ESTIMATION DE L'INCERTAIN DANS LA CLASSIFICATION MULTISOURCE UTILISANT LA REGLE DE DEMPSTER

ZAHZAH E-H, DESACHY J, ZEHANA M.

Université Paul SABATIER , Institut de recherche en informatique de Toulouse, 118 route de Narbonne Toulouse 31062 France.
Tel (+33)(61-55-65-99) Fax (+33)(61-55-62-58)

Introduction

Un des plus grand problème auquel sont confrontés aujourd'hui les décideurs, est la gestion d'informations provenant de diverses sources inégalement importantes, fiables, imprécises, incertaines, etc...

Parmi les problèmes de gestion posés , celui de la combinaison d'informations est le plus crucial .

Par combinaison , on entend que selon le problème posé, un système doit pouvoir extraire une ou plusieurs informations, concernant une entité pouvant être suivant le niveau où l'on se place, une proposition, un événement, objet, etc..., ces informations, peuvent provenir de plusieurs sources, toutes ayant un avis à donner sur l'entité en question. Ces avis sont exprimés soit sous forme d'information directement mesurable (numérique) soit sous forme symbolique, en termes de présence/absence, vrai/faux etc...

Une application concernant notre problème particulier est celle de la représentation et la gestion d'un système d'informations géographiques, ce dernier contient pour une région donnée, une quantité spatiale d'information (images satellites multi-spectrales et multi-temporelles, cartes topographiques, géographiques, géologiques, météorologiques etc...).

Mise à part le problème de fiabilité, on remarquera que ce type d'information est très hétérogène du point de vue mesure d'information, les réponses des sources "satellites" sont physiques, et donc facilement mesurables, par contre les sources "géographiques" peuvent contenir des informations très complexes, relevant de la géographie humaine, économique, politique, allant jusqu'aux habitudes culturelles, la quantification de ces informations est plus complexe.

La quantité et la disparité d'information qu'on peut avoir à manipuler pour résoudre un problème donné est respectivement importante et complexe.

Il nous arrive souvent de faire une relation entre deux événements par de simples observations, mais que cette relation soit difficilement quantifiable. Par exemple:

La relation existant entre la réponse spectrale d'un objet, et son contexte topographique:

On sait que la réponse d'un objet terrestre, captée par un satellite, dépend de son contexte topographique, et que la réponse intrinsèque de l'objet constitue la partie la plus importante de la réponse donnée par le capteur du satellite, le contexte topographique, influe mais à un degré moindre que l'objet lui même, nous définissons donc ici, un ordre d'importance pour les sources.

Le problème que l'on se pose, est la caractérisation d'un objet en ayant à la fois des informations exprimées en unités différentes provenant de sources différentes, et une relation d'importance et de fiabilité de ces sources les unes par rapport aux autres.

Le problème présente une difficulté supplémentaire, quand ces informations sont données sous forme symbolique.

Les méthodes statistiques actuelles, ont montré leurs limites concernant la combinaison d'informations multisources , les principaux inconvénients sont:

1- impossibilité de manipuler des informations non-numériques , ce qui est gênant pour des problèmes utilisant des données symboliques,

2- Pour la plupart des méthodes, les données manipulées , même si elles sont numériques, doivent présenter une fonction de distribution identique, pour toutes les sources, ce qui n'est pas vrai dans tout ce qu'on rencontre dans la réalité,

3- les notions d'importance et de fiabilité, des sources les unes par rapport aux autres, même si on sait les quantifier, ne sont pas bien intégrées dans les méthodes existantes, permettant de les utiliser à bon escient .

De nouvelle théories ont fait leur apparition, se proposant de dépasser les handicaps posés par les méthodes statistiques, la théorie des fonctions de crédibilité proposée par G.Shafer en 1976, est destinée à ce type de traitement.

Les détails sur le développement de cette théorie sont donnés dans [Dubois88(1)] [Dubois88(2)] [Smets88] [Guan91].

Théorie des croyances, Application à la classification d'images multisource en télédétection.

La théorie mathématique des croyances appelée aussi théorie de l'évidence est un domaine dans lequel, les informations provenant de sources disparates, peuvent être combinées, pour produire un résultat d'inférence, pouvant aider à la prise de décision avec un certain degré de certitude, pour un problème quelconque du monde réel, où l'on est constamment confronté au problème d'utilisation de plusieurs sources d'information, provenant soit de jugements humains, des statistiques etc...[Sandri91]

La théorie de l'évidence permet de fusionner l'information en tenant compte des relations qui existent entre les sources, leur importance et fiabilité, et cela quelque soit l'unité dans laquelle est exprimée l'information constituant la source (numérique ou symbolique).

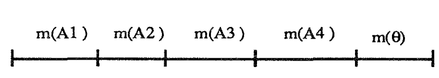
Ceci est d'un très grand intérêt pour les systèmes d'informations géographiques en général, et pour l'interprétation des images satellites en particulier car ces systèmes contiennent une quantité spatiale d'informations, stockée dans une base de donnée, qui necessite une mise à jour régulière.

Cette mise à jour consiste en la fusion

d'informations hétérogènes et incertaines, dans le but de résoudre un problème particulier.
Prenons par exemple, le cas de la classification de données satellites, où il s'agit de rassembler tous les objets qui ont des caractéristiques communes , afin d'étudier les comportements et évolutions de populations (estimations de récoltes, suivi de phénomènes naturels etc...).
L'utilisation des seules données satellite ne peut résoudre le problème car :
1- les données sont entachées de défauts, dûs à des effets topographiques et atmosphériques, cela affecte directement l'information mesurée, qu'il faudra corriger, ici c'est le contenu de l'information qui est mis en cause,
2- si l'on fait contribuer d'autres sources de données à la résolution du problème, il faudrait alors pouvoir quantifier leur contribution du point de vue importance et fiabilité, ici c'est la source elle même qui est mise en cause et non son contenu, qui peut être correct, mais la source peut ne rien à voir, ou ne contribue en rien à la résolution du problème.
C'est ce type de problème que nous allons aborder dans cet article, à l'aide de la théorie de l'évidence pour améliorer les résultats de la classification de données multisources .

Théorie des croyances

Supposons dans un premier temps que l'on traite une source unique de données, et en particulier, qu'on veuille examiner l'appartenance d'un pixel à une classe .
L'information dont on dispose est basée , sur un vecteur de données .
De ces données, on peut concevoir certains degrés d'appartenance de ce pixel aux classes, pour cela, on quantifie nos impressions dans la théorie de l'évidence en affectant une mesure de masse ou masse "d'évidence" à chaque proposition pour le pixel.
On introduit aussi un degré d'incertain permettant de préciser si ces masses sont fiables ou pas c'est-à dire que l'on soit pas totalement sûr, dans le fait que le pixel appartienne à l'une des classes au vu de la distribution des masses (degré de confiance par rapport aux possibilités d'appartenances des pixels) que l'on a choisie.
On peut noter que pour une source, la somme de toutes ces masses avec l'incertain constitue la totalité de l'information disponible , par conséquent les mesures numériques de masses indiquent les proportions de crédibilité.
En représentant la totalité de nos connaissances par une unité, la somme de toutes les possibilités (incluant le cas incertain) exprimées par leurs masses, est égale à 1 et toutes les masses seront comprises entre 0 et 1.



A_i : Classe
m: fonction de masse
$m(\theta)$: l'incertain

Représentation de l'incertain par les fonction de croyances [Dubois88]

Soit g(A), un nombre correspondant, à la réalisation de la proposition A, ce nombre prend ses valeurs dans l'intervalle [0,1], et répond aux conditions limites c'est à dire g(0)=0 et g(1)=1, où 1 représente la tautologie et 0 la contradiction, si g est une probabilité, alors
$g(A)=0 \rightarrow$ A est faux
$g(A)=1 \rightarrow$ A est vrai
Quand la connaissance de A est incertaine, la probabilité est alors exprimée par un intervalle [Bel(A), Pl(A)], où **Bel(A)** représente le degré de **certitude** de A et **Pl(A)** représente le degré de **plausibilité** ou de possibilité de A.
En faisant l'hypothèse que Bel(a)+ Pl(non A)=1 , Les situations suivantes sont les plus remarquables.

Bel(A)=0 et Pl(A)=1 i.e Bel(non A)=0 et Pl(non A)=1	aucune information sur A .	Pl(A)=1 Bel(A)=0
Bel(A)=1 ⇒ Pl(A)=1	A est totalement VRAI.	0 Pl(A) 1 Bel(A)
Pl(A)=0 ⇒ Bel(A)=0	A est totalement FAUX	Pl(A) 0 1 Bel(A)
Bel(A)=0.5 et Bel(non A)=0.5 ⇔ Bel(A)=0.5 et Pl(A)=0.5	A est aléatoire	Pl(A) 0 0.5 1 Bel(A)

Dans la théorie de l'évidence, la fonction de crédibilité Bel est aussi appelée fonction de masse et elle est notée m. On la définit de la manière suivante :

$$\forall A_i \in \Delta, 0 \le m(A_i) \le 1, \quad \sum_{A_i \in \Delta} m(A_i) = 1$$

$$\text{et} \quad m(\emptyset) = 0$$

Δ: représente l'ensemble de toutes les valeurs possibles
A_i: Sous ensemble d'un réferentiel de Δ
Le symbole $m(\theta)$ est utilisé pour représenter l'incertain, qui représente en fait le reste de l'information, pouvant être considéré comme une classe résiduelle.
Nous définissons maintenant les supports et les plausibilités de chaque proposition.
La fonction de croyance (support), pour une proposition A, est défini comme étant la somme de toutes les masses (crédibilités) affectées à A et ses sous-ensembles . Le support mesure à quel point, les informations données par une source, supportent A.
La plausibilité de A est définie comme étant un minimum des supports affectés à une proposition en contradiction avec la proposition A, elle mesure en fait à quel point, les informations données par une source, ne contradisent pas A. Ces mesures mettent en rapport un evenement A, avec l'evenement contraire non A, la fonction de croyance Bel et la fonction de plausibilité sont liées par la relation :
Bel(A)= 1- Pl(non A)
En termes algébriques, cela peut s'écrire

$$s(A) = \sum_{X \subset A} m(X) \quad \text{et} \quad p(A) = 1 - s(\bar{A})$$

Les supports peuvent être interprétés comme le minimum de crédibilité ou la plus petite valeur de

vraisemblance qu'un pixel appartienne à la classe A tandis que les plausibilités comme la possibilité maximum.
La "vraie" vraisemblance se trouve entre ces deux valeurs dans l'intervalle [s(A),pl(A)] qui est souvent appelé intervalle évidentiel.

Combinaison d'informations multi-sources (règle orthogonale de Dempster)
Considérons, maintenant, le cas où on dispose de plusieurs sources de données :
pour simplifier, nous n'allons prendre que deux sources avec les fonctions de masses suivantes:
$m1(<A,B,C,\theta_1>) = (m1(A),m1(B),m1(C),m1(\theta))$
$m2(<A,B,C,\theta_2>) = (m2(A),m2(B),m2(C),m2(\theta))$
La question, maintenant, est de savoir comment nous pouvons combiner ces deux sources pour calculer le degré d'appartenance de ce pixel aux différentes classes, et de réduire l'incertain de telle manière que l'intervalle évidentiel devienne suffisamment petit, pour qu'on puisse choisir la classe appropriée avec un degré de confiance élevé .
Le moyen de combiner ces deux sources se fait dans la théorie de l'évidence par la règle orthogonale ou règle de combinaison de Dempster .
Une première source produit des informations qui seront représentées par le vecteur de masse m1, où chaque composante indique le degré de confiance qu'a la source concernant l'appartenance de l'objet à la classe correspondante.
La seconde source , produit quant à elle des informations qui seront contenues dans le vecteur m2, et le vecteur m sera le résultat de la combinaison des deux vecteurs précédants ($m1 \oplus m2$), Pour un sous-ensemble X, m(x) est donné par la somme du produit de toutes les valeurs affectées aux sous ensembles Y et Z , appartenant respectivement à la première et seconde source, et qui par leur intersection produisent le sous ensemble X :

$$m(X) = K \sum_{Y \cap Z = X} m1(Y).m2(Z)$$

$$K^{-1} = 1 - \sum_{Y \cap Z = \varnothing} m1(Y).m2(Z)$$

K est appelé constante de normalisation de la somme orthogonale de ($m1 \oplus m2$), il mesure le conflit entre les deux fonctions de masses.
La quantité log(K) notée wett(m1,m2) par [Guan91] est appelée le degré de conflit entre m1 et m2.
S'il n'y a pas de conflit entre m1 et m2, alors wett(m1,m2)=0, si les deux sources sont en conflit total i.e (la somme orthogonale n'existe pas), alors $wett(m1,m2)=\infty$.
Pour avoir $m(\Delta)=0$, on mutiplie cette somme par le facteur K qui est calculé à partir du degré de conflit entre les deux sources et qui est donné par

$$1 - \sum_{Y \cap Z = \varnothing} m1(Y).m2(Z)$$

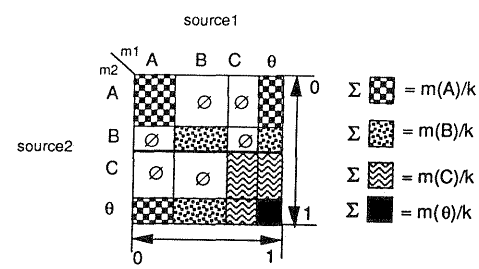
$$\text{si } \sum_{Y \cap Z = \varnothing} m1(Y).m2(Z) = 1$$

alors les deux sources sont
totalement inconsistantes (Aucune étiquette en commun).

L'application de la règle de Dempster, pour la classification d'images satellites, se simplifie, si l'on suppose que les classes sont exclusives, les propositions sont donc considérées comme des classes ou sous ensembles disjoints, (leur intersection est nulle):
Si la source S1 produit $m1(<A,B,C,\theta>)$, et la source S2 produit $m2(<A,B,C,\theta>)$,
et m le vecteur résultat de la combinaison $m(<A,B,C,\theta>)$ alors

$$m(x) = K.(m1(x).m2(x) + m1(x).m2(\theta) + m2(x).m1(\theta))$$

$$K = \frac{1}{1 - \sum_{y \cap z = \varnothing} m1(y).m2(z)}$$



Estimation de l'incertain :
Une condition préalable pour appliquer la règle de Dempster est de déterminer quel degré d'incertain, on va rattacher aux données de chaque source?.
Lee[Lee87] a proposé deux méthodes pour calculer l'incertain de chaque source:
Un choix possible est de réaliser une classification pour chaque source afin d'en estimer le pourcentage d'erreurs et les utiliser comme valeur de l'incertain. Cet incertain serait dû à la modélisation des classes ou à la confusion entre les classes.
Une autre méthode , consiste à considérer les vecteurs de mesures et d'estimer les erreurs engendrées lors du calcul des paramètres statistiques ayant servi pour définir les classes.

L'estimation de l'incertain en fonction de la séparabilité des classes
L'estimation de l'incertain qu'on considére comme étant le reste des cas pour lesquels on ne dispose d'aucune information, et que l'on associe à une classe résiduelle, peut aussi être calculé par une technique basée sur le calcul de la séparabilité des classes, ou bien le degré de confusion inter-classes.
On considère ici que l'incertain, ou le degré de vraisemblance affecté à la classe résiduelle est nulle, si la source affecte la totalité de l'information à une et une seule classe, cest à dire que toutes les autres classes possédent un degré de vraisemblance minimum . Dans ce cas la séparabilité entre la classe la plus favorable et le reste des classes est maximum (i.e confusion nulle).
Si on considére que l'on dispose de n classes et que les mesures de vraisemblance sont des probabilités, alors ce cas serait représenté par:

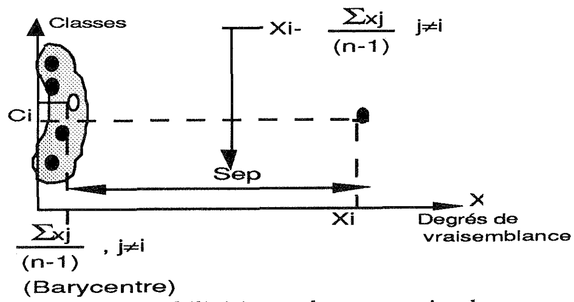
$$P(C_i) = 1 \text{ et on a } P(C_j) = 0 \quad \forall j \neq i$$

L'incertain, présente par contre un degré maximun, quand la totolité de l'information est distribuée équitablement, entre les différentes classes, dans ce cas les informations provenant de la source

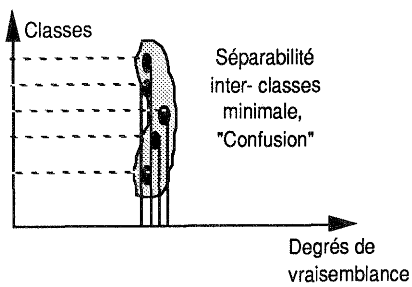
corréspondante sont sans effet et ne contribuent en rien à la prise de décision.
Si on considére que l'on dispose de n classes et que les mesures de vraisemblance sont des probabilités, alors ce cas serait représenté par:

$$P(x/Ci) = 1/n \qquad \forall i.$$

Ces deux cas de figures peuvent être shématisés par :



Séparabilité inter-classes maximale
"Incertain minimal"



"Incertain maximal"

A partir de ces valeurs extrêmes, les valeurs intermédiaires de l'incertain, seront calculées pour des situations intermédiaires corréspondantes. On rappelle que les définitions données plus haut, ne sont valables que si les classes sont mutuellement exclusives.
Le problème de séparabilité ou de concentration des mesures de croyances, a été soulevé par Sandri[Sandri91] sous le titre d'equi-répartition des croyances.
Sandri fait remarquer dans sa thèse la différence qui existe entre l'equi-répartition des croyances et celle de l'ignorance totale, dont voici les définitions,
soit Ω, un référentiel,
A: un sous ensemble du référentiel Ω
g(A): un nombre réel, permettant de quantifier la confiance qu'a la source en l'occurrence de l'événement A.

L'ignorance
l'état d'ignorance d'une source par rapport à un événement, est caractérisé par la seule croyance que Ω est certain et $\varnothing$ est impossible, et l'absence de tout autre événement. Tout ce que l'on sait, est que la croyance est entre 0 et 1. L'ignorance se caractérise par la certitude que tout autre événement différent de Ω est égale à 0. Le raisonnement avec cette hypothèse est appelé par certains auteurs "raisonnement avec les hypothèses du monde fermé" [Smets88] .
Dans la théorie des croyances, l'ignorance est modélisé par :
$Bel(\Omega) = 1$ et $Bel(A) = 0 \; \forall A \neq \Omega$.

L'equi-répartition des croyances
Cet état est caractérisé par une croyance uniforme de la source en tous les éléments du référentiel :
$g(\{\omega\}) = k \; \forall \omega \in \Omega$
l'equi-répartition des croyances exprime l'absence d'informativité d'une source, cet état inclut l'état d'ignorance. La prise de décision est pratiquement impossible, concernant la véracité d'un événement.
En termes de probabilités, l'equi-répartition des croyances est réalisée par une distribution de probabilité, telle que $p(\omega)= 1/ |\Omega| \;\; \forall \omega \in \Omega$.
En théorie des croyances, l'equi-répartition est représentée par:
$\forall \omega \in \Omega$, $Pl(\{\omega\})=k$, $k=1/|\Omega|$
La remarque, qui peut être faite ici, est que la théorie des croyances, permet de représenter et distinguer entre les deux états (ignorance, et equi-répartition des croyances), tandis que la théorie des probabilités, ne peut représenter que l'equi-répartition des croyances.
Afin d'appliquer la règle de Dempster, l'incertain peut être calculé en se basant sur ces définitions, où au départ, il s'agira de calculer les degrés de réalisation des différentes classes, par une méthode classique probabiliste ($Pr \in [0,1]$) , géométrique ($Dist \in R^+$) ou bien par une autre méthode quelconque donnant par exemple des facteurs de certitudes . ($CF \in [-1 +1]$ dans le système expert MYCIN).
Si on suppose que les mesures de vraisemblances sont représentées par des probabilités, l'estimation de l'incertain se réalise de la manière suivante:
Pour un point, les probabilités d'appartenir aux n classes <x1,x2,....,Xn> sont

$$<p1,p2,...,pn>, \; 0 \le pj \le 1 \;\; \forall j=1,n$$

On calcule la séparabilité par la formule :

$$Sep = p_i - \frac{1}{(n-1)} \sum_{\substack{j=1 \\ j \neq i}}^{n} p_j$$

$p_i = Max \; p_j$, $j=1,n$
l'incertain $m(\theta)$ est donnée par $m(\theta) = 1\text{-}Sep$.
Si la séparabilité est maximale c'est à dire égale à 1, alors l'incertain est nul, par contre si la séparabilité est nulle, alors l'incertain est maximal, c'est à dire égal à 1.
Le vecteur de masse X , aura une composante supplémentaire qui correspondra au cas incertain ou bien à la classe résiduelle, pour laquelle on affecte le reste de l'information.

$$m < x_1, x_2, \ldots, x_n, \theta > \quad \text{avec} \quad m(x_i), m(\theta) \in [0,1]$$

on peut facilement construire à partir du vecteur X, le nouveau vecteur de masse m'(X) tel que

$$m'(X) = m'(< x_1, x_2, \ldots, x_n, \theta >) \quad \text{avec} \quad \sum_i m'(x_i) + m'(\theta) = 1$$

A partir de ce vecteur, on calcule les quantités correspondant à la crédibilité et à la plausibilité, pour ainsi pouvoir prendre une décision, pour l'affectation du point à la classe la plus proche.
On peut aussi introduire la fiabilité des sources dans la même formule calculant l'incertain.
On suppose qu'on soit dans le cas où on a une idée sur l'ordre d'importance des sources dans le

processus de classification. Pour exploiter cette information supplémentaire on a trois choix possibles:

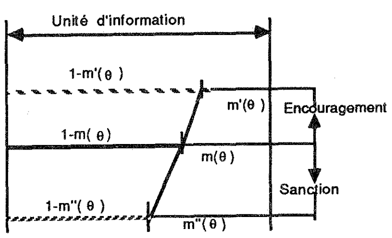
1- encourager les sources les plus importantes, par rapport à celle qui l'est le moins, en conservant l'information globale de la source la moins importante,

2- sanctionner les sources les moins importantes, par rapport à la source la plus importante,. en conservant l'information globale de la source la plus importante,

3- encourager les sources les plus importantes, et sanctionner les moins importantes.

L'encouragement, et la sanction des sources, se font par la réduction ou par l'étalement de la valeur déjà affectée à la partie associée à la classe résiduelle (Incertain).

En étalant cette valeur, on diminue de ce fait, l'information corréspondant aux autres classes. La réduction de cette valeur aura pour effet d'augmenter l'information corréspondant aux autres classes.



On remarque, qu'avec ces hypothèses particulières, si les rapports d'importance sont identiques, l'encouragment (resp la sanction) des deux sources en même temps n'a pas d'effet, car les données pour la prise de décision ne changent pas , elles ne font que se translater.

Si on suppose que deux sources différentes S1, S2 produisent des informations X et Y(degrés de vraisemblance aux différentes classes) concernant un même objet,

S1 produit :

$m1(X)= m1(<x_1, x_2, ..., x_n>) = m1(x), m1(x),....m1(x)$

S2 produit :

$m2(X)=m2(<x_1, x_2, ..., x_n>)= m2(x1),m2(x2),...m2(xn)$

Si on suppose ensuite, que l'on dispose d'une information supplémentaire, nous donnant le rapport d'importance des sources r1/r2 avec r1, r2 nombres positifs quelconques.

On transforme l'intervalle [Min (r1,r2), Max(r1, r2)] dans l'intervalle [k1, k2] tel que:

[k1,k2] = [1, Max(r1,r2)/Min (r1,r2)].

Ce changement d'echelle, est en fait réalisé afin que les rapports d'importance, puissent être intégrés correctement dans le vecteurs de masse de chacune des sources, ainsi que dans le processus de combinaison.

On calcule la quantité correspondant à l'incertain de chaque source par les formules suivantes:

$$m1(\theta) = \begin{cases} \frac{1}{k_2}[1 - Sep1] \text{ si } & r1 > r2 \\ 1 - Sep1 & \text{Sinon} \end{cases}$$

$$m2(\theta) = \begin{cases} \frac{1}{k_2}[1 - Sep2] \text{ si } & r2 > r1 \\ 1 - Sep2 & \text{Sinon} \end{cases}$$

Si r1>r2, alors la source S1, verra la quantité affectée à l'incertain (classe résiduelle) diminuer, et de là sa contribution dans la combinaison d'informations pour les autres classes, sera plus importante, tandis que la quantité affectée à l'incertain (classe résiduelle) de la source S2 gardera la même valeur, et sa contribution restera la même.

Si r2 est supérieur à r1, alors c'est la deuxième source S2 qui contribuera le plus, par rapport à la première source qui gardera sa contribution initiale.

Cette pratique n'est valable que si l'on sait quantifier nos impressions concernant l'importance des sources les unes par rapport aux autres.

On rappelle qu'ici on ne procéde pas à la sanction de la source qui est la moins importante, cette dernière conservera sa contribution, et c'est la source qui est la plus importante qui verra sa contribution augmentée.

La stratégie concernant la conservation de l'information, provenant de la source la plus importante, et la sanction de la source la moins importante, ne peut pas être résolue de la même manière, car si on suppose que les rapports d'importance sont exprimés dans un référentiel quelconque, le contrôle de la nouvelle valeur de l'incertain m''(θ) ne serait plus maîtrisé.

Le problème est en fait de trouver k en fonction des seuls rapports d'importance r1 et r2 tel que:

$m''(\theta)= k. m(\theta)$, avec $0 \leq m''(\theta) \leq 1$ et

$m''(\theta) > m(\theta)$

la résolution de ce systéme donne k en fonction de m(θ), (k>1 et k< 1/m(θ)). Cette méthode est donc inéfficace pour ce problème particulier.

Application

Prétraitements

Nous avons procedé pour l'extraction des mesures de vraisemblances d'un point de l'image aux différentes classes, par calcul de la distance du point aux centres de gravité des différentes classes par la formule:

$$dist(p_j, C_i) = \left| \frac{p_j - \mu_{i,j}}{\sigma_{i,j}} \right|$$

pj: représente la valeur du point dans la source j

$\mu_{i,j}$: représente la moyenne de la classe i pour la source j,

$\sigma^2_{i,j}$: représente la variance de la classe i pour la source j,

Ces mesures nous renseignent sur l'appartenance du point à une classe dans une source donnée .

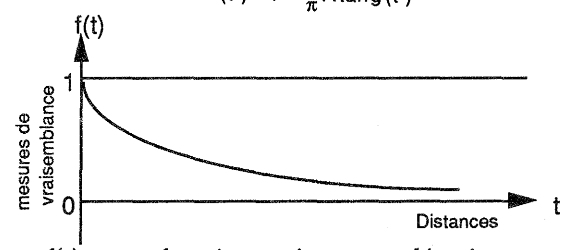
Si la distance se confond avec le centre de la classe i.e distance nulle, alors le point appartient à cette classe, avec un degré maximum, si la distance est par contre importante, le degré d'appartenance de ce point à la classe est minimum.

Ces mesures seront faites pour tous les points de l'image et pour chaque source.

L'utilisation de la règle de Dempster, nécessite un prétraitement de données, qui permettra à toutes ces

mesures d'être dans l'intervalle [0,1], et que la somme de la totalité de l'information y compris le cas incertain d'être égale à 1.
Les mesures auxquelles on a procédé sont des distances et donc définies dans R^+, la distance nulle correspond à un vraisemblance maximale , et une distance→infinie, correspond à une vraisemblance minimum.
La transformation (permettant le passage des distances à des mesures de vraisemblance donnant une valeur maximale 1.0 pour une distance minimale 0.0, et une valeur de vraisemblance minimale 0.0 pour une distance maximale infinie), est réalisée grâce à la fonction suivante par exemple.

$$s = f(t) = 1 - \frac{2}{\pi}\mathrm{Atang}(t)$$



s = f(t): est fonction strictement décroissante et définie de [0, ∞ → [0,1]
A partir de ces valeurs, on procéde au calcul de l'incertain, en fonction de la séparabilité ou de la confusion des classes, en appliquant la formule:

$$m(\theta) = 1 - \mathrm{Sep} = 1 - \left(s_{max} - \frac{1}{(n-1)} \cdot \sum_{\substack{i=1 \\ i \neq max}}^{n} s_i \right)$$

Dans une troisième étape, on recalcule l'ensemble des valeurs de vraisemblance pour chaque classe x_i, y compris le cas incertain θ, de façon à ce que la totalité de l'information soit réprésentée dans l'unité

$$m'(x_i) = \frac{s_i}{\sum_{j=1}^{n} s_j + m(\theta)}$$

$$m'(\theta) = \frac{m(\theta)}{\sum_{j=1}^{n} s_j + m(\theta)}$$

Les $m(x_i)$ ainsi que $m'(\theta)$, seront les nouvelles valeurs correspondant aux mesures qui seront, utilisées lors des traitements qui suivront, à savoir la combinaison multisource par la règle de Dempster, calcul des crédibilités, plausibilités et intervalle évidentiel du vecteur de masse résultat, pour la prise de décision finale, concernant l'appartenance du point à l'une des classes définies au départ.

Résultats

Nous disposons d'images satellites, considérées ici comme sources d'informations, ces images sont prises dans des bandes spectrales différentes, chaque bande possède ses propres caractéristiques, la réponse des objets de l'image est différente selon la bande spectrale utilisée.

Pour la classification, la combinaison de ces sources d'informations est nécessaire, si l'on veut faire ressortir toutes les classes, car une source peut contenir beaucoup d'informations concernant une classe, et peu concernant une deuxième classe, (informations pertinentes liées à une source donnée par exemple)
Nous avons procédé à une classification d'une région dont nous disposons des echantillons sûrs pour neuf classes. La région représente une région a fort relief , ce qui rend la discrimination des classes encore plus difficile. On a utilisé les deux premières composantes du résultat de l'analyse en composante principale (KL1,KL2) effectuée sur les quatre bandes spectrales Mss, qui compresse et décorréle l'information, plus de 90% de l'information du contraste total est contenue dans ces deux composantes.
La classification par la règle de Dempster utilisant ces deux sources spectrales avec le même degré de fiabilité, a permis de faire ressortir les résultats donnés par la table(1) qui indique les pourcentages des pixels bien classés pour les différentes classes, (ce pourcentage est calculé relativement aux échantillons qui ont servi à la caractérisation de la classe.), ainsi que la moyenne de la classification globale, effectuée toujours sur les échantillons.

	Cl1	Cl2	Cl3	Cl4	Cl5	Cl6	Cl7	Cl8	Cl9	MOY
%	27.21	29.17	38.54	21.41	53.39	38.28	61.33	81.25	30.73	42.37

Table 1.
sources d'importance égale
KL1=1, KL2=1

Dans une seconde étape, nous sommes intéressé au problème de l'intégration des informations topographiques dans le même processus de classification, vu que ces informations sont mal intégrées par les méthodes statistiques [Benediktsson90], qui ne manipulent, ou ne combinent que des données ayant une même distribution.
On sait par ailleurs que le contexte topographique, peut contribuer à la discrimination des classes, mais on ignore avec quel degré cela se fait, aussi l'information topographique doit être considérée comme moins importante (moins fiable dans le processus de classification) que l'information provenant des sources radiométriques, qui est en fait plus caractéristique.

L'information topographique, peut contribuer à améliorer le résultat de la classification, mais n'est pas discriminante, dire qu'une classe se trouve toujours à une certaine altitude, n'est pas toujours vrai.
La règle de Dempster, combinant données spectrales et topographiques, a permis d'obtenir les résultats suivants:

	Cl1	Cl2	Cl3	Cl4	Cl5	Cl6	Cl7	Cl8	Cl9	MOY
%	67.84	60.81	65.23	50.31	71.61	78.52	70.57	84.11	26.43	63.94

Table2
Rapport d'importance des sources (KL1, KL2, MNT) = (1,1,1)

On remarque qu'en intégrant cette donnée supplémentaire le résultat s'est nettement amélioré (de 20% environ), cette amélioration s'explique peut être par le fait que l'image représente une zone montagneuse à relief très prononcé, et que les réponses spectrales des classes sont fortement

influencées par la donnée hauteur.
En donnant maintenant, un degré de fiabilité moins important à la source représentant l'information des hauteurs, le résultat obtenu, est :

	Cl1	Cl2	Cl3	Cl4	Cl5	Cl6	Cl7	Cl8	Cl9	MOY
%	68.10	56.12	60.68	48.75	75.26	70.57	78.65	82.29	27.60	63.11

Rapport d'importance des sources (KL1, KL2, MNT) = (2,2,1)

Ici on atteint à un état de tarissement des sources spectrales, la source représentant l'image du MNT, contribue toujours de la même manière, le fait d'augmenter les degrés de fiabilité des sources spectrales, ne sanctionne pas pour autant l'information provenant du MNT, si l'opération devient idempotente, c'est qu'on ne doit plus rien attendre des sources spectrales, elles sont taries.
Enfin, nous avons procédé à une classification en donnant un rapport d'importance entre les sources spectrales et la sources topographique de l'ordre de 1/10 l'image résultat montre l'influence totale de la source MNT, dans le processus global de la classification. Des classifications succéssives, avec des degrés de fiabilité plus importants donnés à la source MNT, à montré que l'opération converge vers un solution stable, puisqu'elle devient là aussi idempotente.

Conclusion
Nous avons abordé, le problème de la classification des images satellite , en utilisant les techniques de la théorie des croyances, et principalement la règle de combinaison de Dempster, ceci nous a permis de surpasser certains obstacles qui étaient posés par les méthodes statistiques classiques.
Les résultats obtenus ont été améliorés, par le fait d'avoir pu intègrer aisément d'autres sources de données hétérogènes dans un même processus. Le résultat global, a été amélioré de plus de 20%, par rapport à l'utilisation des seules données spectrales par la même technique.
Nous avons estimé l'incertain en fonction des mesures fournies par les sources elles mêmes, et de leurs répartitions (séparabilité).
Nous avons supposé, qu'il y a absence d'informativité, quand les croyances sont réparties équitablement, et une informativité maximum correspondant à une séparabilité maximum des croyances.
L'intégration de l'information d'une seconde source hétérogène provenant du MNT dans le processus de classification, est réalisée à partir des mesures effectuées sur les échantillons, cela suppose que des classes différentes se trouvent toujours à priori dans des contextes topographiques différents, ce qui peut être inexact pour certaines régions.

Actuellement, nous pensons aborder le même problème, en considérant chaque classe, par les mesures numériques éffectuées sur les echantillons qui la répresentent, ainsi que par des connaissances symboliques sous forme de règles de production, définissant son contexte général et sa fréquence relative d'apparition.

Références

Barnett J A "calculating Dempster-shafer plausibility " IEEE transactions on pattern analysis and machine intelligence, Vol 13, n° 6, june 1991.

Dubois D, Prade H "Représentation and combination of uncertainty with belief functions and possibility measures" Computational intelligence, Vol 4 n° 4 pp224-264 1988.

Dubois D, Prade H, Sandri S, Testemale C. "Techniques de combinaison d'informations incertaines: comparaison et application à un système d'interrogation multi-source" Rapport interne n°312, LSI Decembre 1988.

Garvey T D, Lawrence J D and Fishler M A " An inference technique for integrating knowledge from disparate sources" VIII th IJCAI Vancouver August 1981.

Guan J, Bell D.A "Evidence theory and its application" Vol I North holland 1991.

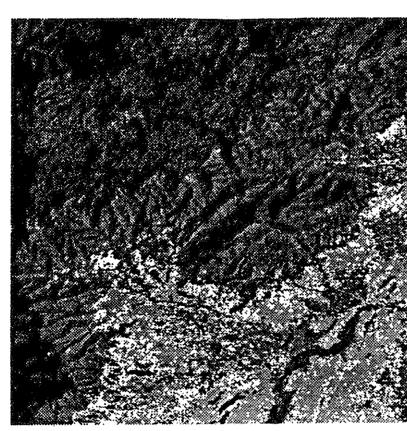
Lee T, Richards J A, Swain P.H "Probabilistic and evidential approaches for multisource data analysis", IEEE Transactions on geoscience and remote sensing, Vol GE-25 N° 3, May 1987.

Sandri S "La combinaison de l'information incertaine et ses aspects algorithmiques", Thèse de doctorat de l'université Paul Sabatier Toulouse III, Decembre 1991.

Shafer G, " probability judgement in artificial intelligent" Uncertainty in artificial intelligence L.N Kanal and J.F Lemmer (Editors) Elsevier Science Publishers B.V (north holland) 1986.pp 127-135.

Shafer G, Logan R, "Implementing Dempster's rule fore hierarchical evidence" Artificial intelligence 33(1987) 271-298.

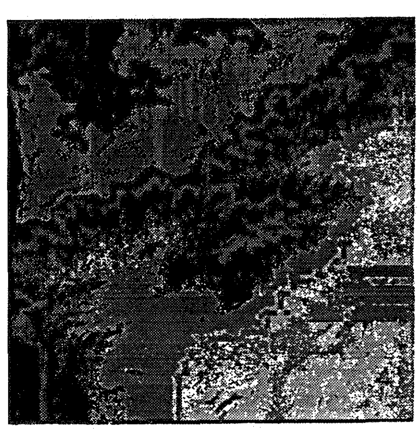
Smets P, "The combination of evidence in the transferable belief model", IEEE transactions on PAMI, Vol 12, N°5, May 1990



KL1=1 KL2=1 MOY=42.37%

KL1=1	KL2=1	MNT=1	MOY=63.94%

KL1=2	KL2=2	MNT=1	MOY=63.11%



KL1=1	KL2=1	MNT=10