m sources among K are reliable (since there is a fully agreeing subset of m experts) and at most n are reliable (since there is a total conflict within all groups of more than n experts). It can be shown that generally $m < n$. Then a natural extension of the adaptive rule (21) to the case of $k > 2$ sources is

$$\pi_{AD}(s) = \max\left(\frac{\pi_{(n)}(s)}{h(n)}, \min(\pi_{(m)}(s), 1 - h(n))\right) \quad (23)$$

where $h(n) = \max\{h(T), |T| = n\}$. It can be checked that when $|K| = 2$, (23) coincides with (21). Indeed if $h(\pi_1, \pi_2) = 1$, then $n = m = 2$ and $\pi_{(2)} = \min(\pi_1, \pi_2)$. If $h(\pi_1, \pi_2) \in (0, 1)$, $m = 1$, $n = 2$, $\pi_{(1)} = \max(\pi_1, \pi_2)$, $\pi_{(2)} = \min(\pi_1, \pi_2)$.

For instance, consider the three possibility distributions on Figure 3. It can easily be checked that $m = 2$, $n = 3$, $\pi_{(3)} = \min_{i=1,3} \pi_i$, $\pi_{(2)} = \max(\min(\pi_1, \pi_2), \min(\pi_2, \pi_3), \min(\pi_3, \pi_1))$. The resulting distribution may have a complex shape; however it highlights which values are most plausible for the parameter. On Figure 3, one immediately sees that the most plausible values are the ones where all experts partially agree (between s_2 and s_3) and to a lesser extent, the ones where two experts strongly agree (between s_4 and s_5).

5.6. Prioritized Aggregation of Expert Opinions

As pointed out earlier, the fuzzy set R of more or less reliable experts obtained from the assessment procedure is useful to partition the set K of experts into classes K_1, K_2, \dots, K_q of equally reliable ones, where K_j corresponds to a higher reliability level than K_{j+1} , for $j = 1, q$. Then the above symmetric aggregation schemes can be applied to each class K_j . The combination between results obtained from the K_j 's can be performed upon the following principle: the response of group K_2 is used to refine the response of group K_1 insofar as it is consistent with it. If π_1 is obtained from K_1 and π_2 from K_2 , the degree of consistency of π_1 and π_2 is again $h(\pi_1, \pi_2) = \sup_s \min(\pi_1(s), \pi_2(s))$ and the following combination rule has been proposed (Dubois and Prade, 1988c; Yager, 1991):

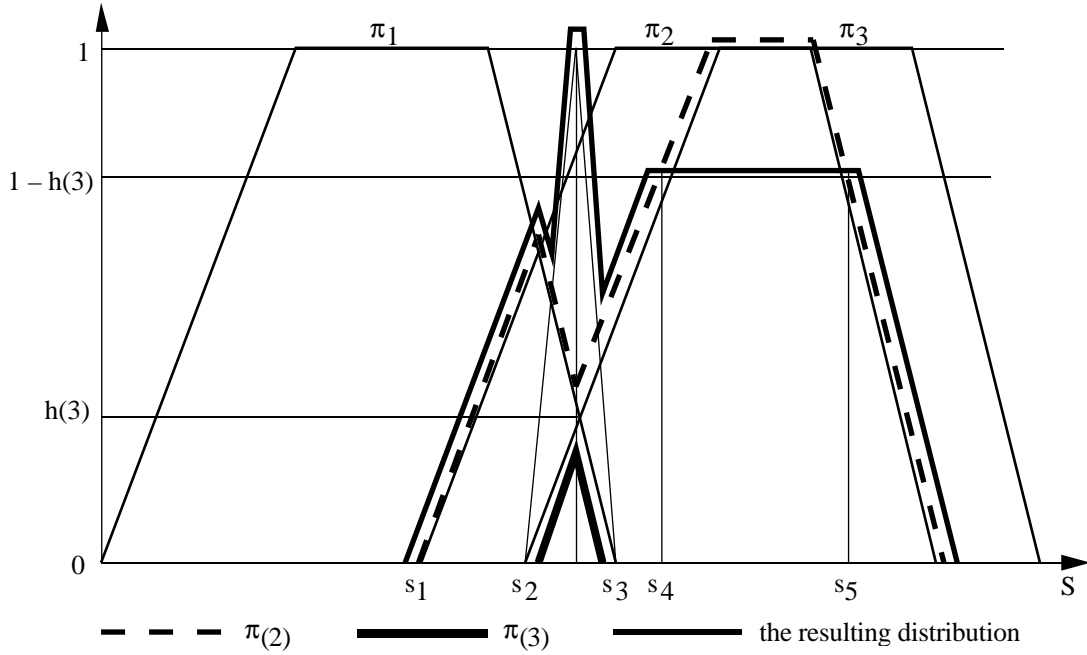


Fig. 3. The adaptive rule with n sources

$$\pi_{1-2} = \min(\pi_1, \max(\pi_2, 1 - h(\pi_1, \pi_2))). \quad (24)$$

Note that when $h(\pi_1, \pi_2) = 0$, K_2 contradicts K_1 and only the opinion of K_1 is retained ($\pi_{1-2} = \pi_1$), while if $h(\pi_1, \pi_2) = 1$ then $\pi_{1-2} = \min(\pi_1, \pi_2)$.

(24) is easily interpreted as the conjunctive combination of the information supplied by source 1 and the information supplied by source 2, the latter being discounted by a certainty coefficient $h(\pi_1, \pi_2)$, such that the degree of possibility that source 2 is wrong is $1 - h(\pi_1, \pi_2)$. $\pi_{1-2}(s)$ evaluates, in the sense of fuzzy logic, the expression "x = s is possible for source 1 and either x = s is possible for source 2 or source 2 is wrong (with certainty $h(\pi_1, \pi_2)$ that source 2 is not wrong)". As suggested by Figure 2, π_{1-2} is subnormalized when it differs from $\min(\pi_1, \pi_2)$.

Hence a renormalization, as in (18) can be used. Moreover (24) is a conjunctive prioritized rule. Its disjunctive counterpart has been proposed by Dubois and Prade (1988d) and Yager (1991)

$$\pi_{1-2} = \max(\pi_1, \min(\pi_2, h(\pi_1, \pi_2))). \quad (25)$$

The effect of this rule is to truncate the information supplied by the less prioritary source, while disjunctively combined with source 1. Again if the two sources disagree ($h(\pi_1, \pi_2) = 0$) then $\pi_{1-2} = \pi_1$; if $h(\pi_1, \pi_2) = 1$ then $\pi_{1-2} = \max(\pi_1, \pi_2)$. Rather than choosing between (24) and (25) it seems more promising to integrate these expressions in the adaptive rule (21) in place of symmetrical conjunctive and disjunctive combinations.

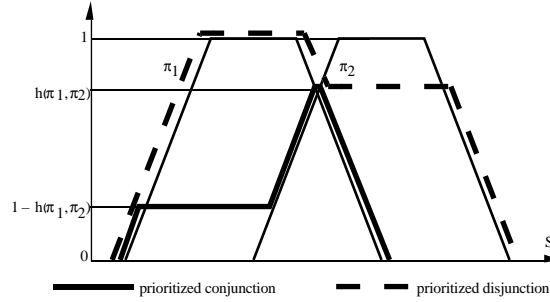


Fig. 4. Prioritized conjunctive and disjunctive rules

6. CONCLUSION

The framework proposed in this paper for data fusion makes sense when it seems difficult to represent the information supplied by the sources by means of single probability distributions, due to imprecision and/or lack of statistical evidence. Regarding the conjunctive and disjunctive pooling modes that possibility theory provides, they are particularly suitable when sources are heterogeneous, i.e., cannot be viewed as instances of a single random source. The proposed combination rules are optimistic in the sense that they always try to assume that as many sources as possible are reliable, as usually found conjunctive rules (like Dempster rule, MYCIN rule) do. However when the assumption that all sources tell the truth is not tenable, it is possible to express that a certain proportion of the sources are faithful, without being forced to point them out. The disjunctive rule, for instance, systematically assumes that at least one source is right. This optimistic prejudice in the combination modes tries to extract the most useful part of the data, by remaining as informative as possible, while avoiding the systematic use of a conjunctive rule. An adaptive rule has been proposed, and extended to multiple sources, in order to achieve a balance between the conjunctive and the disjunctive combination modes, the trade off being driven by the amount of consistency between the sources.

In this paper, the possibility distribution synthesized from the information given by an expert is supposed to restrict *for sure* all the possible values of the variable to estimate (up to the residual uncertainty when the support of the possibility distribution is the whole variable domain S). In this perspective the conjunctive aggregation is natural when there is no conflict between the expert opinions. However we may also imagine situations where experts do not intend to provide a whole range of possible values but only point out some plausible (usually small) subsets of values for the variable; then two experts may provide disjoint subsets without being

necessarily in conflict (they just did not give the same subset of the set of plausible values) and then disjunctive combination is very natural.

The probabilistic setting (at least the Bayesian one) is not very well adapted to the expression of conjunctive and disjunctive combination modes. This is no longer true when upper and lower probabilities are allowed (e.g., Chateauneuf, 1991). Contrastedly, the possibilistic setting does not lend itself very easily to averaging combination modes. Possibilistic data fusion is much more related to logical approaches which cope with inconsistency in multiple source databases using notions of maximal consistent subbases (see Dubois et al., 1992; Cholvy, 1993, among others). For instance rule (23) involves maximal subsets of coherent sources. Our approach looks promising in problems such as knowledge base aggregation, expert judgment pooling, and data fusion with heterogeneous sensors.

As an example, a procedure for processing human-originated information has been devised, based on the above described principles (Kalfsbeek, 1990), implemented on a computer (Sandri, 1991) and tested on actual reliability data (Sandri, 1990; Sandri et al., 1993). Its basic steps go as follows:

- i) elicitation of seed variables by experts and assessment using the $Q(E,x)$ indices;
- ii) comparison of the performance of pooling methods on seed variables;
- iii) elicitation of unknown parameter values by experts;
- iv) pooling of these pieces of information using the best pooling method.

Some of the combination rules (especially conjunction and disjunction) have been tested. Due to the lack of space it is not possible to present the results obtained by this method on the analysis of actual reliability data. Such results appear in a report (Sandri et al., 1993) to be published elsewhere. It turns out that the proposed method seems to compare favorably with respect to purely probabilistic approaches on the studied examples: all methods give similar results when experts agree, and our method does somewhat better when experts considered as reliable disagree, because the combination rules proposed here are more flexible and avoid the averaging effect.

ACKNOWLEDGEMENTS

This paper owes much to discussions with Henk Kalfsbeek who introduced us to the problem of expert opinion pooling and with whom the possibilistic approach to this problem has been developed, and implemented in the framework of Sandra Sandri's PhD thesis. Discussions with Sandra Sandri are also gratefully acknowledged.

REFERENCES

- Buchanan, B.G. and E.H. Shortliffe (1984). *Rule-Based Expert Systems – The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley, Reading.
- Chateaufneuf, A. (1991). Combination of compatible belief functions and relation of specificity. In: *Proc. of the 3rd IPMU Conf.*, ENSTA, Paris, 105-107. Revised version to appear in: *Advances in the Dempster-Shafer Theory of Evidence* (M. Fedrizzi, J. Kacprzyk and R.R. Yager, Eds.), 1994. Wiley, New York
- Cheng, Y.Z. and R.L. Kashyap (1989). A study of associative evidential reasoning. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **11**(6), 623-631.
- Cholvy, L. (1993). Proving theorems in a multi-source environment. In: *Proc. of the 13th Inter. Joint Conf. on Artificial Intelligence (IJCAI'93)*, Chambéry, France, Aug. 28-Sept. 3, 66-71.
- Cooke, R.M. (1988). Uncertainty in risk assessment: A probabilist's manifesto. *Reliability Engineering and Systems Safety*, **23**, 277-283.
- Cooke, R.M. (1991). *Experts in Uncertainty*. Oxford University Press, Oxford, UK.
- Coolen, F.P.A. (1994). *Statistical Modeling of Expert Opinions Using Imprecise Probabilities*. PhD Thesis, Eindhoven University of Technology, Eindhoven, The Netherlands.
- Deveughèle, S. and B. Dubois (1993). Using possibility theory in perception: An application in artificial vision. In: *Proc. of the 2nd IEEE Inter. Conf. on Fuzzy Systems (FUZZ-IEEE'93)*, San Francisco, CA, March 28-April 1st, 821-826.
- Dubois, D. and H.W. Kalfsbeek (1990). Elicitation, assessment and pooling of expert judgments using possibility theory. In: *Proc. of the 8th Inter. Congress of Cybernetics and Systems* (C.N. Manikopoulos, Ed.), New York, June 11-15, 360-367. Published by New Jersey Institute of Technology Press.
- Dubois, D. and J.L. Koning (1991). Social choice axioms for fuzzy set aggregation. *Fuzzy Sets and Systems*, **49**(3), 257-274.
- Dubois, D., J. Lang and H. Prade (1992). Dealing with multi-source information in possibilistic logic. In: *Proc. of the 10th Europ. Conf. on Artificial Intelligence (ECAI'92)* (B. Neumann, Ed.), Vienna, Austria, Aug. 3-7, 1992, 38-42. Wiley, New York.
- Dubois, D. and H. Prade (1986). Weighted minimum and maximum operations in fuzzy set theory. *Information Sciences*, **39**, 205-210.
- Dubois, D. and H. Prade (1987). Une approche ensembliste de la combinaison d'informations incertaines. *Revue d'Intelligence Artificielle*, **1**(4), 23-42.
- Dubois, D. and H. Prade (1988a). Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, **4**, 244-264.
- Dubois, D. and H. Prade (with the collaboration of H. Farreny, R. Martin-Clouaire and C. Testemale) (1988b). *Possibility Theory – An Approach to the Computerized Processing of Uncertainty*. Plenum Press, New York.
- Dubois, D. and H. Prade (1988c). Default reasoning and possibility theory. *Artificial Intelligence*, **35**, 243-257.
- Dubois, D. and H. Prade (1988d). On the combination of uncertain or imprecise pieces of information in rule-based systems – A discussion in the framework of possibility theory. *Int. J. of Approximate Reasoning*, **2**, 65-87.
- Dubois, D. and H. Prade (1990) Aggregation of possibility measures. In: *Multiperson Decision Making using Fuzzy Sets and Possibility Theory* (J. Kacprzyk, M. Fedrizzi, eds.), 55-63. Kluwer, Dordrecht.
- Dubois, D. and H. Prade (1992a). On the relevance of non-standard theories of uncertainty in modeling and pooling expert opinions. *Reliability Engineering and Systems Safety*, **36**, 95-107.

- Dubois, D. and H. Prade (1992b). Combination of fuzzy information in the framework of possibility theory. In: *Data Fusion in Robotics and Machine Intelligence* (M.A. Abidi, R.C. Gonzalez, eds.), 481-505. Academic Press, New York.
- Dubois, D. and H. Prade (1992c). When upper probabilities are possibility measures. *Fuzzy Sets and Systems*, **49**, 65-74.
- Dubois, D., H. Prade and C. Testemale (1988). Weighted fuzzy pattern matching. *Fuzzy Sets and Systems*, **28**, 313-331.
- Hajek, P. (1985). Combining functions for certainty degrees in consulting systems. *Int. J. of Man-Machine Studies*, **22**, 59-76.
- Kalfsbeek, H. (1990). Elicitation, assessment and pooling of expert judgment using possibility theory. Working Paper PER1829/89 (revised version), Joint Research Center of the EEC, Ispra, Italy.
- Klir, G.J. and T. Folger (1988). *Fuzzy Sets, Uncertainty and Information*. Prentice Hall, Englewood Cliffs, NJ.
- Mosleh, A. and G. Apostolakis (1984). Models for the use of expert opinions. In: *Low Probability/ High Consequence Risk Analysis* (R.A. Waller and V.T. Covello, Eds.). Plenum Press, New York.
- Sandri, S. (1990). Fuzzy sets and possibility theory for RPES development. Final Report, Contract n° 3309-87-12, JRC-Ispra and IRIT, Toulouse.
- Sandri, S. (1991). La combinaison de l'information incertaine et ses aspects algorithmiques. Thèse de Doctorat de l'Université Paul Sabatier, Toulouse, France.
- Sandri, S., A. Besi, D. Dubois, G. Mancini, H. Prade and C. Testemale (1989). Data fusion problems in an intelligent data base interface. In: *Reliability Data Collection and Use in Risk and Availability Assessment* (U. Colombari, Ed.), 655-670. Springer Verlag, Berlin.
- Sandri, S., D. Dubois and H. Kalfsbeek (1993). Elicitation, assessment and pooling of expert judgement using possibility theory. In: Tech. Report IRIT/93-24-R, IRIT, Univ. P. Sabatier, Toulouse, France.
- Schweizer, B. and A. Sklar (1983). *Probabilistic Metric Spaces*. North-Holland, New York.
- Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton Univ. Press, Princeton, NJ.
- Wagner, C. and K. Lehrer (1981). *Rational Consensus in Science and Society*. D. Reidel, Dordrecht.
- Yager, R.R. (1985). Aggregating evidence using quantified statements. *Information Sciences*, **36**, 179-206.
- Yager, R.R. (1987). Quasi-associative operations in the combination of evidence. *Kybernetes*, **16**, 37-41.
- Yager, R.R. (1991). Non-monotonic set-theoretic operators. *Fuzzy Sets and Systems*, **42**, 173-190.
- Yager, R.R. (1992). On the specificity of a possibility distribution. *Fuzzy Sets and Systems*, **50**, 279-292.
- Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, **8**, 338-353.
- Zadeh, L.A. (1968). Probability measures of fuzzy events. *J. Math. Anal. Appl.*, **23**, 421-427.
- Zadeh, L.A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, **1**, 3-28.