FORMULATION ET TRAITEMENT DE L'INCERTAIN EN ANALYSE MULTI-SENSEURS

Alain APPRIOU

ONERA
BP 72, 92322 Chatillon Cedex, Tél.: 46734919, Fax.: 46734149

RÉSUMÉ

Un concept de traitement d'informations multi-senseurs, propre à gérer la fiabilité des données délivrées, est proposé sur la base de la théorie de l'évidence. Son intérêt est dégagé dans le cadre de deux types de problèmes complémentaires, à partir de quelques simulations simples.

ABSTRACT

A concept for multisensor informations processing, which aims at managing data reliability, is set up on the basis of the theory of evidence. Its interest is emphasized within the framework of two complementary kinds of problems, thanks to a few simple simulations.

1. BESOINS ET DIFFICULTES

L'intérêt majeur qui justifie l'association de senseurs multiples tient au bénéfice qui peut être tiré de leur complémentarité [2]. Leur vocation est donc avant tout de traiter des situations où une partie des senseurs est plus ou moins en défaut (aptitudes pour la situation à traiter, environnement, contre-mesures, défaillances,...). Un traitement judicieux de ce type de situation exige la prise en compte de toutes les informations disponibles, notamment celles susceptibles de nous renseigner sur la qualité des mesures, sur leur potentiel informatif, sur le contexte des relevés, et donc sur la pertinence relative des différentes évaluations qui peuvent être menées.

Ceci conduit à inférer dans un même système des données de nature particulièrement disparate, et souvent subjectives, incertaines, incomplètes, erronées, imprécises, avec des relations de dépendance très spécifiques. On devra typiquement intégrer le traitement conjoint de mesures, d'apprentissages plus ou moins représentatifs, et d'informations qualitatives sur ces données. Les théories de l'incertain procurent un cadre fédérateur séduisant dans ce contexte, mais leur mise en œuvre pratique se heurte à un certain nombre de difficultés : interprétation et modélisation des informations disponibles dans les cadres théoriques appropriés, choix d'une architecture de fusion et de règles de combinaison, principes de décision à adopter, contraintes sur la rapidité et le volume des calculs nécessaires ([1], [3]).

Une solution à ces différents obstacles est proposée dans le cadre d'un problème générique où l'on dispose d'évaluations concurrentes, et d'informations sur la qualité de ces évaluations. Les deux types de situations pratiques d'intérêt présentées respectivement dans [1] et [3], différentes par la nature des informations considérées, sont fédérés par la solution générale dégagée. L'apport de l'approche préconisée est mis en évidence sur la base de quelques simulations simples de synthèse. Ces techniques ont par ailleurs permis de conduire un certain nombre d'applications pratiques à l'ONERA.

2. THEORIES DE L'INCERTAIN

Pour l'exposé qui suit, nous retiendrons a priori la théorie de l'évidence pour le cadre le plus large et le plus fédérateur des notions qu'elle manipule, et pour les performances de sa loi de combinaison, mises en évidence par ailleurs [3].

Cette théorie, développée dans [4] et brièvement synthétisée dans [1], suppose la définition préalable d'un ensemble E de N éléments Hi ($1 \leq i \leq N$) exclusifs et exhaustifs, appelé cadre de discernement. 2^E désigne alors l'ensemble des 2^N-1 sous-ensembles Aj de E ($1 \leq j \leq 2^N-1$). Une fonction de masse élémentaire m(.) est définie de 2^E sur [0,1] par :

$$\sum_{1 \leq j \leq 2^N-1} m(Aj) = 1 \qquad \text{et} \qquad m(\emptyset) = 0$$

Les éléments focaux sont les éléments Aj de 2^E dont la masse m(Aj) est non nulle. Deux autres fonctions de base sont la crédibilité Cr(.) et la plausibilité Pl(.), qui peuvent être interprétées comme une mesure de vraisemblance, respectivement minimale et maximale, d'un événement :

$$Cr(Bk) = \sum_{Aj \subset Bk} m(Aj) \qquad Pl(Bk) = \sum_{Aj \cap Bk \neq 0} m(Aj)$$

La combinaison m(.) de M sources distinctes mj(.) ($1 \leq j \leq M$) est par ailleurs obtenue par sommation orthogonale :

$$m(A) = (1-K)^{-1} * \sum_{A^1\cap\ldots\cap A^M=A\neq\emptyset} \left\{ \prod_{1\leq j\leq M} mj(Aj) \right\}$$

où Aj désigne un sous-ensemble quelconque de E, considéré par la source mj(.). K est l'inconsistance de la fusion, propre à figurer le degré de conflit entre les sources :

$$K = \sum_{A^1\cap\ldots\cap A^M=\emptyset} \left\{ \prod_{1\leq j\leq M} mj(Aj) \right\}$$

Enfin, l'affaiblissement $m^a(.)$ d'un jeu de masses m(.) permet de réduire globalement dans un rapport a ($0\leq a\leq 1$) son degré de certitude, en fonction de la confiance que l'on est amené à lui accorder par ailleurs :

$$m^a(A) = (1-a)*m(A)\,,\quad \forall A\neq E, \quad \text{et} \quad m^a(E) = a+(1-a)*m(E)$$

3. PROBLEME GENERIQUE

On propose de considérer un problème très général de discrimination entre N hypothèses Hi à partir d'observations relevées par M senseurs Sj qui procurent :

- N*M critères Cij à valeurs dans [0, 1], respectivement délivrés par chaque senseur j pour caractériser la vraisemblance de chaque hypothèse Hi,

- N*M facteurs de qualité qij à valeurs dans [0, 1], relatifs à chaque vraisemblance Cij, et capable de traduire par apprentissage son aptitude à discriminer l'hypothèse Hi.

Il peut par exemple s'agir de la reconnaissance d'une cible observée, parmi N cibles répertoriées a priori. On se place ici dans le cas d'intérêt pratique où les critères Cij sont élaborés par des chaînes d'information distinctes, qui justifient la différenciation de leur pertinence respective par des facteurs qij. on suppose par ailleurs être dans le contexte le plus fréquent où les critères Cij pris isolément ont toujours au moins valeur de réfutation, dans le sens où une valeur minimale garantit que l'hypothèse associée n'est pas vérifiée.

Ceci conduit à formaliser le problème sur la base de deux axiomes :

1-Chacun des N*M couples [Cij, qij] constitue une source d'information distincte, ayant pour éléments focaux: Hi, ¬Hi, et E, où le cadre de discernement E représente l'ensemble des N hypothèses.

2-Cij = 0, lorsqu'il est valide (qij = 1), permet d'affirmer que Hi n'est pas vérifiée.

L'axiome 1 impose d'élaborer N*M jeux de masses mij(.). Pour chacun, la masse des éléments focaux Hi, ¬Hi, et E, est d'abord définie par la valeur du critère Cij correspondant, qui ne peut être interprété qu'en termes de crédibilité ou de plausibilité de Hi. mij(.) est alors obtenu en affaiblissant le résultat dans le rapport (1-qij), où qij est le facteur de confiance associé à Cij. L'axiome 2 limite cependant à 2 le nombre des modèles admissibles :

<u>Modèle 1</u> : $mij(Hi) = 0$, $mij(\neg Hi) = qij*(1-Cij)$, $mij(E) = 1-qij*(1-Cij)$

<u>Modèle 2</u> : $mij(Hi) = qij*Cij$, $mij(\neg Hi) = qij*(1-Cij)$, $mij(E) = 1-qij$

Un jeu de masse m(.) synthétisant l'ensemble des évaluations est ensuite élaboré par sommation orthogonale des différents jeux de masses mij(.), dans le cadre de chaque modèle. Un critère de maximum de plausibilité, justifié dans [1], permet alors de dégager de m(.) l'hypothèse Hi la plus vraisemblable. Les 2 modèles introduits conduisent ainsi respectivement aux deux solutions :

<u>Solution 1</u> : $\max_i \{ \prod_j [1-qij*(1-Cij)] \}$

<u>Solution 2</u> : $\max_i \{ \prod_j [1-qij*(1-Cij)]/[1-qij*Cij] \}$

La détermination pratique des Cij et qij est un problème spécifique au type d'application traité. Deux situations différentes par la nature des informations disponibles sont présentées dans la suite. Elles sont représentatives des deux grandes classes de problèmes rencontrées dans les applications d'intérêt traitées à ce jour.

4. CLASSIFICATION AVEC MATRICE DE CONFUSION

Le problème précédent est particularisé en ce que les cœfficients de confiance qij sont élaborés à partir d'un apprentissage de la matrice de confusion relative à chaque senseur Sj testé isolément. La nature particulière de cette information conduit à considérer que chaque source [Cij, qij] introduite au §2 est elle-même la somme orthogonale de N sources [Ckj, qkij] ($1\leq k\leq N$), strictement de même nature, où :

$$qkij = P[Ckj=\max_h\{Chj\}/Hi]$$

Les deux solutions dégagées s'écrivent alors simplement :

<u>Solution 1</u> : $\max_i \{ \prod_k \prod_j [1-qkij*(1-Ckj)] \}$

<u>Solution 2</u> : $\max_i \{ \prod_k \prod_j [1-qkij*(1-Ckj)]/[1-qkij*Ckj] \}$

A titre de référence, la solution probabiliste adaptée au problème suppose une décision décentralisée au niveau de chaque senseur :

<u>Solution 3</u> : $\max_i \{ \prod_j qkij \}$, avec k tel que: $Ckj = \max_h \{ Chj \}$

Enfin, l'ignorance de la matrice de confusion réduit la décision à :

Solution 4 : $\max_i \{ \prod_j C_{ij} \}$

Ces différentes solutions ont été introduites et analysées parmi d'autres dans [3]. A titre indicatif, la figure 2 présente un résultat de simulation simple en reconnaissance de cibles dans le cas de trois cibles et de deux senseurs. Les taux de reconnaissance moyens pour chaque cible, et pour toutes cibles confondues, ont été obtenus en générant des valeurs aléatoires de critères Cij, ceci selon des lois équiréparties sur [Bij, Bij+0.30] lorsque la cible i est effectivement la cible présentée, et sur [0.35, 0.65] sinon. Les bornes Bij sont choisies de façon à satisfaire les matrices de confusion données pour chaque senseur. les taux de reconnaissance de ces matrices sont visualisés en figure 1, les taux de confusion étant équirépartis sur les cibles concernées. L'intérêt pratique du cas présenté réside dans la complémentarité du pouvoir discriminant des deux senseurs.

Les résultats mettent en évidence tout l'intérêt d'une décision centralisée, de la prise en compte d'informations sur la validité des données disponibles, et surtout de l'approche proposée sur la base de la théorie de l'évidence, qui permet d'intégrer ces différents aspects de façon pertinente.

5. DISCRIMINATION STATISTIQUE SUPERVISEE

A l'inverse de l'exemple précédent, les cœfficients de confiance qij conservent toute leur généralité, mais les critères Cij doivent être élaborés à partir de mesures mj délivrées par les différents senseurs Sj, et d'un apprentissage des densités de probabilité P(mj/Hi) des mesures mj sous les différentes hypothèses Hi. Ces densités sont en pratique plus ou moins représentatives des densités réellement rencontrées compte tenu de l'évolution des conditions d'observation. Les cœfficients qij caractérisent ici ce degré de représentativité.

Dans le cas ou les densités apprises sont parfaitement représentatives de la statistique réelle (qij=1, $\forall$i,j), les modèles retenus doivent rester cohérents avec la solution optimale donnée par les probabilités. Ils doivent en particulier respecter 2 autres axiomes :

3- Les modèles retenus conduisent au résultat fourni par l'inférence bayésienne, lorsque les distributions P(mj/Hi) apprises sont parfaitement représentatives des densités réellement rencontrées (qij=1, $\forall$i,j), et lorsque les probabilités a priori P(Hi) sont connues.

4- Les sources Sj étant indépendantes, si les densités P(mj/Hi) sont bien représentatives de la réalité, le résultat est le même lorsqu'on applique la modélisation retenue aux P(mj/Hi) et que l'on en fait la somme orthogonale selon j, ou lorsqu'on l'applique à la densité conjointe P(m1,... ,mM/Hi).

Une recherche exhaustive des modèles qui satisfont ces deux axiomes 3 et 4, ainsi que l'axiome 1 du §3, est menée dans [1]. Deux modèles s'en dégagent, qui peuvent être directement rattachés aux deux modèles généraux introduits au §3, et respectent par le fait l'axiome 2 de ce même §; il convient pour cela d'adopter les définitions suivantes :

Pour le modèle 1 : $C_{ij} = R_j * P(m_j/H_i)$

Pour le modèle 2 : $C_{ij} = R_j * P(m_j/H_i)/[1+R_j * P(m_j/H_i)]$

Dans les deux cas Rj est un gain de normalisation tel que :

$$0 \leq R_j \leq [\max_i \{ P(m_j/H_i) \}]^{-1}$$

Il est à noter que, par construction, les modèles résultants incluent la solution bayésienne, obtenue en posant qij=1 pour tout i et j, et optimale en l'absence d'erreurs.

La figure 3 présente les probabilités moyennes de bonne reconnaissance fournies par la simulation d'un problème à 2 senseurs et 2 hypothèses. Les deux senseurs sont identiques, avec un bon pouvoir discriminant a priori, mais aussi avec un apprentissage douteux sous l'hypothèse H2, en liaison avec une évolution supposée possible du contexte attaché à cette hypothèse. En pratique les apprentissages considérés sont des gaussiennes :

$P(m1/H1) = P(m2/H1) = N(0,1)$, avec $q11 = q12 = 1$
$P(m1/H2) = P(m2/H2) = N(6,1)$, avec $q21 = q22 = q$

Les mesures réellement simulées satisfont en revanche :

$P(m1/H1) = P(m2/H1) = N(0,1)$
$P(m1/H2) = N(2,1)$, $P(m2/H2) = N(S,1)$

On teste donc un cas où le senseur 1 est effectivement mauvais, et où la qualité du senseur 2 varie en fonction du signal S. Le choix des facteurs qij traduit en effet que l'on est dans une situation où l'erreur risque d'être importante simultanément sur les deux capteurs.

Les résultats présentés sont significatifs de la robustesse apportée, sur l'intervalle des valeurs de S alors concernées, par l'approche proposée (courbe q=0,9) par rapport à l'approche bayésienne (courbe q=1), et même par rapport à un quelconque fonctionnement mono-capteur, ce que ne satisfait absolument pas l'inférence bayésienne.

6. PERSPECTIVES

L'intérêt de la solution générale proposée au §3 a été mis en évidence sur la base de deux types de problèmes complémentaires. Il reste maintenant à en étendre le domaine de mise en œuvre de ce concept à la prise en compte d'informations plus variées, en recherchant une méthodologie plus systématique d'élaboration des critères Cij et des cœfficients qij, qui utilise notamment les cadres théoriques compatibles (probabilités, ensembles flous, théorie des possibilités, théorie des fonctions d'utilité, mesures d'information,...).

Les domaines applicatifs pour lesquels le bénéfice de ce type d'approche a pu être dégagé concernent essentiellement :

- la reconnaissance de cible, pour traiter la pollution de l'apprentissage, son manque de représentativité ou son insuffisance, et la disparité du potentiel informatif ;

- le guidage d'engin pour assurer la robustesse de l'extraction de cibles ou du recalage de la navigation par reconnaissance d'amers ;

- l'intégration dans les méthodes de poursuite d'attributs d'identité, ou de règles contextuelles floues ;

- l'analyse de situations pour l'aide à la décision.

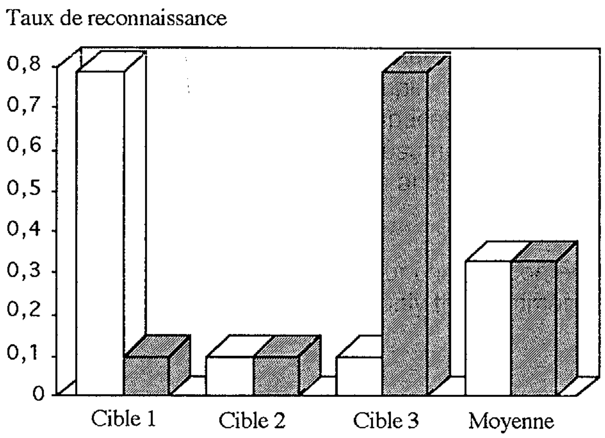

Figure 1 - Classification avec matrice de confusion
Senseur 1 Senseur 2

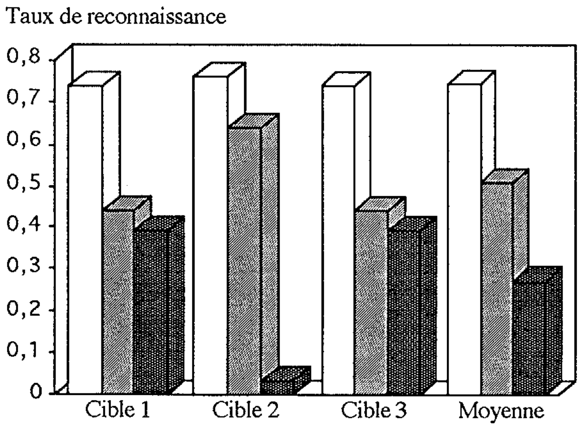

Figure 2 - Classification avec matrice de confusion
Solution 1 Solution 3 Solution 4

BIBLIOGRAPHIE

[1] A. APPRIOU, "Probabilités et incertitude en fusion de données multi-senseurs", Revue Scientifique et Technique de la Défense, n°11, 1991-1, pp27-40.

[2] A. APPRIOU, "Perspectives liées à la fusion de données", Science et Défense, Paris, mai 1990.

[3] A. APPRIOU, "Intérêt des théories de l'incertain en fusion de données", Conférence invitée, Colloque International sur le Radar, Paris, 24-28 avril 1989.

[4] G. SHAFER, "A mathematical theory of evidence", Princeton University Press, Princeton, New Jersey, 1976.

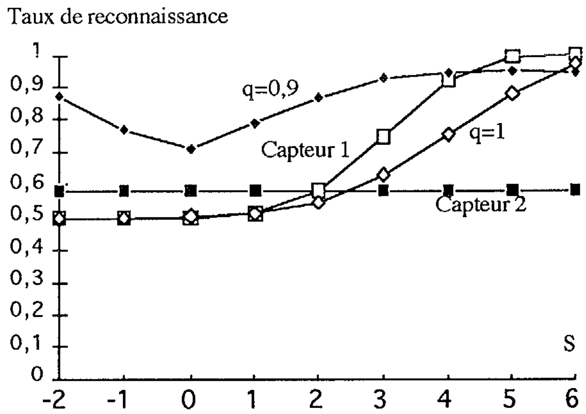

Figure 3 - Discrimination par apprentissage statistique.
Evolution du signal S sur le premier capteur.