La fusion d'informations imprécises*

Imprecise Data Fusion

par D. DUBOIS, H. PRADE
I.R.I.T., Université Paul Sabatier,
31062 Toulouse Cedex, France.

Résumé

La théorie des possibilités offre un cadre formel naturel pour la représentation de données imprécises, d'informations pauvres. Cette théorie prend tout son intérêt quand il s'agit d'agréger des informations issues de plusieurs sources (par exemple un groupe d'experts, un ensemble hétérogène de capteurs, plusieurs bases de données). En effet elle s'avère être beaucoup plus souple que la théorie des probabilités pour décrire des modes d'agrégation qui ne correspondent pas à des moyennes. Dans cet article on tente d'expliquer pourquoi la théorie des possibilités est intéressante dans le problème de fusion d'informations imprécises, et on décrit les modes d'agrégation qu'elle permet de représenter, avec les hypothèses qui les sous-tendent. On indique notamment l'existence d'opérations de combinaison adaptatives qui prennent en compte le niveau de conflit entre les sources. Cette approche semble justifiée pour l'agrégation d'opinions d'experts. On suggère ici qu'elle peut, dans certaines conditions, être utilisée pour la fusion multi-capteurs.

Mots clés : Théorie des possibilités, fusion de données, combinaison adaptative

Abstract

Possibility theory offers a natural setting for representing imprecise data and poor information. This theory turns out to be quite useful for the purpose of pooling pieces of information stemming from several sources (for instance, several experts, sensors, or databases). Indeed it looks more flexible than probability theory for the representation of aggregation modes that do not express averaging processes. This paper tentatively explains why possibility theory is appealing for the fusion of impresise data, and it describes several aggregation modes it allows, along with their underlying assumptions. The existence of adaptive combination rules are pointed out, that take into account the level of conflict between the sources. This approach sounds natural in the pooling of expert opinions. It is suggested here that, under some assumptions, it might also be useful in sensor data fusion.

***Key words :** Possibility theory, data fusion, adaptive combination*

1. Introduction

Le problème de fusion de données imprécises issues de plusieurs sources d'information se rencontre dans divers champs d'application et notamment

i) pour agréger des mesures issues de plusieurs capteurs, en robotique par exemple (Luo and Kay [22]).

ii) dans les systèmes d'information quand on interroge plusieurs bases de données et qu'il faut fournir une seule réponse à l'utilisateur (Cholvy [3]).

iii) pour synthétiser les opinions d'un groupe d'experts qui fournissent chacun une estimation de la valeur d'un paramètre mal connu (taux de panne d'un équipement dans un environnement particulier, par exemple) (Cooke [5]).

Notre thèse est que d'une part, il ne semble pas toujours raisonnable de représenter une information imprécise par une distribution de probabilité. D'autre part, il n'existe pas une opération de combinaison universelle qui traiterait le problème de fusion de manière satisfaisante dans tous les cas de figure. En effet la théorie des probabilités, notamment bayésienne, est trop riche pour pouvoir exprimer fidèlement des informations pauvres reflétant une ignorance partielle. Par ailleurs, la palette d'opérations d'agrégation qu'elle offre (essentiellement des moyennes) semble très limitée. En revanche, la théorie des possibilités s'avère beaucoup plus souple pour modéliser des informations imprécises, et elle offre des modes d'agrégation de type ensembliste, qui sont absents des modèles probabilistes usuels, mais qu'on connaît bien en logique (le "et" et le "ou"). Une mesure de possibilité peut également représenter de façon commode une famille de mesures de probabilité entre lesquelles on ne sait pas choisir. Ces idées sont présentées plus en détail dans cet article, qui n'a d'autre ambition que de susciter l'intérêt des chercheurs du traitement du signal, pour une façon d'aborder les problèmes de fusion de données qui diffère notablement de leur pratique usuelle, souvent

* Il s'agit pour l'essentiel de la traduction remaniée d'un article intitulé "Possibility theory and data fusion in poorly informed environments" invité par la revue IFAC "Control Engineering Practice", vol. 2(5), 811-823, 1994.

influencée par une approche probabiliste pure et dure et l'emploi de techniques de filtrage.

Cet article est construit comme suit : d'abord on tente de préciser les contours du problème de fusion que l'on va aborder. Puis on explique pourquoi la théorie des probabilités, utilisée de façon traditionnelle, semble limitée pour traiter ce problème. En section 3, on décrit le modèle possibiliste pour la représentation d'informations imprécises et ses liens avec la théorie des probabilités. En fait ce n'est pas tant la théorie des probabilités en elle-même qui est remise en cause, que certaines hypothèses qui sous-tendent des façons de l'utiliser, et notamment celle de la distribution unique. En section 4, on aborde les modes d'agrégation d'informations imprécises dont on dispose en théorie des possibilités. En section 5, on montre que ces modes d'agrégation peuvent être intégrés dans le cadre d'une règle adaptative dont le comportement dépend du niveau de conflit entre les sources.

2. Le problème de combinaison d'informations incertaines

Le problème que nous allons considérer peut se formuler abstraitement de la façon suivante : considérons un paramètre, que nous désignerons par x, dont la valeur est inconnue. Un ensemble de sources d'information fournit des estimations de la valeur de x. Ces estimations ne sont pas forcément des valeurs précises mais peuvent être très pauvres. A la limite, on admet que certaines sources déclarent ignorer la valeur de x. Il s'agit alors de synthétiser ces informations afin de déterminer les valeurs les plus plausibles du paramètre x, compte tenu de la confiance éventuelle que l'on peut avoir sur la fiabilité des sources. Une source peut représenter un capteur, un expert ou une base de données. L'information qu'une source fournit peut donc être de nature diverse : dans le cas d'un capteur on dispose souvent d'une valeur nominale et d'un intervalle d'erreur. Dans le cas d'un expert, il peut s'agir d'une information linguistique, ou d'un ensemble d'intervalles de confiance avec les niveaux correspondants. Dans le cas d'une base de données, l'information peut prendre la forme d'un histogramme (base de données statistique), ou d'une formule logique (base de données déductive). On voit que notre notion de signal est relativement générale.

Ce problème de fusion n'est qu'un parmi d'autres. Il s'agit ici d'agréger des informations et non des préférences. Dans le cas de l'agrégation des préférences, il fait alors sens de chercher à définir une "source moyenne" qui reflète l'opinion du groupe en faisant des compromis entre les sources individuelles. Au contraire l'agrégation d'information est affaire de fiabilité et de vérité, non de préférence. Dans ce cas il est important de détecter quelles sont les sources fiables, et quelles sont les sources qui sont dans l'erreur, et il n'est pas question de faire "la moyenne" entre les unes et les autres. Les opérations de type moyenne peuvent néanmoins se justifier si on peut considérer l'ensemble des sources comme une seule source aléatoire dont les informations à agréger sont les réalisations. Dans ce cas l'ensemble d'information est interprété comme une statistique standard, et son caractère dissonant est interprété comme une variabilité de nature aléatoire. Il n'est pas évident qu'on puisse toujours se ramener à cette situation, notamment dans le cas de sources très hétérogènes ou encore de sources dépendantes.

Lorsqu'on dispose de plusieurs sources d'information, une idée naturelle est de chercher à recouper ces informations afin de mettre en valeur les zones de consensus. Plus ces zones de consensus sont étendues, plus les sources sont concordantes, et plus le recoupement s'avère justifié. En revanche le caractère conflictuel des informations issues de plusieurs sources suggère qu'au moins l'une de ces sources donne une information fausse. C'est sur ces principes que reposent les méthodes de fusion de données présentées ici.

D'autres aspects peuvent être pris en compte : la dépendance et la fiabilité des sources. Deux sources jugées dépendantes, et de surcroît concordantes, ne doivent pas compter pour deux. Par ailleurs, si une source est jugée plus fiable qu'une autre, les informations fournies par la première doivent être prioritaires sur celles fournies par la seconde lorsque cette dernière est en désaccord avec les informations de la première source. Parfois on ne connaît pas la fiabilité relative des sources. Il convient alors de faire des hypothèses sur la proportion de sources qui ne se trompent pas. Cela donne des méthodes de fusion très proches de pratiques courantes dans les bases de connaissances inconsistantes, où on recherche des sous-ensembles maximaux "consistants" (c'est-à-dire cohérents).

Enfin notons que le problème abordé ici est celui de l'agrégation d'informations reçues en parallèle. Il diffère notablement du cas où un ensemble de connaissances est mis à jour par une information nouvelle, ce que l'on appelle la révision. Dans la révision les deux sources ne jouent pas le même rôle. Le principe de la révision est de recalculer un état de croyance qui accepte la nouvelle information, tout en restant aussi proche que possible de l'état de croyance originel : la révision obéit à un principe de changement minimal, et donne en général la priorité au nouvel élément d'information. En revanche la combinaison d'informations incertaines est un processus symétrique; la seule dissymétrie qui puisse y apparaître est due à la présence de sources de fiabilités diverses. On peut bien sûr combiner révision et fusion en considérant la mise à jour d'une connaissance à partir d'un ensemble de sources. Dans cet article on considère essentiellement le problème de combinaison parallèle de sources hétérogènes, c'est-à-dire lorsqu'on ne peut supposer que les informations à agréger proviennent d'une source aléatoire unique; on ne supposera pas non plus que les sources sont indépendantes.

Pour mettre en œuvre une procédure de fusion, plusieurs problèmes doivent être résolus :

– comment modéliser l'imperfection des informations produites par les sources?

– comment connaître et modéliser la fiabilité des sources?

– comment combiner les informations disponibles?

Pour modéliser l'imperfection des informations on a souvent recours à la théorie des probabilités, ou aux intervalles d'erreur. L'emploi d'une distribution de probabilité est parfois dû au fait qu'on ne connaît pas d'autre théorie de l'incertain. Par ailleurs, l'intervalle d'erreur est fort peu expressif : trop étroit il est peu fiable, trop large il n'est plus informatif. En revanche la distribution de probabilité contient une information très riche, trop riche parfois pour traduire fidèlement la pauvreté relative des informations issues d'experts ou de capteurs peu fiables. Une manière de compromis entre les deux est d'adopter une représentation à base d'intervalles de probabilité, ce qui correspond à une famille de distributions de probabilité, plutôt qu'à une probabilité unique. On a alors une représentation plus générale que les intervalles d'erreurs (un intervalle d'erreur peut être vu comme l'ensemble des distributions de probabilité dont cet intervalle est le support), mais fort complexe à manipuler. La théorie des possibilités offre un modèle plus simple, compatible avec l'interprétation probabiliste imprécise, et plus expressif que les intervalles d'erreur car distinguant des valeurs plus plausibles que d'autres.

La détermination de la fiabilité des sources est très dépendante du contexte d'application. Dans le cas de capteurs, la fiabilité peut être connue d'avance (donnée par le constructeur de l'appareil de mesure), ou affaire d'usage. Dans le cas d'experts, des protocoles d'évaluations à base de questions-tests ont été établis; voir (Cooke [5]), (Sandri [27]). Dans cet article, ce problème ne sera donc pas abordé. En revanche on s'attardera sur les divers modes de combinaison et les hypothèses qu'ils sous-tendent sur la fiabilité des sources.

3. Limitations des approches probabilistes classiques

Dans cette partie, on envisage essentiellement le cas d'informations fournies par des experts. Les modélisations classiques d'opinions d'experts tentent de calibrer une densité de probabilité sur la base de quelques informations telles que des fractiles et des valeurs plausibles du paramètre étudié. Par exemple l'expert fournit des valeurs x_i et x_s d'un paramètre x, tel que $P(x \leq x_i) = 0,05$ et $P(x \leq x_s) = 0,95$, ainsi qu'une évaluation du mode, de la médiane ou de la moyenne de la distribution. Sur cette base on cherche à définir une densité de forme donnée (par exemple une bêta-distribution) qui soit calibrée au mieux par rapport à ces données. Notons que la nature de telles distributions est matière à controverse de par leur caractère subjectif. Les probabilités obtenues sont parfois vues comme des "degrés de certitude" (subjectifs) que le paramètre x prenne sa valeur dans telle ou telle zone; elles peuvent aussi être considérées comme des estimations subjectives de fréquence : $P(x \leq x_i) = 0,05$ est la proportion des situations (connues de l'expert) où il a observé $x \leq x_i$. Cette dernière interprétation est plus aisée à utiliser en pratique. Enfin certaines approches bayésiennes telles que celle de Mosleh et Apostolakis[23] supposent que l'expert fournit une valeur précise du paramètre. Quand plusieurs experts fournissent de telles informations, les distributions de probabilité sont agrégées en une seule qui reflète l'opinion du groupe, en donnant plus d'importance à l'opinion des experts fiables. Il y a deux méthodes principales pour agréger les opinions d'experts : la méthode du consensus justifiée théoriquement par Wagner et Lehrer[31] et utilisée par Cooke[5], et l'approche bayésienne, dont un exemple est la méthode de Mosleh et Apostolakis [23].

Dans la méthode du consensus chaque expert E_i fournit une distribution de probabilité p_i et on lui attribue un poids w_i qui reflète sa fiabilité. La probabilité résultante est la moyenne pondérée $\sum w_i p_i$ (avec $\sum w_i = 1$). Cooke [4] a développé une théorie de la pondération, à base de règles d'évaluation qui obligent les experts à être calibrés et informatifs. Dans cette approche, l'ensemble des experts est vu comme une source aléatoire et le poids w_i est interprété comme la fréquence avec laquelle cette source fournit p_i comme réponse.

La méthode bayésienne suppose que l'analyste qui recueille les opinions des experts possède lui-même une opinion sur la valeur du paramètre, cette opinion étant représentée par une distribution de probabilité. Les valeurs précises fournies par les experts sont alors utilisées pour mettre à jour l'opinion de l'analyste. La fiabilité des experts est capturée à l'aide de fonctions de vraisemblance, à savoir les probabilités conditionnelles $P_i(x_i|x_{\text{vrai}})$ pour que l'expert E_i fournisse la valeur $x = x_i$ sachant que la vraie valeur de x est x_{vrai}. Connaissant $P(x)$ et $P_i(x_i|x)$ on peut appliquer le théorème de Bayes pour calculer la distribution a posteriori. Si les experts ne sont pas indépendants, on utilise directement une distribution jointe $P(x_1...x_n|x)$ au travers de coefficients de corrélation. D'autres méthodes bayésiennes existent (voir Cooke [5]) mais elles sont plus complexes.

Notons qu'une pratique courante dans la fusion d'informations issues de capteurs est de modéliser les mesures par des gaussiennes (calibrées sur l'intervalle d'erreur) que l'on va multiplier point par point et renormaliser. Cette technique, qui suppose les sources indépendantes, a son origine dans une règle de combinaison due à Bernoulli (voir Shafer[30]) et est un cas particulier de la règle de Dempster (Shafer, [29]).

Dans le cas de l'agrégation d'opinions d'experts, ces méthodes peuvent être critiquées pour plusieurs raisons (Dubois and Prade [15]) :

– L'identification d'une distribution de probabilité demande plus d'information que ce qu'un expert peut fournir. Il y a plus d'une distribution de probabilité de mode donné et de fractiles à 5 % et 95 % donnés. Le choix d'une famille paramétrée de distributions de probabilité est affaire de commodité de calcul. En conséquence le modèle de l'opinion d'expert ainsi obtenu n'est pas fidèle aux connaissances de l'expert, lequel est peut être bien incapable de choisir une distribution de probabilité dans la famille. De même l'utilisation de gaussiennes dans la représentation des erreurs dues aux capteurs n'est pas toujours justifiée, sinon par un souci d'efficacité du calcul.

– Un expert préférera toujours fournir un intervalle plutôt qu'une valeur précise, parce que son savoir est non seulement incertain, mais il est imprécis. De même, un capteur fournit une valeur imprécise, représentée par un intervalle d'erreur. La théorie des probabilités n'a pas été conçue pour représenter l'imprécision, autrement dit, les connaissances incomplètes. Dans le cas des bases de données, cela correspond à une information manquante, qui empêche de répondre à certaines questions.

– La méthode du consensus par somme pondérée de probabilités présente un inconvénient notoire lorsqu'il s'agit d'agréger des opinions conflictuelles sur la valeur de x (l'une des sources affirme que x est grand et l'autre que x est petit). La méthode du consensus fournira une distribution de probabilité dont l'espérance mathématique ne sera ni petite ni grande, c'est-à-dire une valeur que les experts s'accordent à rejeter. Il semble plus raisonnable de choisir entre les deux valeurs proposées pour x en fonction de la fiabilité de chaque expert, ou si celle-ci est inconnue, de fournir une réponse prudente qui soit fidèle aux informations obtenues (par exemple x est soit petit, soit grand, mais pas moyen). La moyenne pondérée des valeurs obtenues pour x semble naturelle si x mesure un degré de préférence, mais semble peu adaptée à la recherche d'une valeur vraie dans un ensemble de valeurs proposées comme possibles.

– Une autre interprétation de l'opération de moyenne provient de la considération du groupe de sources en tant que réalisations d'une source aléatoire unique. Cette hypothèse implique une homogénéité des sources qui est valide si les observations proviennent d'une source unique consultée à des instants différents. En revanche, elle est très contestable dans le cas de plusieurs capteurs distincts, de natures diverses, dont certains fonctionnent mal, ou dans le cas d'experts dont certains peuvent se tromper. De plus la moyenne arithmétique influe sur la variance du résultat, à savoir : la variance d'une moyenne pondérée de distribution de probabilité est plus petite que celle de chacune des distributions. Ce phénomène est naturel dans le cas d'observations indépendantes, mais cette hypothèse d'indépendance est douteuse dans le cas d'opinions d'experts, et parfois aussi pour les capteurs fournissant des signaux ou des images .

– Le principal inconvénient des méthodes bayésiennes est qu'elles requièrent de la connaissance a priori sur la valeur étudiée. La distribution de probabilité a priori doit être fournie soit par l'analyste qui reçoit les opinions d'experts, soit par un expert dont l'information joue un rôle particulier. De façon générale les approches bayésiennes sont gourmandes en données et ne peuvent pas être initialisées à partir d'un état d'ignorance.

L'approche proposée dans cet article tente de surmonter les difficultés rencontrées par les approches probabilistes pures sur le problème de fusion de données issues de plusieurs sources hétérogènes. Ses atouts principaux sont : la capacité de représenter l'information imprécise, le caractère non-impératif de la connaissance a priori, et la présence de plusieurs modes de combinaison. Le choix d'un mode de combinaison dépend de la fiabilité des sources et du niveau de conflit entre les observations qu'elles fournissent.

4. Théorie des possibilités et informations imprécises

Dans la suite on supposera que l'information dont on dispose sur la valeur d'un paramètre x est représentée par une *distribution de possibilité* π_x (Zadeh [36]). Pour tout élément s du domaine S de x, on appellera $\pi_x(s)$ le "degré de possibilité" pour que $x = s$, avec les conventions suivantes : $\pi_x(s) = 0$ indique l'impossibilité que x vaille s. En revanche $\pi_x(s) = 1$ indique que rien n'empêche x de valoir s. Les valeurs telles que $\pi_x(s) > 0$ sont supposées mutuellement exclusives et forment l'ensemble (flou [35]) des valeurs (plus ou moins) possibles de x. La vraie valeur de x existe mais elle est supposée mal connue.

Les valeurs s telles que $\pi_x(s) = 1$ sont les plus plausibles. De plus, $\pi_x(s) > \pi_x(s')$ indique que $x = s$ est plus plausible que $x = s'$. On suppose qu'au moins un des éléments de S est la valeur de x, ce qui se traduit par $\pi_x(s) = 1$ pour au moins une valeur s^*. Notons qu'une distribution de possibilité n'obéit pas aux mêmes conventions qu'une distribution de probabilité. Dans le cas probabiliste, savoir $\mathrm{Pr}_x(s^*) = 1$ implique $x = s^*$, ce qui n'est pas le cas lorsque $\pi_x(s^*) = 1$. En particulier on peut avoir $\pi_x(s) = 1 \, \forall s \in S$ et cela traduit l'ignorance totale sur la valeur de x. Si deux distributions de possibilité π'_x et π_x sont telles que $\forall s, \pi_x(s) \leq \pi'_x(s)$ on dira que π_x est au moins aussi spécifique ("informante") que π'_x. La distribution de possibilité la moins informante est celle telle que $\pi_x(s) = 1, \forall s$, qui affirme que toutes les valeurs dans S sont également possibles. Inversement toute distribution de possibilité telle que $\pi_x(s^*) = 1$ pour $x = s^*$ et $\pi_x(s) = 0$ sinon, est maximalement informante.

Connaissant une distribution de possibilité π_x, deux fonctions d'ensemble permettent d'évaluer la confiance que l'on peut avoir sur l'affirmation $x \in B$ pour tout sous-ensemble B de S : le degré de possibilité $\prod(B)$ et le degré de nécessité $N(B)$ sont définis par (Zadeh [36], Dubois et Prade [11]) :

$$\begin{aligned} \Pi(B) &= \sup_{s \in B} \pi_x(s) && (1) \\ N(B) &= \inf_{s \notin B} 1 - \pi_x(s) && (2) \\ &= 1 - \Pi(\overline{B}) \end{aligned}$$

où $\overline{B}$ est le complément de B dans S. Cette terminologie se justifie comme suit : supposons que π_x soit la fonction caractéristique du sous-ensemble $E \subseteq S$, soit $\pi_x(s) = \mu_E(s)$. Cela veut dire que $x \in E$ est la seule information dont on dispose. Alors $\prod(B) = 1 <=> E \cap B \neq 0$: sachant $x \in E, x \in B$ est possible. De même $N(B) = 1 <=> E \subseteq B$: sachant $x \in E$, $x \in B$

est certain. La relation de dualité entre possibilité et nécessité $N(B) = 1 - \prod(\overline{B})$ veut dire que $x \in B$ est certain si et seulement si $x \in B$ est impossible.

On peut concevoir une distribution de possibilité comme une famille d'ensembles emboîtés $\{A_i, i = 1, n\}$ avec des niveaux de confiance associés λ_i, avec $A_i \subseteq A_{i+1}, \forall i = 1, n-1$ et $\lambda_i \leq \lambda_{i+1}$. Cela permet de voir une distribution de possibilité comme le codage simple d'une famille de distributions de probabilité. Soit $\mathbb{P} = \{P, P(A_i) \geq \lambda_i, i = 1, n\}$ une telle famille. Tout événement B peut être évalué sur la base de bornes supérieures et inférieures de probabilité

$$P^*(B) = \sup\{P(B), P \in \mathbb{P}\},$$
$$P_*(B) = 1 - P^*(\overline{B}) = \inf\{P(B), P \in \mathbb{P}\}$$

De par l'emboîtement de la famille $\{A_i, i = 1, n\}$, on peut montrer que si l'on considère la distribution de possibilité définie par

$$\begin{aligned} \pi_x(s) &= \min_{i=1,n} \max(1 - \lambda_i, \mu_{A_i}(s)) \\ &= 1 \text{ si } s \in A_1 \\ &= \min \ \{1 - \lambda_i, s \in A_i\} \text{ sinon} \end{aligned} \tag{3}$$

alors on a bien $P^*(B) = \prod(B)$, $P_*(B) = N(B)$, $\forall B \subseteq S$ (Dubois et Prade, [17]). Donc les degrés de possibilité peuvent être vus comme des bornes supérieures de probabilité. Par ailleurs, π_x donné par (3) est également la distribution la moins informative telle que les contraintes $N(A_i) \geq \lambda_i$, $i = 1, n$ soient satisfaites, c'est-à-dire celle qui accorde les degrés de possibilité maximaux aux valeurs $s \in S$. Inversement toute distribution de possibilité π_x telle que l'image $\pi_x(S)$ de S par π_x est un sous-ensemble fini de [0,1], est représentable par une famille pondérée $\{(A_i, \lambda_i), i = 1, n\}$ d'ensembles emboîtés. Soit $\{\alpha_1 = 1 > \alpha_2 ... > \alpha_n\} = \pi_x(S)$. Alors on a $A_i = \{s, \pi_x(s) \geq \alpha_i\}$ et $\lambda_i = 1 - \alpha_{i+1}$, pour $i = 1, m$, en posant $\alpha_{m+1} = 0$.

On peut également interpréter π_x dans le cadre de la théorie des croyances de Shafer [29]. En posant $p_i = \alpha_i - \alpha_{i+1}$, $i = 1, m$, on vérifie qu'on a également

$$\pi_x(s) = \sum_{i:s \in A_i} p_i$$

Une distribution de possibilité peut donc être vue comme un ensemble aléatoire $\{(A_i, p_i), i = 1, n\}$ avec $\sum p_i = 1$, qui code l'incertitude (les valeurs de probabilité p_i) et l'imprécision (la taille des sous-ensembles A_i). Notons que p_i est la probabilité pour que la source fournisse exactement A_i comme représentation de sa connaissance sur x (ce n'est pas la probabilité pour que $x \in A_i$). Dans ce contexte si on note $\prod_i(B)$ la possibilité binaire pour que $x \in B$ lorsque la source affirme $x \in A_i(\prod_i(B) = 1$ si $A_i \cap B \neq \emptyset$ et 0 sinon) alors $\prod(B)$ calculé avec (1) est une espérance mathématique de possibilité puisque

$$\Pi(B) = \sum_{i=1,n} p_i \Pi_i(B)$$

et de même (2) est une espérance mathématique de nécessité :

$$N(B) = \sum_{i=1,n} p_i N_i(B) \tag{4}$$

avec $N_i(B) = 1$ si $B \supseteq A_i$ et 0 sinon. Même lorsque les ensembles A_i ne sont pas emboités, (4) définit une fonction de croyance au sens de Shafer, et c'est encore un cas particulier d'enveloppe inférieure de probabilité.

Une troisième interprétation de la distribution de possibilité π_x consiste à la voir comme une fonction de vraisemblance probabiliste, c'est-à-dire à identifier $\pi_x(s)$ à $P(F|s)$, c'est-à-dire la probabilité pour que la source indique $x \in F$, la vraie valeur de x étant s. Alors supposons qu'on connaisse $P(F|s)$, $\forall s \in S$, et seulement cela. Si la source affirme $x \in F$, on interprète la probabilité $P(F|s)$ comme la vraisemblance de $x = s$, au sens où plus $P(F|s)$ est grand plus $x = s$ est plausible, mais ce raisonnement abductif ne permet pas de calculer la probabilité $P(s|F)$ par le théorème de Bayes, puisqu'on ne suppose pas connue une distribution a priori sur la valeur de x (c'est-à-dire avant de recevoir l'information par la source). En particulier (cf. [9]) $\forall A \subseteq S$

$$\min_{s \in A} P(F|s) \leq P(F|A) \leq \max_{s \in A} P(F|s) \tag{5}$$

Si on identifie $\pi_x(s)$ à $P(F|s)$, on voit que $\prod(A)$ est bien la borne supérieure de la fonction de vraisemblance $P(F|A)$. La borne inférieure dans (5) correspond au degré de "possibilité garantie" $\Delta(A) = \inf_{s \in A} \pi(s)$ (cf Dubois et Prade [19]) qui évalue à quel point toute valeur dans A est plausible.

Ces façons d'interpréter la distribution de possibilité montrent que l'approche possibiliste ne s'inscrit pas en opposition avec la théorie des probabilités mais vient plutôt la prolonger, dans le cas où l'hypothèse de la distribution unique doit être rejetée. Néanmoins on peut concevoir la théorie des possibilités indépendamment de toute interprétation probabiliste si on la voit comme une approche ordinale de l'incertain. L'intervalle [0,1] est alors vu comme une simple échelle ordinale qui peut être exploitée à l'aide des degrés $\prod(A)$, $N(A)$ et $\Delta(A)$. La représentation possibiliste semble plus réaliste que l'emploi d'une distribution de probabilité unique dans un certain nombre de situations, pour modéliser l'information provenant d'une source. Supposons que le domaine de S de x soit un intervalle réel $[s_o, s_m]$. Un expert fournira plus volontiers des intervalles censés contenir la valeur de x plutôt que des valeurs précises. On peut par exemple lui demander de fournir un intervalle, le plus petit possible, où il est sûr que cette valeur se trouve. Pour être informatif cet intervalle devra être strictement inclus dans S. L'expert peut aussi fournir des intervalles encore plus petits avec des niveaux de confiance plus faibles. En termes probabilistes, on peut interpréter ces informations comme des intervalles de confiance, c'est-à-dire des intervalles A_i les plus petits possibles tels que l'expert pense que $x \in A_i$ avec une probabilité au moins égale à une valeur fixée λ_i. Il est clair qu'un ensemble pondéré $\{(A_i, \lambda_i), i = 1, n\}$ de cette sorte ne suffit pas à isoler une distribution de probabilité unique. En revanche cette information est fidèlement représentée à l'aide d'une distribution de possibilité unique. En pratique (cf. Sandri, [25]) on demande 3 intervalles A_1, A_2, A_3, à l'expert, correspondant à des bornes inférieures de probabilité $\lambda_1 = 0,05$, $\lambda_2 = 0,5$ et $\lambda_3 = 0,95$.

A_1 correspond à un intervalle de valeurs plausibles pour x, et A_3 à un intervalle contenant x avec certitude pour l'expert. Mais on admet une incertitude résiduelle pour ne pas exclure l'éventualité qu'il se trompe. L'idée de représenter des opinions d'experts par des familles de distributions de probabilité a été étudiée particulièrement par Coolen [6]. Ce type de représentation peut être complexe, et c'est tout l'intérêt de la démarche possibiliste que de fournir des modèles simples de familles non-paramétrées.

Dans le cas d'un capteur, la distribution de possibilité permet de généraliser la notion d'intervalle d'erreur, car on peut exprimer de l'incertitude sur cet intervalle. L'utilisation de bornes de probabilité pour modéliser les erreurs dues à des capteurs permet de généraliser à la fois les modèles probabilistes de l'erreur, et l'approche par intervalles. Mais la détermination du type de distribution de possibilité adaptée à l'étude des erreurs dues aux capteurs reste à faire. En pratique, les modèles courants d'erreurs de mesure sont souvent probabilistes (des gaussiennes, par exemple). Il est souvent difficile de savoir si ce type de modèle est utilisé parce qu'il capture bien la réalité des phénomènes (on peut justifier, sous certaines hypothèses, la distribution gaussienne), ou simplement parce qu'il simplifie les calculs (c'est le cas pour l'hypothèse de distributions gaussiennes dans le filtre de Kalman). Pour définir une distribution de possibilité attachée à une mesure, on peut penser utiliser une valeur modale et une évaluation de la dispersion de cette mesure. Une autre idée serait de partir de la représentation habituelle probabiliste, et de transformer la probabilité P en distribution de possibilité. Ceci peut se faire en calculant la distribution de possibilité π la plus précise, dont la famille de distribution de probabilités associée $\mathbb{P}(\pi) = \{P', P'(A) \leq \prod(A), \forall A\}$ contient P (voir Dubois et al. [20]). Enfin on peut attacher des niveaux de confiance respectivement à la valeur nominale mesurée et à l'intervalle d'erreur autour de cette valeur, obtenu à partir d'informations sur la précision du capteur. On se retrouve alors dans l'hypothèse d'intervalles emboîtés.

Si on considère le cas des bases de données logiques vues comme sources d'informations, il est facile de voir que la représentation possibiliste est particulièrement bien adaptée. En effet l'information fournie par une base de données incomplète est de type disjonctive, c'est-à-dire $x \in E$ où E est un ensemble de valeurs mutuellement exclusives. Si la base de données contient des informations de fiabilités diverses, elle peut être codée en logique possibiliste (Dubois, Lang et Prade [8]) et les renseignements qu'elle fournira sont représentables à l'aide d'une distribution de possibilité qualitative sur les interprétations du langage. La théorie des possibilités offre donc un cadre unificateur pour la représentation de données incomplètes. Ce cadre est plus pauvre que le cadre probabiliste, bien qu'ayant un pouvoir expressif distinct, mais il peut être suffisant pour beaucoup d'applications.

5. La combinaison d'informations en théorie des possibilités

La théorie des possibilités utilise la notion d'intervalle, et plus généralement la notion d'ensemble, pour représenter des informations imprécises. Ces ensembles ne sont plus vus comme des collections d'objets, mais comme contenant plutôt des valeurs mutuellement exclusives. Ce recours aux concepts ensemblistes, et notamment aux ensembles flous [35], conduit tout naturellement à adopter une vision logique de la combinaison, en termes d'union et d'intersection ensemblistes (Dubois et Prade [12]). L'union et l'intersection sont les deux façons fondamentales de combiner symétriquement des ensembles. Soit deux sources 1 et 2 qui renseignent sur la valeur de x sous la forme suivante :

$$\begin{array}{ll} \text{Source 1 :} & x \in E_1 \subseteq S \\ \text{Source 2 :} & x \in E_2 \subseteq S \end{array}$$

Que peut-on dire sur x ? La thèse de cet article est que la réponse à cette question est affaire de contexte. Il ne peut exister de théorie susceptible de fournir une méthode universelle pour combiner ces deux éléments d'information, méthode qui s'appliquerait dans tous les cas. La méthode adoptée dépend d'une part des propriétés formelles qu'on souhaite lui voir satisfaire; elle dépend d'autre part du niveau de conflit entre les sources et de leurs fiabilités respectives. Il faut en fait proposer une batterie de modes de combinaison avec les hypothèses sous-jacentes. Dans l'ensemble élémentaire ci-dessus, on a intérêt à supposer les sources fiables. Tant que cela est crédible c'est-à-dire si $E_1 \cap E_2$ est suffisamment grand (en particulier si $E_1 = E_2$), on peut alors conclure $x \in E_1 \cap E_2$. En revanche si $E_1 \cap E_2 = \emptyset$, il est clair qu'au moins une des deux sources se trompe, et on peut alors supposer que l'autre est fiable. On en conclura alors que $x \in E_1 \cup E_2$. Cette combinaison ne suppose pas que l'on sache quelle est la source qui se trompe. Il est clair que la conclusion $x \in E_1 \cap E_2$ est informative mais parfois risquée, alors que la conclusion $x \in E_1 \cup E_2$ est plus sûre mais moins informative. Toute la difficulté de la combinaison d'informations incertaines consiste à chercher un compromis entre un résultat précis mais sûrement faux, et un résultat certain mais trop imprécis. Cette vision de la fusion de données a été appliquée à des systèmes d'interrogation de bases de données (Sandri et col.[26]), à la combinaison d'opinions d'experts (Sandri[25]), (Sandri et col.[27]) et au traitement du signal (Deveughèle [7], Kewley [21],Park et Lee [24]).

5.1. MODES DE COMBINAISON FONDAMENTAUX EN THÉORIE DES POSSIBILITÉS

Puisqu'une distribution de possibilité généralise une fonction caractéristique d'ensemble, les opérations de combinaison en théorie des possibilités seront soit conjonctives (intersections généralisées) soit disjonctives (réunions généralisées). Soit π_i la distribution de possibilité fournie par la source i, pour $i = 1, 2$. On définit $\pi_\wedge$ et $\pi_\vee$ comme suit :

$$\forall\, s \in S,\ \pi_\wedge(s) \quad = \quad \pi_1(s) * \pi_2(s) \qquad (6)$$

$$\forall\, s \in S,\ \pi_\vee(s) \quad = \quad \pi_1(s) \perp \pi_2(s) \qquad (7)$$

où $*$ et $\perp$ sont des opérations internes de [0,1] qui généralisent l'intersection et l'union à des ensembles flous [35]. En particulier $*$ coïncide avec un "et" logique sur $\{0, 1\}$. $*$ est commutatif, associatif, possède un élément neutre 1 (car $A \cap S = A$) et un élément absorbant 0, $*$ est aussi non-décroissante selon ses arguments; $*$ est ce qu'on appelle une norme triangulaire (Schweizer et Sklar [28]). L'opération $\perp$ vérifie les mêmes propriétés sauf que les rôles joués par 0 et 1 sont échangés, de sorte que $\perp$ coïncide avec un "ou" logique sur $\{0, 1\}$, et que cette opération est de la forme $a \perp b = 1 - (1 - a) * (1 - b)$ (loi de de Morgan). Les conjonctions $*$ principales sont le minimum, $a * b = \min(a, b)$ (la plus grande), le produit $a * b = a.b$, et la conjonction linéaire $a*b = \max(0, a+b-1)$; les disjonctions principales sont le maximum $a \perp b = \max(a, b)$ (la plus petite), la somme probabiliste $a \perp b = a + b - ab$, et la somme bornée $\min(a + b, 1)$.

La commutativité de ces opérations est justifiée pour des sources considérées comme équivalentes. Mais l'associativité n'est pas indispensable : elle est n'utilisée que pour généraliser ces combinaisons à n sources de façon modulaire. La quasi-associativité suffit (Yager [33]), c'est-à-dire une combinaison de la forme $f_n(a_1 * a_2 * \ldots a_n)$ où l'opération $*$ est associative et f_n est inversible. Par exemple, la moyenne arithmétique est de ce type. Une autre hypothèse implicite dans (6,7) est celle de fermeture, qui stipule que la combinaison de deux mesures de possibilité redonne une mesure de possibilité. Cela ne va pas forcément de soi. Par exemple si on combine deux mesures de possibilité, vues comme des fonctions de plausibilité au sens de Shafer [29], par la règle de Dempster, on n'obtient plus une mesure de possibilité car la propriété de consonance est perdue. Plus précisément si π_1 est représentable par $\{(A_i, p_i), i = 1, m\}$ et π_2 par $\{(B_j, q_j), j = 1, n\}$, avec $\sum p_j = \sum q_j = 1$, alors le résultat de la règle de Dempster est

$$\{(A_i \cap B_j, k.q_i.q_j);\ i = 1,\ m;\ j = 1,\ n;\ A_i \cap B_j \neq \emptyset\}$$

où k est un coefficient de normalisation tel que $\sum k.p_i.q_j = 1$. Lorsque les A_i et les B_j forment des familles emboîtées, il n'en est plus de même pour les $A_i \cap B_j \neq \emptyset$. Néanmoins si la fonction de plausibilité est appliquée aux singletons, $\pi_\wedge(s) = Pl(\{s\}) = \pi_1(s).\pi_2(s)$, lorsque $k = 1$ (c'est-à-dire $A_i \cap B_j \neq \emptyset \forall i, \forall j$), ce qui justifie la combinaison conjonctive avec le produit.

Comme il a été dit plus haut, l'agrégation conjonctive s'applique au cas où les deux sources sont considérées comme totalement fiables. L'agrégation disjonctive correspond à une hypothèse plus faible à savoir qu'une des deux sources est fiable alors que l'autre fournit une information erronée, sans que l'on puisse l'identifier.

Parmi les combinaisons conjonctives, le minimum correspond à une vision "logique" de la fusion multisources : la source qui octroie le plus faible degré de possibilité à une valeur du paramètre x est considérée comme la mieux informée à cet égard. Avec l'opération minimum, il n'y a pas d'effet de renforcement entre les sources lorsqu'elles fournissent la même information ($\pi_1 = \pi_2$). Cette combinaison généralise celle qui, lorsqu'on combine deux bases de connaissance cohérentes, consiste à mettre en commun les informations qu'elles contiennent. Dans le cas d'agrégation d'opinions d'experts, ceux-ci ont souvent des connaissances communes, et le caractère idempotent du minimum prend en compte cette redondance. En revanche la combinaison avec le produit conduit à un effet de renforcement entre les sources qui rejettent des valeurs du paramètre x jugées peu possibles : cette opération n'est justifiée que si les sources sont indépendantes. Cette hypothèse semble mieux adaptée pour la fusion multicapteurs que pour l'agrégation d'opinions d'experts. L'opération minimum et l'opérateur produit peuvent se justifier dans la cadre des enveloppes de probabilité comme meilleures approximations respectant la propriété de fermeture (Dubois et Prade [17]).

Une remarque importante relative à la combinaison conjonctive est que le résultat qu'elle fournit peut être sous-normalisé, c'est à dire $\forall s \in S,\ \pi_\wedge(s) < 1$. Cela traduit un conflit entre les sources, puisqu'aucune valeur n'est trouvée simultanément complètement possible par ces deux sources. Moins le résultat est normalisé, plus le conflit entre les sources est important, et moins il est légitime de supposer que les sources sont toutes les deux fiables. Deux hypothèses permettent de restaurer une distribution normalisée. On peut continuer à supposer que les deux sources sont fiables et renormaliser le résultat de la combinaison conjonctive. Au contraire on peut considérer que, dès qu'il y a conflit, aussi faible soit-il, l'une des deux sources se trompe. Dans le premier cas on est amené à renormaliser $\pi_\wedge$:

$$\pi(s) = \pi_\wedge(s)/h(\pi_1, \pi_2) \qquad (8)$$

avec $\pi_\wedge(s) = \pi_1(s)*\pi_2(s)$ et $h(\pi_1, \pi_2) = \sup_s \pi_1(s)*\pi_2(s)$ qui représente le degré de cohérence entre les sources. $h(\pi_1, \pi_2)$ est une mesure naturelle de chevauchement entre les distributions de possibilité π_1 et π_2. Cette règle, analogue à la règle de Dempster pour la combinaison de fonctions de croyance (Shafer [29]), prend le parti d'éliminer les informations affirmées par une source et totalement rejetées par l'autre; une partie de l'information est donc oubliée. Notons que lorsque $*$ = produit, (8) demeure associatif et coïncide, lorsque S ne contient que 2 éléments, avec le règle de combinaison de faits incertains issues de plusieurs règles dans les systèmes experts tels que MYCIN (Buchanan et Shortiffe [1]) : voir Dubois et Prade [12], [13]). La règle (8) avec $*$ = minimum n'est pas associative, mais il est facile de l'étendre à n sources en combinant d'abord et en renormalisant. La règle (8) n'est plus définie dès que $h(\pi_1, \pi_2) = 0$. Dans ce cas, seule la règle disjonctive (7) peut s'appliquer. On peut vouloir l'utiliser également lorsque $h(\pi_1, \pi_2)$ est faible. En effet il est facile de voir que la règle de combinaison (8) est discontinue au voisinage du conflit, ce qui implique que les résultats numériques obtenus seront quelque peu arbitraires quand $h(\pi_1, \pi_2)$ est proche de 0 (Dubois, Prade [12], [13]).

5.2. UNE RÈGLE DE COMBINAISON ADAPTATIVE

L'existence de deux modes de combinaison possibles (conjonctif ou disjonctif) liés à deux hypothèses antagonistes (les deux sources sont fiables; l'une des deux sources est fiable) pose un problème de choix qui dépend du niveau de consensus $h(\pi_1, \pi_2)$ entre les sources : si $h(\pi_1, \pi_2)$ est élevé, on utilise la règle conjonctive que l'on peut normaliser. Si $h(\pi_1, \pi_2)$ est faible, la règle disjonctive est recommandée. Cette situation est peu satisfaisante, car il est difficile de fournir un seuil pour discriminer la zone où la règle conjonctive s'applique de celle où la règle disjonctive s'applique. Il paraîtrait plus satisfaisant de passer continûment du mode conjonctif au mode disjonctif. Pour atteindre ce but on peut utiliser la façon dont la théorie des possibilités modélise l'information incertaine, en considérant $h(\pi_1, \pi_2)$ comme le degré de nécessité (certitude) pour que les deux sources soient fiables. Quand $h(\pi_1, \pi_2) = 1$ on suppose que les sources sont fiables. Sinon, on considèrera qu'il y a un degré de possibilité égal à $1 - h(\pi_1, \pi_2)$ pour que la valeur du paramètre x soit ailleurs que dans le support de $\pi_\wedge$. Par ailleurs on supposera également que l'une des sources au moins est fiable, c'est-à-dire que le support de $\pi_\vee$ contient nécessairement la valeur de x. Au total, on arrive à la règle de combinaison adaptative suivante [16] :

$$\pi_{AD}(s) = \max(\pi_\wedge(s)/h(\pi_1, \pi_2),\ \min(1 - h(\pi_1, \pi_2),\ \pi_\vee(s))) \tag{9}$$

qui traduit en logique floue l'hypothèse selon laquelle soit les deux sources sont fiables (et on normalise), soit elles ne sont pas fiables, et l'une des deux est fiable. Le résultat fourni par une telle règle de combinaison est illustré en Figure 1. On privilégie systématiquement la zone où les deux sources sont d'accord, aussi petite soit-elle. Si $h(\pi_1, \pi_2) = 1$ on retrouve la règle conjonctive. Si $h(\pi_1, \pi_2) = 0$ on retrouve la règle disjonctive. Mais cette zone est d'autant plus noyée dans l'incertitude que le degré de consensus entre les sources est faible. Cette règle a été testée par Deveughèle [7] sur un problème (élémentaire) de fusion d'informations visuelles issues de plusieurs caméras. L'extension de cette règle au cas de plus de deux sources n'est pas triviale car on a perdu l'associativité pour prix de l'adaptativité. Une idée évidente est de remplacer $\pi_\wedge$ et $\pi_\vee$ dans (9), par la combinaison conjonctive, et respectivement disjonctive, des distributions de possibilité $\pi_i, i = 1, k > 2$. Mais cette méthode n'est pas très convaincante. En général, plus il y a de sources plus elles seront

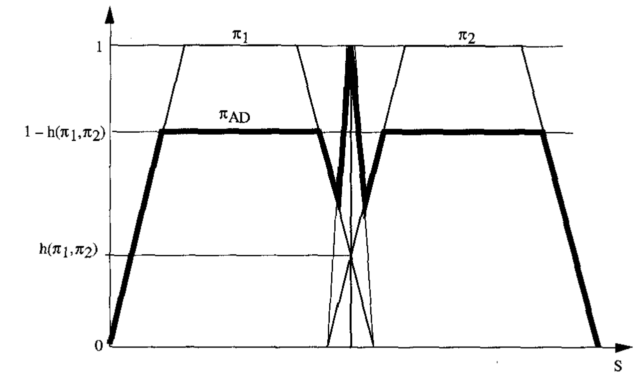


Figure 1. – Combinaison adaptative de deux sources.

conflictuelles ($\sup \pi_\wedge = 0$), et on obtiendra la combinaison disjonctive $\pi_\vee$. Cette combinaison repose alors sur une hypothèse excessivement faible (une des sources parmi les k ne se trompe pas); en conséquence, plus il y a de sources moins $\pi_\vee$ sera informative. Pour extraire des sources une information pertinente, il convient de faire des hypothèses plus hardies que celle qui consiste à supposer qu'une seule source parmi k est fiable.

5.3. AGRÉGATIONS ADAPTATIVES QUANTIFIÉES

Une hypothèse intermédiaire consiste à supposer que, parmi les k sources, j sont fiables, sans nécessairement savoir désigner celles qui le sont. Soit K l'ensemble des sources, et J un sous-ensemble de K. Si J est l'ensemble des sources fiables on peut combiner conjonctivement les distributions π_j, $j \in J$. Comme par hypothèse, c'est l'un des sous-ensembles J qui contient les sources fiables, on doit combiner disjonctivement les résultats intermédiaires. On obtient donc la règle de combinaison quantifiée suivante (Dubois et col. [18])

$$\pi_{(j)}(s) = \max_{\substack{J \subseteq K \\ |J| = j}} \ \mathop{*}_{i \in J} \pi_i(s) \tag{10}$$

Il est facile de voir que $\pi_{(1)} = \pi_\vee$ alors que $\pi_{(k)} = \pi_\wedge$. Autrement dit ce mode d'agrégation généralise les modes conjonctifs et disjonctifs. Quand $* = \min$, ce genre d'opération a été également proposé par Yager [32] sous une autre forme. Le calcul pratique du résultat apparaît combinatoire mais peut être simplifié comme suit :

Calcul de $\pi_{(j)}(s)$:

i) ranger $\{\pi_i(s),\ i = 1, k\}$ dans un ordre décroissant $\pi_{i_1}(s) \geq \pi_{i_2}(s) ... \geq \pi_{i_k}(s)$

2) $\pi_{(j)}(s) = \pi_{i_j}(s)$ (si $* = \min$)
$\quad = \pi_{i_1}(s)\pi_{i_2}(s)...\pi_{i_j}(s)$ (si $*$ = produit)

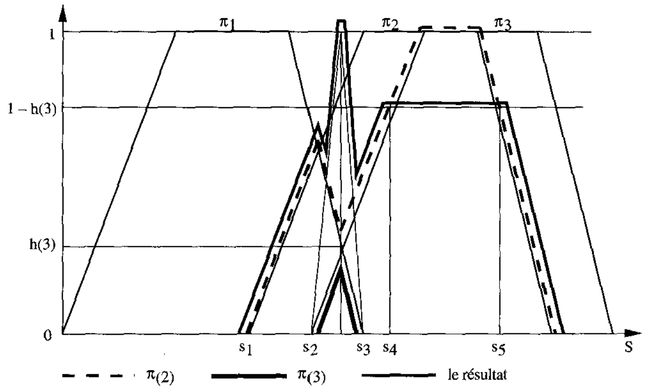


Figure 2. – Combinaison adaptative de trois sources.

Il est facile de vérifier que $\pi_{(k)} \leq \pi_{(k-1)} \leq ...\pi_{(2)} \leq \pi_{(1)}$. Cette règle peut être modulée en introduisant des quantificateurs flous (Yager [32], Dubois et col. [18]), par exemple "environ j sources sont fiables". La règle (10) n'est pas facile à utiliser si on ne dispose pas d'informations sur le nombre de sources fiables. Pour émettre une hypothèse plausible à ce propos on va à nouveau utiliser des degrés de consensus entre les sources, et calculer une estimation optimiste du nombre de sources fiables en partant du principe que des sources fiables sont nécessairement concordantes, et une estimation pessimiste qui prend acte du fait que des sources en total conflit ne peuvent être toutes fiables. Soit J un sous-ensemble de sources, et $h(J) = \sup_s \min_{i \in J} \pi_i(s)$ leur degré de consensus. Soient m et n, tels que $m \leq n$, ces estimations. On pose

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

Alors il est plausible de supposer qu'il y a un groupe de m sources fiables, et qu'en aucun cas il n'y en a plus de n. Il est alors naturel de généraliser à k sources la règle adaptative comme suit.

$$\pi_{AD}(s) = \max(\pi_{(n)}(s)/h(n),\ \min(1 - h(n),\ \pi_{(m)}(s)) \quad (11)$$

avec $h(n) = \max\{h(J), |J| = n\}$. On peut vérifier que si $|K| = 2$, on retrouve bien la règle adaptative (9). En effet si $h(\pi_1, \pi_2) = 1$, alors $m = n = 2$ et $\pi_{AD} = \pi_{(2)} = \min(\pi_1, \pi_2)$. Si $0 < h(\pi_1, \pi_2) < 1$, alors $m = 1$, $n = 2$, $\pi_{(1)} = \max(\pi_1, \pi_2), \pi_{(2)} = \min(\pi_1, \pi_2)$.

A titre d'exemple, considérons les trois distributions de possibilité de la Figure 2. On peut voir facilement que $m = 2$, $n = 3$, $\pi_{(3)} = \min(\pi_1, \pi_2, \pi_3)$, $\pi_{(2)} = \max(\min(\pi_1, \pi_2), \min(\pi_2, \pi_3), \min(\pi_3, \pi_1))$. La forme de la distribution résultante n'est pas simple, mais elle indique clairement les valeurs les plus plausibles du paramètre décrit par les trois sources : là où les trois sources sont tant soit peu d'accord, et dans une moindre mesure, là où deux des sources sont en fort consensus.

5.4. SOURCES PLUS OU MOINS FIABLES

Dans un certain nombre de situations, certaines sources sont jugées plus fiables que d'autres. Par exemple si on cherche à connaître l'âge d'une personne on peut utiliser des informations indirectes ("il est chauve donc il a plus de cinquante ans") issues de considérations générales non liées à la personne, et d'observations d'autres paramètres que celui que l'on veut identifier, et des informations directes (dans l'exemple, avoir accès à une carte d'identité). Il est clair que les informations directes sont plus crédibles que les conclusions issues d'un raisonnement plausible. Dans ce cas il conviendra de rejeter ces dernières en cas de conflit avec des informations plus spécifiques de la situation étudiée. Si on dispose d'un ensemble K de sources d'information, on peut classer cet ensemble selon la fiabilité des sources. Soit $\{K_1...K_n\}$ une partition de K telles que les sources réunies dans K_i sont jugées de fiabilité égale, et plus fiables que celles de K_j dès que $j > i$. La combinaison d'informations issues des sources $K_1...K_n$ peut s'effectuer comme suit : combiner symétriquement (par la règle adaptative) les informations relatives à K_1 puis ne raffiner le résultat par les informations issues de K_2 que lorsque celles-ci ne contredisent pas les sources de K_1, etc. Soit π_i la distribution de possibilité construite à partir de K_i et considérons à nouveau le niveau de consensus $h(\pi_1, \pi_2)$ entre π_1 et π_2 dans le cas où $n = 2$. Alors on peut utiliser la règle suivante (Dubois-Prade [14]; Yager [34]) :

$$\pi_\wedge^{1>2} = \min(\pi_1, \max(\pi_2, 1 - h(\pi_1, \pi_2))) \quad (12)$$

Elle exprime que les sources de K_1 fournissent une information fiable tandis que l'information issue de K_2 n'est considerée comme fiable qu'avec un degré de certitude égal au degré de consensus entre K_1 et K_2. Quand $h(\pi_1, \pi_2) = 0$ on constate que $\pi_\wedge^{1>2}(s) = \pi_1$, c'est-à-dire que l'on ne tient pas compte des informations provenant de K_2. Au contraire, si $h(\pi_1, \pi_2) = 1$ alors $\pi_\wedge^{1>2} = \min(\pi_1, \pi_2)$. (12) est appelée conjonction non-monotone par Yager [34]. C'est aussi une opération de conjonction pondérée (Dubois Prade [10]) où les pondérations valent 1 sur π_1 et $h(\pi_1, \pi_2)$ sur π_2. Comme indiqué sur la Figure 3, le résultat de cette opération est sous-normalisé, dès que (12) ne coincide pas avec l'opération minimum. On pourrait donc vouloir renormaliser la distribution obtenue. Par ailleurs, (12) possède une contre-partie disjonctive (Yager [34])

$$\pi_\vee^{1>2}(s) = \max(\pi_1,\ \min(\pi_2, h(\pi_1, \pi_2))) \quad (13)$$

Cette opération tronque la distribution π_2 pour indiquer que si l'on suppose qu'une des deux sources se trompe, on tend à penser que c'est la source 2. Tout comme (12), la disjonction à priorité élimine π_2 dès que $h(\pi_1, \pi_2) = 0$. Plutôt que de choisir entre (12) et (13), on peut les intégrer dans une règle adaptative comme on l'a fait plus haut pour $\pi_\wedge$ et $\pi_\vee$.

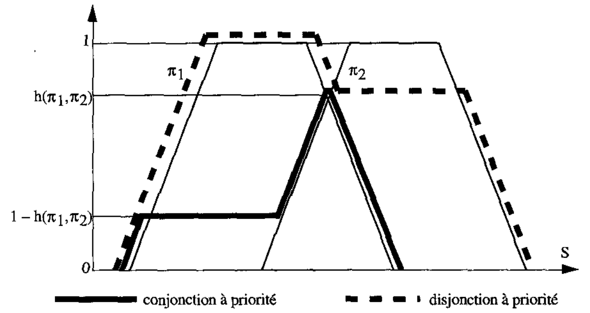


Figure 3. – Combinaison avec priorité.

6. Conclusion

Le cadre proposé dans cet article pour la fusion de données imprécises prend tout son sens quand il paraît difficile de représenter l'information fournie par chaque source à l'aide d'une seule distribution de probabilité, et lorsque l'ensemble des informations à agréger ne peut être vu comme le résultat d'expériences aléatoires indépendantes. On a en revanche considéré le cas de sources peu informées, individualisées, et sujettes à l'erreur. Cette situation est typique de la fusion de bases de données et de l'agrégation de savoirs issus d'experts. D'ailleurs, les modes fondamentaux d'agrégation présentés ici généralisent des approches courantes dans les bases de données déductives. Une base de connaissances $\mathbb{B}$ contenant des formules logiques $p_1, ..., p_n$ est systématiquement assimilée en logique à la conjonction $p_1 \wedge p_2 ... \wedge p_n$ de ces formules, ce qui revient à calculer une distribution de possibilité de type $\pi_\wedge$. Si $\mathbb{B}$ est "inconsistante" au sens logique (c'est à dire incohérente), une méthode classique, et très prudente, consiste à effectuer la disjonction des formules obtenues comme conjonction des formules de chaque sous-base maximale "consistante" (cohérente). Cela revient à considérer le cas de n distributions de possibilités $\pi_1 ... \pi_n$ tout ou rien, assimilables à des ensembles non flous $A_1 ... A_n$ de valeurs possibles, tels que $\cap_i A_i = \emptyset$. Alors l'approche logique précitée consiste à repérer les ensembles maximaux au sens de l'inclusion dans l'ensemble $\{J, \cap_{i \in J} A_i \neq \emptyset\}$ et à calculer $\cup_{J \text{ maximal}} \cap_{i \in J} A_i$. Notre règle adaptative revient à ne sélectionner parmi ces éléments maximaux, que les ensembles J de cardinalité maximale. En effet, avec les notations du 5.3 on voit que $n = m = \max\{|J|, \cap_{i \in J} A_i \neq \emptyset\}$, $h(n) = 1$ et (11) se réduit à (10) avec $j = n$. On obtient $\cup_{|J|=n} (\cap_{i \in J} A_i)$, qui est la version ensembliste classique de (10).

Les hypothèses de l'approche possibiliste semblent bien adaptées à la fusion d'opinions d'experts, comme l'ont montré des études précédentes (Sandri [25]). Dans le cas de capteurs, il semble que l'approche possibiliste puisse être envisagée pourvu qu'on représente l'information qu'ils fournissent à l'aide d'intervalles d'erreur avec leurs niveaux de confiance associés, qu'on ait affaire à des capteurs distincts, éventuellement très différents, et que l'hypothèse que certains d'entre eux fonctionnent mal ne soit pas exclue. Néanmoins, cette proposition a besoin d'être validée sur des expériences concrètes qui vont bien au-delà de la tentative de Deveughèle [7].

Les règles de combinaison envisagées ici sont optimistes au sens où elles postulent toujours que le plus grand nombre possible de sources est fiable, dans l'esprit des règles de combinaison conjonctives telles que la règle de Dempster par exemple. Néanmoins lorsqu'il est peu plausible, voire impossible que toutes les sources disent vrai, des hypothèses plus faibles sont émises, sans pour autant qu'on soit obligé d'identifier les sources dans l'erreur. Par ailleurs, la possibilité d'utiliser des opérations idempotentes s'avère cruciale pour absorber sans dommage les dépendances entre les sources. La règle adaptative proposée cherche à extraire le plus d'information possible de l'ensemble des observations émises par les sources, grâce à l'emploi du degré de consensus pour contrôler le caractère plus ou moins conjonctif de la fusion.

On a vu que l'approche probabiliste est peu adaptée à la représentation de modes disjonctifs et conjonctifs de fusion; cette vision logique est mieux prise en compte par la théorie des possibilités. En revanche les agrégations de type "moyenne" conviennent mal aux représentations possibilistes. Si on veut récupérer les modes de fusion conjonctifs et disjonctifs en théorie des probabilités, il faut pouvoir représenter l'information issue d'une source à l'aide de familles de mesures de probabilité (cf. Coolen [6]). Dans ce cas on retrouve naturellement des règles de combinaison logiques, de type conjonctif ou disjonctif, en faisant l'intersection ou la réunion d'ensembles de mesures de probabilité (par exemple Chateauneuf [2]). L'approche possibiliste peut être vue comme une approximation, facile à calculer, des méthodes de combinaison ensemblistes de familles de probabilité mal connues (cf Annexe). En cela, elle ne vient pas remettre en cause des méthodes probabilistes éprouvées sur des problèmes classiques, mais elle vient élargir le spectre de ces méthodes afin de traiter des problèmes de fusion de données nouveaux parce que peu envisagés auparavant, problèmes qu'on rencontre en robotique, en bases de données et en intelligence artificielle.

Annexe : Justification probabiliste d'opérations ensemblistes floues

Soit π une distribution de possibilité numérique sur un ensemble fini S. Soit $\mathbb{P}(\pi) = \{P, \forall A \supset S, P(A) \leq \prod(A)\}$ l'ensemble des mesures de probabilité P représentables par π sur S. On vérifie que l'on a

$$\forall A \in S,\ \Pi(A) = \sup\{P(A),\ P \in \mathbb{P}(\pi)\}$$

ce qui justifie que $\prod$ représente une famille de mesures de probabilité. Soient π_1 et π_2 deux distributions de possibilité. On peut envisager de combiner conjonctivement π_1 et π_2 en considérant $\mathbb{P}(\pi_1) \cap \mathbb{P}(\pi_2)$. En général, la fonction d'ensemble P^* définie par $P^*(A) = \sup\{P(A),\ P \in \mathbb{P}(\pi_1) \cap \mathbb{P}(\pi_2)\}$ n'est pas une mesure de possibilité, même si $\mathbb{P}(\pi_1) \cap \mathbb{P}(\pi_2) \neq \emptyset$, ou si $\min(\pi_1, \pi_2)$ est normalisé. En particulier on n'a pas $\mathbb{P}(\pi_1) \cap \mathbb{P}(\pi_2) = \mathbb{P}(\min((\pi_1, \pi_2))$. Néanmoins on peut montrer que $\min(\pi_1, \pi_2)$ fournit une bonne approximation de $\mathbb{P}(\pi_1) \cap \mathbb{P}(\pi_2)$, ensemble qu'il n'est pas aisé de représenter exactement.

Soit $\mathbb{P}$ un ensemble de mesures de probabilité, qu'on supposera convexe ($P_1 \in \mathbb{P}, P_2 \in \mathbb{P} \Rightarrow \alpha P_1 + (1-\alpha)P_2 \in \mathbb{P}, \forall \alpha \in [0,1]$), et π une distribution de possibilité.

Définition 1 π est appelée approximation intérieure de $\mathbb{P}$ ssi $\mathbb{P}(\pi) \subseteq \mathbb{P}$. Dans ce cas $\prod(A)$ sert de minorant à $P^*(A) = \sup\{P(A), P \in \mathbb{P}\}$

Définition 2 π est une meilleure approximation intérieure de $\mathbb{P}$ ssi $\mathbb{P}(\pi) \subseteq \mathbb{P}$ et $\forall \pi', \mathbb{P}(\pi) \subseteq \mathbb{P}(\pi') \subseteq \mathbb{P}$ implique $\pi = \pi'$.

Notons que $\mathbb{P}(\pi) \subseteq \mathbb{P}(\pi') \Leftrightarrow \pi \subseteq \pi'$ (inclusion ensembliste floue). On a pu prouver le résultat suivant (Dubois, Prade [17]) :

Théorème 1 : $\min(\pi_1, \pi_2)$ est l'unique meilleure approximation intérieure de $\mathbb{P}(\pi_1) \cap \mathbb{P}(\pi_2)$, dès que $\min(\pi_1, \pi_2)$ est normalisé.

La notion d'approximation extérieure d'un ensemble de probabilités P est définie de la même façon en renversant les inclusions ($\mathbb{P}(\pi) \supseteq \mathbb{P}$). On définit la combinaison disjonctive de deux famille de probabilité en considérant l'union ensembliste de celles-ci. Ainsi $\mathbb{P}(\pi_1) \cup \mathbb{P}(\pi_2)$ correspond à la combinaison disjonctive de π_1 et π_2. Notons que $\mathbb{P}(\pi_1) \cup \mathbb{P}(\pi_2)$ n'est pas convexe généralement. La fermeture convexe de $\mathbb{P}(\pi_1) \cup \mathbb{P}(\pi_2)$ est contenue dans l'ensemble $\mathbb{P}$ de distributions de probabilité engendrées par les bornes

$$P^*(A) = \sup\{P(A), P \in \mathbb{P}(\pi_1) \cup \mathbb{P}(\pi_2)\}.$$

C'est le plus petit ensemble convexe de probabilités contenant $\mathbb{P}(\pi_1) \cup \mathbb{P}(\pi_2)$ et représentable par des bornes de probabilité. Il est facile de vérifier que

$$\begin{aligned} P^*(A) &= \sup\{P(A),\ P(A) \leq \Pi_1(A),\ \text{ou } P(A) \leq \Pi_2(A)\} \\ &= \sup\{P(A),\ P(A) \leq \max(\Pi_1(A),\ \Pi_2(A))\} \end{aligned}$$

et aussi que $\max(\prod_1(A), \prod_2(A)) = \sup_{s \in A} \max(\pi_1(s), \pi_2(s))$ c'est-à-dire que $\max(\prod_1, \prod_2)$ est une mesure de possibilité. Donc $P^* = \max(\prod_1, \prod_2)$. On a donc montré le résultat suivant :

Théorème 2 $\max(\pi_1, \pi_2)$ est l'unique meilleure approximation extérieure possibiliste de $\mathbb{P}(\pi_1) \cup \mathbb{P}(\pi_2)$, et $\mathbb{P}(max(\pi_1, \pi_2))$ contient la fermeture convexe de $\mathbb{P}(\pi_1) \cup \mathbb{P}(\pi_2)$.

BIBLIOGRAPHIE

[1] B.G. Buchanan, and E.H. Shortliffe. *Rule-Based Expert Systems - The MYCIN Experiments of the Stanford Heuristic Programming Project.* Addison-Wesley, Reading, Mass, 1984.

[2] A. Chateauneuf. Combination of compatible belief functions and relation of specificity. In : *Advances in the Dempster-Shafer Theory of Evidence* (M. Fedrizzi, J. Kacprzyk and R.R. Yager, Eds.),Wiley, New York, 1994, 97-114.

[3] L. Cholvy. 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 1993, 66-71.

[4] R.M. Cooke. Uncertainty in risk assessment : A probabilist's manifesto. *Reliability Engineering and Systems Safety*, 23, 1988, 277-283.

[5] R.M. Cooke. *Experts in Uncertainty*. Oxford University Press, Oxford, Grande Bretagne, 1991.

[6] F.P.A Coolen. *Statistical Modeling of Expert Opinions Using Imprecise Probabilities.* PhD Thesis, Eindhoven University of Technology, Eindhoven, The Netherlands, 1994.

[7] S. Deveughèle. *Etude d'une méthode de combinaison graduelle d'informations incertaines dans un cadre possibiliste.* Thèse de doctorat de l'Université de Technologie de Compiègne, 1993

[8] D. Dubois, J. Lang and H. Prade. 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, Wiley, New York, 1992, 38-42.

[9] D. Dubois, S. Moral, H. Prade. A semantics for possibility theory based on likelihoods, In : *Rapport de Recherche IRIT 93-51-R*, Toulouse, 1993. A paraître dans J. Math. Analysis and Applications.

[10] D. Dubois et H. Prade. Weighted minimum and maximum operations in fuzzy set theory. *Information Sciences*, 39, 1986, 205-210.

[11] D. Dubois et H. Prade. *Théorie des Possibilités*, 2ème édition, Masson, Paris, 1987.

[12] D. Dubois et H. Prade. Une approche ensembliste de la combination d'informations incertaines. *Revue d'Intelligence Artificielle*, 1(4), 1987, 23-42.

[13] D. Dubois et H. Prade. Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, 4, 1988, 244-264.

[14] D. Dubois et H. Prade. Default reasoning and possibility theory. *Artificial Intelligence*, 35, 1988, 243-257.

[15] D. Dubois et H. Prade. On the relevance of non-standard theories of uncertainty in modeling and pooling expert opinions. *Reliability Engineering and Systems Safety*, 36, 1992, 95-107.

[16] D. Dubois et H. Prade . 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.). Academic Press, New York, 1992, 481-505.

[17] D. Dubois et H. Prade. When upper probabilities are possibility measures. *Fuzzy Sets and Systems*, 49, 1992, 65-74.

[18] D. Dubois, H. Prade et C. Testemale. Weighted fuzzy pattern matching. *Fuzzy Sets and Systems*, 28, 1988, 313-331.

[19] D. Dubois et H. Prade (1994). Ensembles flous et mesures de possibilité : notions de base. In : *Logique Floue* (ARAGO 14, OFTA), 29-62, Masson, Paris.

[20] D. Dubois, H. Prade, S. Sandri. On possibility-probability transformations. In : *Fuzzy Logic* (R. Lowen et M. Roubens), Kluwer Academic, Dordrecht, Pays-Bas, 1993, 103- 112.

[21] D. J. Kewley. A model for evaluating data fusion systems. *27ème IEEE Asilomar Conference on Signal*, Systems and Computers., 1993.

[22] R.C. Luo, M.G. Kay. Data fusion and sensor integration : state of the art 1990's. In : *Data Fusion in Robotics and Machine Intelligence* (M.A. Abidi, R.C. Gonzalez, eds.), Academic Press, New York, 1992, 7-135.

[23] A. Mosleh, G. Apostolakis. 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, 1984.

[24] S. Park, C.S.G. Lee. Uncertainty fusion of sensory information using fuzzy numbers. *Actes du 5ème Congrès Mondial de l'IFSA* (International Fuzzy System Association), Séoul, 1993, 1001-1004

[25] S. Sandri. *La combinaison de l'information incertaine et ses aspects algorithmiques*. Thèse de Doctorat de l'Université Paul Sabatier, Toulouse, France, 1991.

[26] S.Sandri, A. Besi, D. Dubois, G. Mancini, H. Prade et C. Testemale. Data fusion problems in an intelligent data base interface. In : *Reliability Data Collection and Use in Risk and Availability Assessment* (V. Colombari, Ed.), Springer Verlag, Berlin, 1989, 655-670.

[27] S. Sandri, D. Dubois et H. Kalfsbeek. Elicitation, assessment and pooling of expert judgement using possibility theory. In : Tech. Report IRIT/93-24-R, IRIT, Univ. P. Sabatier, Toulouse, France, 1993. IEEE Trans. on Fuzzy Systems, à paraître.

[28] B. Schweizer et A. Sklar. *Probabilistic Metric Spaces*. North-Holland, New York, 1983.

[29] G. Shafer. *A Mathematical Theory of Evidence*. Princeton Univ. Press, Princeton, NJ., 1976.

[30] G. Shafer. Non-additive probabilities in the works of Bernoulli and Lambert. *Archives for the History of Exact Sciences*, 19, 1978, 309-370

[31] C. Wagner et K. Lehrer. *Rational Consensus in Science and Society*. D. Reidel, Dordrecht, 1981.

[32] R.R. Yager. Aggregating evidence using quantified statements. *Information Sciences*, 36, 1985, 179-206.

[33] R.R. Yager. Quasi-associative operations in the combination of evidence. *Kybernetes*, 16, 1987, 37-41.

[34] R.R. Yager. Non-monotonic set-theoretic operators. *Fuzzy Sets and Systems*, 42, 1991, 173-190.

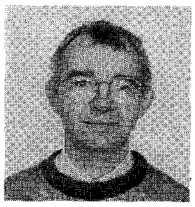
[35] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8, 1965, 338-353.

[36] L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1, 1978, 3-28.

Manuscrit reçu le 19 mai 1994.

LES AUTEURS

D. DUBOIS



H. PRADE

Didier Dubois et Henri Prade sont directeurs de recherche au CNRS et travaillent ensemble à l'Institut de Recherche en Informatique de Toulouse qui dépend également de l'Université Paul Sabatier, et de l'Institut National Polytechnique de Toulouse. Leurs centres d'intérêt sont la logique floue, l'intelligence artificielle, la recherche opérationnelle et la théorie de la décision. Leurs thèmes de recherche concernent la modélisation de l'imprécis et de l'incertain dans divers domaines tels que le raisonnement automatisé, la fusion d'information, le diagnostic de pannes et les problèmes de satisfaction de contraintes. Ils ont écrit depuis 1980, deux monographies spécialisées autour de la logique floue et de la formalisation du raisonnement, dont une en français sur la théorie des possibilités.