Utilisation de la théorie de l'évidence pour la restauration de chaînes de Markov cachées multispectrales

Nathalie Giordana et Wojciech Pieczynski

Département Signal et Image
Institut National des Télécommunications
9, rue Charles Fourier
91011 EVRY cedex France

RÉSUMÉ

Nous présentons dans cet article certaines possibilités d'utilisation de la théorie de l'évidence de Dempster-Shafer dans le contexte de la restauration des chaînes de Markov cachées. Nous introduisons une nouvelle notion de chaîne évidentielle de Markov cachée et montrons, par une étude des simulations, la pertinence de son utilisation dans certaines situations.

ABSTRACT

We present in this paper the use of theory of evidence by Dempster Shafer in the framework of restauration of hidden Markov chains. We propose a new evidential hidden Markov chain model and we show on simulations results the interst of this approach in particular cases.

1 Introduction

La théorie de l'évidence [6], suscite un interêt grandissant dans le domaine de la fusion de données. Citons notamment les travaux de Bloch [3], Nimier [5] et Appriou [1]. L'objectif de notre travail est de présenter certaines possibilités d'utilisation de la fusion de Dempster-Shafer dans le contexte de la restauration des chaînes de Markov cachées. Nous considérons pour cela un système d'acquisition de données multisenseurs, nous supposons que nous avons accès dans chaque senseur, à une version bruitée du signal modélisée par une chaîne de Markov. Le problème général est celui de l'estimation de la réalisation inobservable de la chaîne à partir de toutes les sources d'informations disponibles. Nous introduisons une nouvelle notion de chaîne évidentielle de Markov cachée et montrons, par une étude des simulations, la pertinence de son utilisation dans certaines situations.

2 Les chaînes de Markov cachées

Nous précisons dans cette section le modèle de chaînes de Markov cachées et rappelons le principe et le déroulement de la méthode de restauration bayesienne MPM.

2.1 Le modèle

Nous considérons une suite de variables aléatoires $\{X_n,\ n \in S\}$, $card(S) = N$, à valeurs dans $\Omega = \{\omega_1, \ldots, \omega_k\}$ espace des classes. Nous ne disposons pas directement des observations des réalisations de $X = (X_1, \ldots, X_N)$ mais d'une suite de variables aléatoires $Y = (Y_n)_{n\in S}$, chaque Y_i étant à valeurs dans $\mathbb{R}^d$, dont les réalisations sont les observations obtenues par d senseurs. Le problème est d'estimer l'état $X = x$ de la chaîne au vu des observations $(Y_1, \ldots, Y_N)$. La loi *a priori* de X notée P_X et les lois conditionnelles de Y sachant $X = x$ notées $P_{Y|X=x}$ sont supposées connues. Nous pouvons alors obtenir sous les hypothèses (H1) et (H2) ci-dessous, la loi du couple et la loi *a posteriori* de X sachant $Y = y$.

Plus précisement, $(X_n)_{n\in S}$ est une chaîne de Markov cachée stationnaire d'ordre 1 à valeurs dans un espace $\Omega = \{\omega_1, \ldots, \omega_k\}$. Sa loi *a priori* est déterminée par

$$c_{ij} = P(X_n = \omega_i, X_{n+1} = \omega_j), \quad (\omega_i, \omega_j) \in \Omega^2 \tag{1}$$

dites probabilités conjointes. Ces probabilités définissent la loi initiale et les probabilités de transition :

$$\pi_i = P(X_1 = \omega_i) = \sum_{j=1}^{k} c_{ij}, \quad \omega_i \in \Omega \tag{2}$$

$$\begin{aligned} a_{ij} &= P(X_{n+1} = \omega_j \mid X_n = \omega_i) \\ &= \frac{c_{ij}}{\sum_{j=1}^{k} c_{ij}}, \quad (\omega_i, \omega_j) \in \Omega^2 \end{aligned}$$

Nous émettons les hypothèses suivantes :

(H1) Les variables aléatoires $(Y_n)_{n\in S}$ sont indépendantes conditionnellement à X.

(H2) La loi conditionnelle de Y_n sachant X est égale à la loi conditionnelle de Y_n sachant X_n. On note f_x la densité de probabilité de la loi de Y conditionnelle à $X = x$. Nous pouvons écrire :

$$\begin{aligned} P(Y = y \mid X = x) &= \prod_{t=1}^{N} P(Y_t = y_t \mid X_t = x_t) \\ &= \prod_{t=1}^{N} f_{x_t}(y_t) \end{aligned}$$

La loi du couple (X, Y) est alors définie par la densité $h(x, y) = P_X(x) \times P_{Y/X=x}(y)$ qui est une densité par rapport au produit de la mesure de dénombrement sur Ω^N par la mesure de Lebesgues sur $\mathbb{R}^{dN}$.
De plus nous supposons que les observations conditionnellement à une classe donnée sont spectralement indépendantes :

(H3) les variables aléatoires $Y_n^1, \ldots, Y_n^d$ sont indépendantes conditionnellement à X_n.

Soit f_i la densité de probabilité sur $\mathbb{R}^d$ de la loi de Y_n conditionnelle à $X_n = \omega_i$. Alors f_i est donnée par d densités sur $\mathbb{R}$, qui sont les lois de Y_n^j conditionnelles à $X_n = \omega_i$ et que l'on notera f_i^j.

$$f_i(y_n^1, \ldots, y_n^d) = \prod_{j=1}^{d} f_i^j(y_n^j) \tag{3}$$

2.2 La restauration

La méthode de restauration de chaînes de Markov cachées présentée dans ce paragraphe est la méthode *Maximum du Mode a Posteriori* (MPM). Elle consiste à chercher pour chaque variable aléatoire X_n, l'état ω_j qui maximise la probabilité *a posteriori* marginale :

$$X_n = \omega_j \Leftrightarrow P(X_n = \omega_j \mid Y_1 = y_1 \ldots Y_N = y_N) = \max_{i\in\{1\ldots k\}} P(X_n = \omega_i \mid Y_1 = y_1 \ldots Y_N = y_N)$$

La probabilité *a posteriori* de chaque X_n est évaluée par l'algorithme *Forward Backward* de Baum [2].
Soit F et B les probabilités "Forward" et "Backward" :

$$F_n(\omega_i) = P(X_n = \omega_i \mid Y_1 = y_1 \ldots Y_n = y_n)$$
$$B_n(\omega_i) = P(Y_n = y_n \ldots Y_N = y_N \mid X_n = \omega_i)$$

$$F_n(\omega_j) = N_n \pi_j f_j(y_n) \quad \text{pour} \quad n = 1 \tag{4}$$
$$F_n(\omega_j) = N_n \left(\sum_{i=1}^{k} F_{n-1}(\omega_i) a_{ij}\right) f_j(y_n) \quad \text{pour} \quad n > 1$$
$$B_n(\omega_i) = 1 \quad \text{pour} \quad n = N \tag{5}$$
$$B_n(\omega_i) = N_{n+1} \sum_{j=1}^{k} a_{ij} f_j(y_{n+1}) B_{n+1}(\omega_j) \quad \text{pour} \quad n < N.$$

N_n est un coefficient de normalisation :

$$N_n = \left(\sum_{j=1}^{k} F_n(\omega_j)\right)^{-1}$$

3 Les chaînes de Markov cachées évidentielles

Nous présentons dans cette section des éléments de la théorie de l'évidence, puis nous explicitons la notion de chaînes de Markov cachées évidentielles.

3.1 Quelques éléments de la théorie de l'évidence

Notons $E = \{C_1, \ldots, C_k\}$ l'ensemble des classes encore appelé cadre de discernement, et $E^* = \{A_1, \ldots, A_{2^k}\}$, où A_j sont les sous ensembles de E.

Définition 3.1 — *On appelle fonction de masse une application*

$$m: \begin{array}{rcl} E^* & \longmapsto & [0, 1] \\ A_j & \longmapsto & m(A_j) \end{array} \tag{6}$$

vérifiant :
$m(\emptyset) = 0$ *et* $\sum_{j=1}^{2^k} m(A_j) = 1, \quad m(A_j) \geqslant 0$

Définition 3.2 — *On appelle fonction de plausibilité notée Pl la quantité*

$$Pl(B_k) = \sum_{A_j \cap B_k \neq \emptyset} m(A_j) \tag{7}$$

Les éléments focaux sont les éléments de E^* de masse non nulle. Lorsque les éléments focaux sont réduits aux seuls singletons C_i du cadre de discernement E, les notions de masse élémentaire et de plausibilité sont assimilables à celle de probabilité :

$$\forall A_j \neq C_i \quad i \in \{1, \ldots, k\} \quad m(A_j) = 0 \quad \textit{et}$$
$$Pl(A_j) = m(A_j) = P(A_j)$$

Dans ce cas, la masse sera dite "bayesienne".

Dans le cas de d sources d'informations, nous associons à chaque source i une fonction de masse M_i. La fusion des informations consiste alors à combiner les d fonctions de masses. La règle générale de combinaison de d sources distinctes par sommation orthogonale est la suivante :

$$M(.) = M_1(.) \oplus M_2(.) \oplus \cdots \oplus M_d(.) \quad \text{où}$$

$$M(\emptyset) = 0 \quad \text{et} \quad M(A) = (1-K)^{-1} \sum_{A_1 \cap \cdots \cap A_d = A} \left(\prod_{j=1}^{d} M_j(A_j)\right) \tag{8}$$

$$\text{avec } K = \sum_{A_1 \cap \cdots \cap A_d = \emptyset} \left(\prod_{j=1}^{d} M_j(A_j)\right)$$

Propriété 1 *La combinaison d'une fonction de masse M quelconque avec une fonction de masse bayesienne M_b donne une fonction de masse bayesienne M_b' tel que :*

$$M_b'(C_i) = \frac{M_b(C_i) Pl(C_i)}{\sum_{l=1}^{k} M_b(C_l) Pl(C_l)} \qquad C_i \in E \tag{9}$$

où Pl et la fonction de plausibilité associée au jeu de masse M.

3.2 Chaîne de Markov cachée évidentielle

Considérons $\Omega^* = \{\Omega_1, \ldots, \Omega_{2^k}\}$ où les ensembles Ω_i, $i = 1, \ldots, 2^k$, sont des sous-ensembles de l'ensemble des classes $\Omega = \{\omega_1, \ldots, \omega_k\}$.

Une fonction de masse étant une probabilité, nous pouvons considérer une suite de variables aléatoires $(X_n^*)_{n\in S}$, chaque X_n^* prenant ses valeurs dans Ω^*, dont la loi est une probabilité sur Ω^{*^N} avec $card(S) = N$.

Définition 3.3 **—** *Une fonction de plausibilité définie sur Ω^N sera dite "loi d'une chaîne de Markov évidentielle $(X_n)_{n\in S}$", si la fonction de masse qu'elle définit sur Ω^{*^N} est une loi d'une chaîne de Markov $(X_n^*)_{n\in S}$.*

Notons $A^* = [a_{ij}^*]$, $i, j \in \{1, \dots, 2^k\}^2$ la matrice de transition et π_i^*, $i \in \{1, \dots, 2^k\}$ la probabilité initiale de la chaîne. La loi de X^* est notée M^0.
Soit $(Y_n)_{n\in S}$ l'observation issue des d canaux d'observations. Chaque $Y_n \in \mathbb{R}^d$, $Y_n = (Y_n^1, \dots, Y_n^d)$. Les fonctions de masses associées aux observations sont notées M^i, $i = 1, \dots, d$. Nous nous plaçons dans le cadre des hypothèses (H1)-(H3).
Ainsi la loi de Y conditionnelle à X^* est donnée par $d \times 2^k$ densités de probabilités des lois de Y_n^j conditionnelles à $X_n^* = \Omega_i$, que l'on note g_i^j. Nous définissons les masses $M^1, \dots, M^d$ par :

$$M_y^j[(\Omega_1, \dots, \Omega_N)] = \frac{\prod_{i=1}^{N} g_{\Omega_i}^j(y_i^j)}{\sum_{(\Omega_1,\dots,\Omega_N)\in(\Omega^*)^N} \prod_{i=1}^{N} g_{\Omega_i}^j(y_i^j)} \tag{10}$$

Proposition 3.4 **—** *Pour chaque $n \in S$ et chaque $A = (\Omega_1, \dots, \Omega_N) \in (\Omega^*)^N$, nous avons :*

$$M_y^j[(\Omega_1, \dots, \Omega_N)] = \prod_{n\in S} M_{n,y}^j(\Omega_n) \tag{11}$$

La démonstration de la proposition 1 est donnée dans [4].

En plus de l'information contenue dans les observations, nous considérons l'information *a priori* modélisée par la fonction de masse M^0, qui est la loi de X^*.
Nous fusionnons ces informations en supposant toujours au moins l'une des sources d'information bayesienne, de sorte que la fusion donne une masse bayesienne. Après fusion, qui peut être faite site par site, en ce qui concerne les observations, l'ensemble des observations est représentée par M^{sens}, $M^{sens} = M^1 \oplus \cdots \oplus M^d$ et $M = M^0 \oplus M^{sens}$.
Deux cas de figure sont alors envisageables, l'information *a priori* est évidentielle et l'information issue des observations est bayesienne, ou l'information *a priori* est bayesienne et l'information issue des observations est évidentielle. Nous étudions les deux cas et nous montrons que dans chacun d'entre eux le résultat de la fusion est une probabilité égale à la loi *a posteriori* d'une chaîne de Markov cachée classique.

3.2.1 Information a priori évidentielle

Soit (X^*) une chaîne de Markov cachée telle que l'information *a priori* soit évidentielle de fonction de masse M^0 et l'observation bayesienne de fonction de masse M^{sens}. Nous avons le résultat suivant :

Proposition 3.5 **—** *M est une distribution* a posteriori *d'une chaîne de Markov cachée classique dont la loi est donnée par :*

$$p_i = \sum_{\Omega_i\cap\omega_i\neq\emptyset} \pi_i^* \qquad b_{ij} = \sum_{\Omega_i\cap\omega_i\neq\emptyset} a_{ij}^* \tag{12}$$

bruitée de façon classique par M^{sens}.

3.2.2 Informations conditionnelles évidentielles

Soit X^* une chaîne de Markov cachée telle que l'information *a priori* soit bayesienne de fonction de masse M^0, qui est la probabilité P_{X^*}, et l'information conditionnelle soit évidentielle de fonction de masse M^{sens}. Nous avons le résultat suivant :

Proposition 3.6 **—** *M est une distribution* a posteriori *d'une chaîne de Markov cachée classique dont la loi est donnée par M^0 et le bruit est défini par :*

$$g_{x_n}(y_n) = \sum_{x_n\cap A_n\neq\emptyset} g_{A_n}(y_n) \quad A_n \in \Omega^* \tag{13}$$

où $x = (x_1, \dots, x_N)$ et $y = (y_1, \dots, y_N)$ sont les réalisations de X et Y.

Les démonstrations des propositions [2,3] sont présentées dans [4].
Ainsi d'après les propositions 2 et 3, la fusion de plusieurs sources d'information dont l'une au moins est bayesienne, donne une chaîne de Markov. Tous les traitements valables dans le cas des chaînes de Markov cachées classiques restent donc applicables au modèle considéré.

4 Restauration de chaînes de Markov cachées évidentielles

Nous étudions, dans cette section, la fusion de deux canaux d'observation. Le processus après fusion étant une chaîne de Markov, nous pouvons notamment appliquer l'algorithme de restauration bayesienne MPM présenté à la section 2.2. Nous noterons MPM1 l'algorithme MPM lorsque M^0 est évidentiel et MPM2 l'algorithme MPM lorsque M^{sens} est évidentiel. Le nombre de classes sera pris égal à deux, $\Omega = \{0, 1\}$ et $\Omega^* = \{0, 1, \Omega\}$, l'ensemble $\emptyset$ ayant été rejeté ($card(\Omega^*) = 2^k - 1$).
Nous simulons une chaîne A de taille 1000 et de paramètres *a priori*

$$A = \begin{pmatrix} 0,98 & 0,02 \\ 0,02 & 0,98 \end{pmatrix} \qquad \pi = (0,5 \quad 0,5)$$

Puis nous segmentons la chaîne A bruitée, avec les algorithmes MPM, MPM1, MPM2 pour différentes valeurs des paramètres.

4.1 Paramètres a priori mal définis

Nous bruitons la chaîne A avec les bruits gaussiens N(1,1) et N(3,3) dans le canal 1 et N(1,1) et N(2,2) dans le canal 2. Nous fixons les paramètres des lois conditionnelles aux vraies valeurs utilisées pour le bruitage et nous faisons varier les paramètres des matrices de transition. Les probabilités initiales n'ont pas d'influence significative sur la qualité de la segmentation aussi les fixons nous à $\pi = (0,5 \quad 0,5)$ pour MPM et à $\pi^* = (0,33 \quad 0,33 \quad 0,33)$ pour MPM1. Les taux d'erreur après restauration pour différentes matrices de transition, notés τ_{MPM} et τ_{MPM1}, sont présentés dans la table 1.

Les taux d'erreur MPM se sont dégradés lorsque les paramètres de la loi de X s'éloignent de leurs vraies valeurs. L'introduction d'une classe évidentielle peut améliorer le taux de bonne classification de façon significative. Par ailleurs, l'utilisation du modèle évidentiel ne détériore jamais les résultats obtenus avec le modèle probabiliste utilisant les paramètres erronés.

4.2 Paramètres du bruit mal définis

Nous nous intéressons dans ce paragraphe à l'apport d'une information conditionnelle évidentielle. Pour cela nous fixons les valeurs des paramètres *a priori* aux vraies valeurs et nous faisons varier les paramètres du bruit autour des leurs vraies valeurs. La chaîne A est bruitée dans les canaux 1 et 2, avec les bruits suivants : N(1,2) pour la classe 0 et N(2,2) pour la classe 1. Nous étudions le comportement de MPM et MPM2 lorsque les moyennes s'éloignent de leurs vraies valeurs, les variances étant fixées à 2 pour les classes 0 et 1. Nous notons m_0 et m_1 les moyennes des classes 0 et 1 dans les deux canaux, m_Ω la moyenne "évidentielle" et var_Ω la variance "évidentielle". Ainsi MPM utilise comme densités de probabilités N(m_0,2) et N(m_1,2) alors que MPM2 utilise comme densités de probabilités N(m_0,2), N(m_1,2) et N(m_Ω,var_Ω) pour les deux canaux. Nous présentons dans la table 2 les taux d'erreur τ_{MPM} et τ_{MPM2} lorsque m_0 et m_1 varient.
Comme pour les paramètres des lois *a priori*, lorsque les paramètres des lois conditionnelles sont mal déterminés, il est intéressant d'utiliser la modélisation évidentielle. Si les paramètres considérés sont les vrais paramètres, MPM2 dégrade légèrement la segmentation par rapport à MPM, sinon il peut apporter une forte amélioration.

5 Conclusion

Nous avons présenté dans cet article un modèle utilisant l'approche évidentielle. Nous avons pour cela introduit la notion de chaîne de Markov cachée évidentielle. L'utilisation de cette modélisation pour une segmentation à deux classes semble donner des résultats prometteurs. En effet, l'étude numérique a montré que la modélisation par chaînes de Markov cachées évidentielles améliore la segmentation par rapport à la modélisation bayesienne classique lorsque les paramètres *a priori* ou les paramètres des densités conditionnelles sont connus avec insuffisament de précisions. Les améliorations, en termes de sites bien classés, peuvent être très importantes et les taux obtenus peuvent approcher le taux théorique, à savoir le taux probabiliste obtenu avec les vrais paramètres.

Références

[1] Appriou. A. Probabilités et incertitude en fusion de données multi-senseurs. Revue scientifique et technique de la Défense, pages 27-40, 1er trimestre 1991.

[2] Baum. L. E, Petrie. T, Soules. G and Weiss. N. A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains. Annals of Mathematical Statistics, 41(1) :164-171, 1970.

[3] Bloch. I. Information combination operators for data fusion : a comparative review with classification. IEEE Transactions on Systems, Man and Cybernetics -Part A Systems and Humans, 26(1) :52-67, jan 96.

[4] Giordana. N. Segmentation non supervisée d'images multi-spectrales par chaînes de Markov cachées. Thèse de Doctorat de l'Université Technologique de Compiègne, Déc 96.

[5] Nimier. V et Appriou. A. Utilisation de la théorie de Dempster-Shafer pour la fusion d'informations. Quinzième colloque GRETSI, pages 137-140, Juan-Les-Pins Sept 95.

[6] Shafer. G. A mathematical theory of evidence. Priceton University Press, 1976.

$\begin{pmatrix} 0,98 & 0,02 \\ 0,02 & 0,98 \end{pmatrix}$ $\tau_{MPM} = 1,5\%$		
$\begin{pmatrix} 0,8 & 0,2 \\ 0,2 & 0,8 \end{pmatrix}$ $\tau_{MPM} = 5\%$	$\begin{pmatrix} 0,8 & 0,1 & 0,1 \\ 0,1 & 0,8 & 0,1 \\ 0,1 & 0,1 & 0,8 \end{pmatrix}$ $\tau_{MPM1} = 2,3\%$	
$\begin{pmatrix} 0,7 & 0,3 \\ 0,3 & 0,7 \end{pmatrix}$ $\tau_{MPM} = 9,5\%$	$\begin{pmatrix} 0,7 & 0,2 & 0,1 \\ 0,2 & 0,7 & 0,1 \\ 0,2 & 0,1 & 0,7 \end{pmatrix}$ $\tau_{MPM1} = 5,6\%$	$\begin{pmatrix} 0,7 & 0,1 & 0,2 \\ 0,1 & 0,7 & 0,2 \\ 0,1 & 0,2 & 0,7 \end{pmatrix}$ $\tau_{MPM1} = 2,5\%$

TAB. 1 — Taux d'erreur après segmentation par MPM et MPM1

m_0	m_1	m^*_Ω	var^*_Ω	τ_{MPM}	τ^*_{MPM2}
0	3	1,8	2	8,2%	3,6%
0,5	2,5	1,8	0,5	5 %	3,5%
1	2	1,8	2	3,5%	3,8%
1,5	2	1,8	2	3,7%	4,1%
2	2,5	1,2	1,5	10,7%	9,2%

ligne 3 : vraies valeurs des moyennes

TAB. 2 — Taux d'erreur après segmentation par MPM et MPM2