UTILISATION DE LA THÉORIE DE DEMPSTER-SHAFER POUR LA FUSION D'INFORMATIONS

Vincent NIMIER, Alain APPRIOU

ONERA,
BP 72
92322 CHATILLON Cedex
FRANCE

RÉSUMÉ

La fusion de données, et plus généralement la fusion d'informations nécessite un cadre théorique commun pour manipuler et combiner des sources d'information dont l'origine est différente. Nous montrons dans cette communication que la théorie de Dempster-Shafer peut être utilisée en ce sens dans un problème de classification et qu'elle permet en outre de décomposer chaque système d'information en sources élémentaires quelles que soient la nature et l'origine de l'information manipulée. Les sources élémentaires sont ensuite combinées pour accéder à l'information globale fournie par le système.

ABSTRACT

Data fusion and more generally information fusion needs a common theoretical framework to manipulate and combine information sources that can have different natures. This paper shows that Dempster Shafer theory can be used in a classification problem and permits to decompose each information system on several elementary sources whatever the nature and the origin of information to manipulate. All these elementary sources are then combined to get the global information provided by the system.

1.INTRODUCTION

La fusion de données est apparue, ces dernières années comme un domaine d'investigation privilégié. En effet, l'utilisation concomitante de divers senseurs se révèle être un moyen tout à fait approprié pour rendre compte d'un environnement aux facettes multiples, avec une précision et une fiabilité accrues par rapport à l'exploitation d'un unique moyen de mesure doté des traitements les plus sophistiqués.

Cependant la fusion de données n'est pas sans générer certaines difficultés ; leurs origines étant pour l'essentiel nouvelles puisque intrinsèquement liées à l'utilisation simultanée de plusieurs capteurs et de sources d'informations hétérogènes. Les théories probabilistes peuvent dans certains cas apparaître insuffisantes et même conduire à des interprétations paradoxales dès lors qu'elles sont utilisées et confrontées à des problèmes particuliers ; d'où le besoin de théories nouvelles.

Pour le problème générique de classification considéré ici, la théorie de Dempster-Shafer s'inscrit dans cette nécessité. Sans être en contradiction avec les théories probabilistes, elle apparaît comme une extension et une alternative apportant une flexibilité dans la modélisation et une représentativité de l'information plus importante.

Nous commencerons par rappeler les notions de base de la théorie de Dempster-Shafer [7] pour aborder ensuite les problèmes de modélisation de l'information suivant sa nature : objective ou subjective. Nous illustrerons cette modélisation par un exemple de fusion mettant en œuvre deux classifieurs et des données subjectives.

2.THÉORIE DE DEMPSTER-SHAFER

2.1 Notions de base

Soit Θ un ensemble appelé cadre de discernement et contenant N éléments H_i ($i \in \{1, \ldots, N\}$), exclusif et exhaustif. Une masse élémentaire m(.) est une application de 2^Θ (l'ensemble des parties de Θ) dans [0, 1] vérifiant les deux propriétés suivantes :

$$m(\emptyset) = 0 \tag{2.1}$$

$$\sum_{A_i \in 2^\Theta} m(A_i) = 1 \tag{2.2}$$

Les éléments $A_i \in 2^\Theta$ tels que $m(A_i) \neq 0$ sont appelés les éléments focaux. Dès lors que l'ensemble des éléments focaux se réduisent aux seuls singletons H_i du cadre de discernement, la masse élémentaire m(.) est assimilable à une probabilité P(.); la relation 2.2 ci-dessus est alors la relation classique de normalisation des probabilités :

$$\sum_{i=1}^{N} P(H_i) = 1 \tag{2.3}$$

L'apport de la théorie de l'évidence par rapport à la théorie des probabilités tient au fait qu'elle permet une évaluation conjointe d'un ensemble quelconque d'hypothèses H_i. Par exemple, si le cadre de discernement ne contient que deux hypothèses : $\Theta = \{ H_1, H_2 \}$ il est possible d'affecter une masse élémentaire à chacune des hypothèses H_1 et H_2 mais aussi à la partie qui contient les deux hypothèses $\{ H_1, H_2 \}$ qui est ici le cadre de discernement Θ. A titre d'illustration, le cas peut se produire lors d'un sondage effectué avant une élection comprenant deux candidats. $m(H_1)$ est la proportion de personnes interrogées se prononçant en faveur du candidat H_1, $m(H_2)$ la proportion de personnes interrogées se prononçant en faveur du candidat H_2, et $m(\{ H_1, H_2 \})$ la proportion de personnes qui sont encore indécises. Au moment de l'élection et en supposant que l'ensemble des personnes interrogées doivent nécessairement se prononcer pour l'un ou l'autre des candidats, la proportion de personnes indécises pourra se reporter totalement sur le candidat H_1 ou sur le candidat H_2 ou bien partiellement sur l'un et l'autre des deux candidats.

Deux cas limites de ce report d'incertitude peuvent être évalués par les fonctions de crédibilité Cr(.) et de plausibilité Pl(.) qui, à partir du jeu de masse élémentaire, sont définies sur 2^{Θ} grâce aux relations suivantes :

$$Cr(B_j) = \sum_{A_i \subset B_j} m(A_i) \tag{2.4}$$

$$Pl(B_j) = \sum_{A_i \cap B_j \neq \emptyset} m(A_i) \tag{2.5}$$

La crédibilité représente la masse minimum qu'il est possible d'affecter à une partie B de Θ. La plausibilité représente à l'inverse la masse maximum.

Ces fonctions peuvent être considérées, lorsqu'elles sont évaluées sur l'ensemble des singletons $\{H_i\}$, comme étant les probabilités minimum et maximum qu'il est possible d'associer à chaque hypothèse, compte tenu de l'ensemble des incertitudes inhérentes au problème.

En outre, quel que soit $A \in \Theta$, la relation suivante est toujours vérifiée :

$$Pl(A) + Cr(\bar{A}) = 1 \tag{2.6}$$

Cette relation exprime le fait que lorsqu'un report de masse se fait en faveur d'une partie A de Θ, il s'effectue toujours au détriment de son contraire.

2.2 Règle de combinaison

Chaque source d'information, notée S_i, est à l'origine d'un jeu de masse élémentaire noté $m_i(.)$, caractéristique de l'information qu'elle véhicule. En présence de plusieurs sources, la règle de Dempster permet de combiner l'ensemble des jeux de masses et de construire ainsi un jeu de masse unique caractéristique de l'information globale du système.

La combinaison de deux jeux de masses élémentaires $m_1(.)$ et $m_2(.)$ donne lieu à un jeu de masse unique noté m(.) avec comme notation :

$$m(.) = m_1(.) \oplus m_2(.) \tag{2.7}$$

et comme expression:

$$m(A_i) = (1 - K)^{-1} \sum_{B_i \cap B_j = A_i \neq \emptyset} m_1(B_i)\, m_2(B_j) \tag{2.8}$$

où K est l'inconsistance de la fusion qui est calculée par la formule :

$$K = \sum_{B_i \cap B_j \neq \emptyset} m_1(B_i)\, m_2(B_j) \tag{2.9}$$

Une telle règle de combinaison est à la fois commutative et associative.

2.3 Théorème de Bayes généralisé

Soient X et Θ deux ensembles discrets pouvant être utilisés comme cadres de discernement. Considérons le jeu de masse conditionné : $m_X(. /H_i)$ qui représente un jeu de masse sur l'ensemble des variables observables sachant que l'hypothèse H_i est présente. Le théorème de Bayes généralisé défini par Smets [2] prend la forme suivante :

$$m_\Theta(\bar{H}_i / x) = Cr_X(\bar{x} / H_i)$$

$$m_\Theta(\Theta / x) = 1 - Cr_X(\bar{x} / H_i) \tag{2.10}$$

$\bar{H}_i$ représente ici la partie composée de l'ensemble des éléments de Θ à l'exclusion de H_i, et $\bar{x}$ représente la partie composée de l'ensemble des éléments de X à l'exclusion de x. $Cr_X(./H_i)$ est la crédibilité calculée à partir du jeu de masse $m_X(. /H_i)$. On remarque que l'application du théorème de Bayes généralisé ne nécessite aucune loi a priori.

3. SOURCE D'INFORMATION ÉLÉMENTAIRE POUR LA FUSION

Un jeu de masse est représentatif d'une source d'information. Dans beaucoup de problèmes de classification on peut identifier au moins deux types de sources d'information élémentaire suivant que l'information provient d'un apprentissage préalable, qualifiée ici d'objective, ou bien qu'elle est issue d'une expertise humaine, qualifiée ici de subjective.

3.1 Source d'information objective.

3.1.1 Méthode Bayésienne

Les performances globales d'un classifieur sont généralement fournies par sa matrice de confusion. L'estimation de celle-ci se fait grâce à un apprentissage préalable qui consiste à présenter au classifieur un certain nombre d'exemples représentatifs de chacune des hypothèses du cadre de discernement et ensuite d'évaluer la probabilité empirique de chaque décision. Les performances d'un classifieur sont donc totalement déterminées dès lors que l'ensemble des probabilités : $P(d = H_i / H_j)$ $i, j \in [1 \dots N]$, est accessible.

Lors de l'utilisation effective du classifieur, et en présence d'une décision $d=H_1$, la donnée de la loi a priori sur les hypothèses $P(H_i)$ permet, par la formule de Bayes, d'accéder à la probabilité a posteriori $P(H_i \ / d = H_1)$ et de choisir l'hypothèse qui maximise cette probabilité.

Cependant la probabilité a posteriori ne peut se calculer que lorsque l'ensemble de toutes les probabilités $P(d = H_i / H_j)$ sont totalement connues ; ce qui rend inexploitable toute base d'apprentissage incomplète conduisant à une connaissance partielle des probabilités $P(d = H_i / H_j)$. Par ailleurs, la loi a priori est indispensable au calcul des probabilités a posteriori, cependant cette loi est généralement l'une des inconnues du problème. Une hypothèse d'équirépartition des probabilités a priori est donc souvent nécessaire à l'utilisation du théorème de Bayes sous la forme probabiliste.

3.1.2 Utilisation de la théorie de Dempster-Shafer

A contrario, la théorie de Dempster-Shafer permet de construire un jeu de masse sur l'ensemble des hypothèses avec une connaissance même partielle des performances du classifieur, et sans l'utilisation de loi a priori.

Une base d'apprentissage, pour l'évaluation des performances d'un classifieur, peut être considérée comme étant constituée d'autant de bases d'apprentissage élémentaires qu'il existe d'hypothèses à classifier. Chaque base d'apprentissage élémentaire ne contient des exemples que de cette seule hypothèse, si bien qu'il est possible de calculer les probabilités associées aux performances du classifieur pour chaque base élémentaire. Par exemple, une base d'apprentissage contenant des exemples de l'hypothèse H_j permet d'évaluer l'ensemble des probabilités $P(d = H_i / H_j)$ avec $i \in [\ 1 \dots N]$. Le jeu de masse est construit sur le cadre de discernement $D = \{ \ d = H_1 , \dots , d = H_N \}$ par l'identification suivante :

$$m_j(d = H_1 \ / H_j) = P(d = H_1/ H_j)$$
$$\dots$$
$$m_j(d = H_N / H_j) = P(d = H_N/ H_j) \qquad (3.1)$$

et

$m_j(A \ / \ H_j) = 0$ pour toute partie A différente des singletons de D.

Le théorème de Bayes généralisé (formule 2.10) permet de définir un jeu de masse sur Θ :

$$m_j(\overline{H_j} \ / \ d = H_i) = 1 - P(d = H_i \ / \ H_j)$$
$$m_j(\Theta \ / \ d = H_i) = P(d = H_i \ / \ H_j) \qquad (3.2)$$

Pour $i \in [\ 1 \dots N]$

Comme l'a montré par une approche différente A. Appriou dans [4], ce jeu de masse peut être obtenu lorsque X est un espace continu.

L'interprétation de ce jeu de masse est la suivante : supposons que la seule information disponible soit la donnée de la probabilité $P(d = H_1 \ / \ H_1) = 1$ qui signifie qu'à chaque fois que l'hypothèse H_1 est présente devant le classifieur celui-ci fournit la réponse $d = H_1$, les autres probabilités étant par ailleurs toutes inconnues. La donnée de la décision $d = H_1$ pour l'identification de l'hypothèse présente n'apporte aucune information puisque toutes les hypothèses peuvent être à l'origine d'une décision $d = H_1$, ce qui justifie le jeu de masse : $m_1(\Theta \ / \ d = H_1) = 1$. A contrario, si l'unique information est $P(d = H_1 \ / \ H_1) = 0$ alors la donnée de $d = H_1$ exclut nécessairement H_1 du nombre des hypothèses présentes on obtient donc $m_1(\overline{H_1} \ / \ d = H_1) = 1$. Le jeu de masse ci-dessus peut donc être vu comme une généralisation du modus tollens à des valeurs différentes de 0 et de 1.

Si M (M<N) jeux de masse élémentaires sont disponibles, le jeu de masse global sera obtenu par combinaison des M jeux de masse élémentaires.

$$m(./d = H_i) = m_1(./d = H_i) \oplus \dots \oplus m_M(./d = H_i) \qquad (3.3)$$

Pour $i \in [\ 1 \dots N]$

La cohérence avec le théorème de Bayes probabiliste est conservée puisqu'un jeu de masse conforme aux probabilités a posteriori obtenu par la formule de Bayes, est obtenu en posant:

$$m_p(H_1) = P(H_1)$$
$$\dots$$
$$m_p(H_N) = P(H_N) \qquad (3.4)$$

et en combinant ce jeu de masse a priori avec les N jeux de masse issues des N bases élémentaires.

3.2 Source d'information subjective.

Dans de nombreuses applications, l'ensemble des informations disponibles n'est pas nécessairement le résultat d'un apprentissage préalable. La prise en compte de certaines connaissances humaines sous la forme d'heuristiques peut à bien des égards faciliter les prises de décision et augmenter la fiabilité globale du système. Par exemple, dans un problème de reconnaissance d'aéronefs, il peut être judicieux de considérer le paramètre vitesse. Malheureusement on comprend aisément que la constitution d'une base d'apprentissage, pour la caractérisation d'un classifieur basé sur ce type de paramètre, soit totalement irréaliste. Un expert du domaine peut néanmoins fournir des valeurs (même imprécises) maximum et minimum de la vitesse pour chaque aéronef.

Ce type d'information peut être extrait et modélisé par des sous-ensembles flous. On peut considérer par exemple deux aéronefs : Hélicoptère et Avion qui représentent les deux hypothèses $\{ \ H_1 , \ H_2 \ \}$ du problème de classification. Chacune des plages de variation de la vitesse, fournie par l'expert, sera modélisée par des sous-ensembles flous que nous noterons F^1 et F^2 avec :

F^1 = plage de variation de H_1
F^2 = plage de variation de H_2

Soient $\mu_{F^1}(v)$ et $\mu_{F^2}(v)$ les deux fonctions d'appartenance des deux sous-ensembles flous F^1 et F^2. Un jeu de masse élémentaire peut être construit pour chaque fonction d'appartenance [3]. Il convient, pour cela de considérer l'ensemble des α-coupes de chaque fonction d'appartenance $F^j_{\alpha i}$ avec :

$$F^j_{\alpha i} = \{ v / \mu_{F^j}(v) \geq \alpha_i \} \qquad (3.5)$$

avec $\alpha_1 > \alpha_2 > .... > \alpha_n$ et $\alpha_{n+1} = 0$ pour un sous-ensemble flou discret et fini.

L'ensemble des α-coupes vérifie la propriété de consonance: $F^j_{\alpha 1} \subset F^j_{\alpha 2} ..... \subset F^j_{\alpha n}$. Le jeu de masse associé est donc le suivant :

$$m(F^j_{\alpha i} / H_1) = \alpha_i - \alpha_{i+1} \qquad (3.6)$$

Ce jeu de masse est un jeu de masse conditionné puisque le sous-ensemble flous qui lui est associé correspond à la plage de variation du véhicule sachant que celui-ci est du type H_j.

L'application du théorème de Bayes généralisé fournit le jeu de masse élémentaire suivant sur l'espace des hypothèses :

$$\begin{aligned} m_j(\bar{H}_j / v) &= 1 - \mu_{F^j}(v) \\ m_j(\Theta / v) &= \mu_{F^j}(v) \qquad (3.7) \end{aligned}$$

4. FUSION DE SOURCES D'INFORMATIONS

Nous venons de montrer que chaque source d'information peut être modélisée par un jeu de masse élémentaire qui correspond soit à une base d'apprentissage élémentaire pour une source d'information objective, soit à un sous ensemble flou pour une source d'information subjective. Ces jeux de masse peuvent ensuite être combinés en un jeu de masse unique représentatif du système dans sa globalité.

Par exemple, supposons un système composé de deux classifieurs C1 et C2 et d'une connaissance subjective représentée par deux sous-ensembles flous F^1 et F^2. Le cadre de discernement contient deux hypothèses distinctes. Supposons, par ailleurs, que le classifieur C1 fournit une décision d1, que le classifieur C2 fournit une décision d2, et que la vitesse mesurée de la cible est v_m. Le jeu de masse résultant est alors :

$$\begin{aligned} m(./d1, d2, v_m) = &\, m^{C1}{}_1(./d1) \oplus m^{C1}{}_2(./d1) \\ &\oplus m^{C2}{}_1(./d2) \oplus m^{C2}{}_2(./d2) \\ &\oplus m^{F1}(./ v_m) \oplus m^{F2}(./ v_m) \qquad (4.1) \end{aligned}$$

où $m^{Ci}{}_j(.)$ est le jeu de masse du classifieur i obtenu grâce à la probabilité carractéristique de l'hypothèse H_j. $m^{F1}(.)$ et $m^{F2}(.)$ sont les masses obtenues grâce aux deux sous-ensembles flous. Le jeu de masse résultant de cette fusion représente donc toute l'information que l'on peut déduire sur les hypothèses en fonction des variables mesurées.

Le jeu de masse étant construit, le choix de l'hypothèse présente devient alors crucial. Ce problème de décision est plus délicat que dans un contexte purement probabiliste puisqu'il n'est plus possible de choisir l'hypothèse la plus probable. Cette approche fournit en revanche une flexibilité et un degré de liberté plus grand dans la prise de décision puisqu'il est désormais possible de choisir entre plusieurs attitudes suivant que l'on recherche une attitude prudente ou moyenne. Les critères généralement admis pour le choix de l'hypothèse la plus favorable sont le maximum de plausibilité, le maximum de crédibilité ou bien le maximum des probabilités pignistiques [1], [5], [6]. On doit cependant remarquer que la multiplicité des sources d'informations a généralement tendance à favoriser une augmentation des masses affectées aux singletons et une diminution des masses affectées aux autres parties du cadre de discernement ; tous les critères précédents deviennent alors, lorsque le nombre de sources est suffisant, équivalent au maximum de probabilité a posteriori.

5. CONCLUSION

Nous avons montré dans cet article que la théorie de Dempster-Shafer a, entre autres, comme intérêt de décomposer chaque entité qui participe à un système de fusion en autant de jeux de masses élémentaires qu'il existe de sources d'information. Cette formulation est séduisante par sa simplicité mais aussi parce qu'elle permet de fournir un modèle cohérent et compatible avec des informations dont la nature est intrinsèquement différente. Ce cas se produit dès lors qu'une source d'information est accessible par un apprentissage et qu'une autre provient d'une connaissance humaine. Elle permet de la sorte de modéliser, d'une façon exacte et dans un cadre fédérateur commun, toute l'information disponible. En particulier aucune loi a priori au sens des probabilités n'est nécessaire dans ce modèle.

Cette théorie a été utilisée dans des cadres applicatifs différents et des études supplémentaires sont actuellement en cours à l'ONERA, notamment sur les critères de décisions et sur la fusion de classifieurs dont les cadres de discernement sont différents.

6 RÉFÉRENCES

[1] Smets P., Kennes R. "The transferable belief model", Artificial Intelligence, 66, 1994, pp 191-234.

[2] Smets P. "Bielief functions: The disjunctive rule of combination and the generalized bayesian theorem", Int. Jour. Approx. Reason., 9,1993, pp1-35.

[3] Dubois D., Prade H. "Fuzzy sets, probability and measurement", Eur. Jour. Operational Research, 40, 1989, pp135-154.

[4] Appriou A. "Probabilités et incertitude en fusion de données multi-senseurs", Revue Scientifique et Technique de la Défense, 11, 1991, pp 27-40.

[5] Denœux T. "A k-Nearest neighbor Classification rule base on Dempster-Shafer Theory", IEEE trans. SMC, 25, N°5, may 1995, pp 804-813.

[6] Caselton W.F., Luo W.,"Decision making with imprecise probabilites: Dempster Shafer Theory and application" Water Resources Reaserch, 28, N°12,1992, pp 3071-3081.

[7] Shafer G. "A mathematical theory of evidence", Princeton University Press,1976.