Available online at www.sciencedirect.com

SCIENCE DIRECT®

C. R. Acad. Sci. Paris, Ser. I 336 (2003) 869–872

COMPTES RENDUS MATHEMATIQUE

Statistique/Probabilités

Arbres de Markov Triplet et fusion de Dempster–Shafer

Wojciech Pieczynski

GET/INT, département CITI, 9, rue Charles Fourier, 91000 Evry, France

Reçu le 14 novembre 2002 ; accepté après révision le 1er avril 2003

Présenté par Paul Deheuvels

Résumé

Les chaînes de Markov cachées (CMC) (X, Y) ont été récemment généralisées aux chaînes de Markov Triplet (CMT), lesquelles gardent les mêmes pouvoirs de restauration du processus caché X à partir du processus observé Y. Par ailleurs, dans une CMC (X, Y) la loi a posteriori de X, qui est de Markov, peut être vue comme une fusion de Dempster–Shafer (fusion DS) de sa loi a priori avec une probabilité q définie à partir des observations $Y = y$. Lorsque l'on se place dans le contexte de la théorie de l'évidence en remplaçant la loi de X par une fonction de masse M admettant une écriture markovienne similaire (une modélisation plus générale redonnant le modèle classique pour une M particulière), sa fusion DS avec q généralise la probabilité a posteriori. Bien que le résultat de cette fusion ne soit, en général, pas une chaîne de Markov, il a été établi qu'il est une CMT, ce qui autorise les divers traitements d'intérêt. L'objet de cette Note est de présenter diverses généralisations de ce dernier résultat : (i) extension aux CMC plus généraux ; (ii) q, qui peut éventuellement être une fonction de masse Q, est elle même résultat de fusion DS ; enfin, (iii) tous les résultats sont étendus aux arbres de Markov cachés (AMC), qui englobent les CMC. ***Pour citer cet article : W. Pieczynski, C. R. Acad. Sci. Paris, Ser. I 336 (2003).***

Abstract

Triplet Markov trees and Dempster–Shafer fusion. The hidden Markov chains (HMC) (X, Y) have been recently generalized to triplet Markov chains (TMC), which enjoy the same capabilities of restoring a hidden process X from the observed process Y. The posterior distribution of X can be viewed, in an HMC, as a particular case of the so called "Dempster–Shafer fusion" (DS fusion) of the prior Markov with a probability q defined from the observation $Y = y$. As such, when we place ourselves in the Dempster–Shafer theory of evidence by replacing the probability distribution of X by a mass function M having an analogous Markov form (which gives again the classical Markov probability distribution in a particular case), the result of DS fusion of M with q generalizes the conventional posterior distribution of X. Although this result is not necessarily a Markov distribution, it has been recently shown that it is a TMC, which renders traditional restoration methods applicable. The aim of this Note is to present some generalizations of the latter result: (i) more general HMCs can be considered; (ii) q, which can possibly be a mass function Q, is itself a result of the DS fusion; and (iii) all these results are finally specified in the hidden Markov trees (HMT) context, which generalizes the HMC one. ***To cite this article: W. Pieczynski, C. R. Acad. Sci. Paris, Ser. I 336 (2003).***

Adresse e-mail : Wojciech.Pieczynski@int-evry.fr (W. Pieczynski).

1631-073X/03/$ – see front matter © 2003 Académie des sciences. Publié par Éditions scientifiques et médicales Elsevier SAS. Tous droits réservés.
doi:10.1016/S1631-073X(03)00188-2

1. Introduction

On considère un processus inobservable $X = (X_1, \ldots, X_n, \ldots)$, chaque X_i prenant ses valeurs dans un ensemble fini de classes $\Omega = \{\omega_1, \ldots, \omega_k\}$, et un processus observé $Y = (Y_1, \ldots, Y_n, \ldots)$, chaque Y_i prenant ses valeurs dans R. Le processus $Z = (X, Y)$ sera dit « chaîne de Markov cachée avec bruit indépendant » (CMC-BI) si sa loi s'écrit $p(x, y) = p(x_1)p(x_2|x_1) \cdots p(x_n|x_{n-1})p(y_1|x_1) \cdots p(y_n|x_n)$. Il sera dit « chaîne de Markov cachée » (CMC) si X et Z sont de Markov, et il sera dit « chaîne de Markov couple » (CMCouple) s'il est de Markov. Enfin, le processus Z sera dit « chaîne de Markov triplet » (CMT) s'il existe un processus $U = (U_1, \ldots, U_n, \ldots)$, chaque U_i à valeurs dans un ensemble fini $\Lambda = \{\lambda_1, \ldots, \lambda_m\}$, tel que $T = (X, U, Y)$ est une chaîne de Markov. On peut alors montrer qu'une CMC-BI est une CMC, une CMC est une CMCouple, et une CMCouple est une CMT, les réciproques étant fausses [6,8]. Les CMC-BI sont les plus anciens et bien connus, l'utilisation des CMC qui ne soient pas des CMC-BI étant plus rare. L'introduction des CMCouple et CMT qui ne soient pas des CMC est très récente et leur potentiel, qui semble prometteur, commence seulement à être exploré [3]. Par ailleurs, dans le cas de CMC-BI on peut considérer $p(x|y)$ comme le résultat de la fusion de Dempster–Shafer (fusion DS) de la probabilité $p(x) = p(x_1)p(x_2|x_1) \cdots p(x_n|x_{n-1})$ avec la probabilité (y est fixé) $q(x) \propto p(y_1|x_1) \cdots p(y_n|x_n)$ [1,9,10]. Ce type de fusion est valable dans le cadre plus général, lorsque la probabilité $p(x)$ est généralisée à une « fonction de masse » M. On montre alors que le résultat de la fusion de M par q, noté $M \oplus q$ (qui devient $p(x|y)$ lorsque M devient $p(x)$), est une CMT, ce qui permet les traitements bayésiens, malgré le fait que $M \oplus q$ ne soit pas nécessairement une chaîne de Markov [8].

L'objet de cette Note est de présenter diverses généralisations de ce dernier résultat. Dans un premier temps, on remplace CMC-BI par CMC quelconque. Ensuite, on considère les CMC avec les Y_i à valeurs dans R^r, avec la possibilité pour la probabilité q d'être également généralisée à une fonction de masse Q. Enfin, les résultats sont étendus aux Arbres de Markov Cachés (AMC).

2. Fusion DS dans CMC

Soit $\Omega = \{\omega_1, \ldots, \omega_k\}$ et $P(\Omega)$ l'ensemble des parties de Ω. Une fonction de masse (FM) est une application $M : P(\Omega) \to [0, 1]$ vérifiant $M(\emptyset) = 0$ et $\sum_{A \in P(\Omega)} M(A) = 1$ (on retrouve une probabilité classique comme cas particulier de FM nulle en dehors des singletons) [1,9,10], voir également [2] pour l'application de la fusion DS à la segmentation d'images dans un cas simple des champs de Markov cachés. On dira qu'une FM M définie sur $P(\Omega^n)$ est une chaîne de Markov évidentielle (CME) si elle est nulle en dehors de $[P(\Omega)]^n$ et si $M(A_1, \ldots, A_n) = M(A_1)M(A_2|A_1) \cdots M(A_n|A_{n-1})$; ainsi une chaîne de Markov (CM) classique est un cas particulier d'une CME.

Lemme 2.1. *Dans le cadre ci-dessus, soit M une CME et q une CM. Alors la fusion DS $M \oplus q$ est une CMT $T = (X, U, Y)$, où chaque U_i prend ses valeurs dans $\Lambda = P(\Omega)$, et dont la loi est donnée par $p(t_1) \propto 1_{[x_1 \in u_1]} M(u_1)q(x_1)$, et $p(t_{i+1}|t_i) \propto 1_{[x_i \in u_i]} 1_{[x_{i+1} \in u_{i+1}]} M^0(u_{i+1}|u_i)q(x_{i+1}|x_i)$.*

Démonstration. Soit $x = (x_1, \ldots, x_n) \in \Omega^n$. La fusion DS s'écrit (dans les deux premières sommes $x = (x_1, \ldots, x_n)$ est fixée et $u = (u_1, \ldots, u_n)$ varie dans $[P(\Omega)]^n$ avec $x_1 \in u_1, \ldots, x_n \in u_n$) :

$$
\begin{aligned}
(M \oplus q)(x) &\propto \sum_{x \in u} M(u_1, \ldots, u_n)q(x) \propto \sum_{x \in u} M(u_1)M(u_2|u_1) \cdots M(u_n|u_{n-1})q(x) \\
&= \sum_{u \in [P(\Omega)]^n} 1_{[x_1 \in u_1]} \cdots 1_{[x_n \in u_n]} M(u_1)M(u_2|u_1) \cdots M(u_n|u_{n-1})q(x) \\
&= \sum_{u \in [P(\Omega)]^n} p(t_1)p(t_2|t_1) \cdots p(t_n|t_{n-1}) = \sum_{u \in [P(\Omega)]^n} p(t_1, \ldots, t_n)
\end{aligned}
$$

d'où le résultat. □

Considérons une CMCouple $Z=(X,Y)$. Sachant que $p(z_i,z_{i+1})=p(x_i,x_{i+1})p(y_i,y_{i+1}|x_i,x_{i+1})$, la loi de Z peut s'écrire

$$p(x,y)=\underbrace{p(x_1)p(x_2|x_1)\cdots p(x_n|x_{n-1})}_{a(x)}\underbrace{\frac{p(y_1,y_2|x_1,x_2)\cdots p(y_{n-1},y_n|x_{n-1},x_n)}{p(y_2|x_2)\cdots p(y_{n-1}|x_{n-1})}}_{b(x,y)}; \tag{1}$$

On constate qu'une CNS pour que Z soit une CMC est que $a(x)$ soit la loi de X, ou encore que l'intégrale de $b(x,y)$ par rapport à y vaille 1 (ce qui est le cas, par exemple, lorsque pour tout $1\leqslant i\leqslant n-1$, $p(y_i|x_i,x_{i+1})=p(y_i|x_i)$).

Considérons une CMC dont la loi est donnée par (1) (donc, avec $a(x)=p(x)$). L'observation y étant fixée, $b(x,y)$ définit, après la normalisation, une CM que l'on notera $q(x)=q(x_1)q(x_2|x_1)\cdots q(x_n|x_{n-1})$. Nous pouvons donc dire que $p(x|y)\propto p(x)q(x)$, ce qui signifie que $p(x|y)$ est la fusion DS de deux CM $p(x)$ et $q(x)$. Lorsque la CM $p(x)$ devient une CME M nous pouvons énoncer, à partir du Lemme 2.1, le résultat suivant :

Proposition 2.2. *Soit M une CME définie sur $[P(\Omega)]^n$ et, pour $y\in R^n$, soit la CM q définie par $b(x,y)$ ci-dessus. Alors la fusion DS $M\oplus q$ est la loi de X conditionnelle à $Y=y$, où (X,Y) est une CMT $T=(X,U,Y)$, avec U_i à valeurs dans $P(\Omega)$ et dont la loi est définie par $p(t_1)\propto 1_{[x_1\in u_1]}M(u_1)q(x_1)$, et $p(t_{i+1}|t_i)\propto 1_{[x_i\in u_i]}1_{[x_{i+1}\in u_{i+1}]}M^0(u_{i+1}|u_i)q(x_{i+1}|x_i)$.*

Remarque 1. (1) Nous obtenons ainsi une généralisation par rapport aux HMC-BI étudiées dans [8], où la probabilité q est de la forme $q(x)=q(x_1)\cdots q(x_n)$;

(2) Lorsque $T=(X,U,Y)$ est stationnaire (les $p(t_i,t_{i+1})$ ne dépendent pas de i) on peut estimer les paramètres, ce qui permet la mise en place des méthodes de restauration non supervisées, par des adaptations des méthodes classiques comme la méthode « Expectation–Maximization » (EM).

Considérons (1) à nouveau, avec $p(x)$ la loi de X et $q(x)$ définie comme ci-dessus (donc $p(x|y)\propto p(x)q(x)$). Supposons que la CM q devient une CME Q; le lemme s'applique toujours et nous pouvons énoncer la

Proposition 2.3. *Soit $p(x)$ une CM (définie sur Ω^n) et Q une CME définie sur $[P(\Omega)]^n$, pour $y\in R^n$, par $Q(A)=Q(A_1)Q(A_2|A_1)\cdots Q(A_n|A_{n-1})\propto [p(y_1,y_2|A_1,A_2)\cdots p(y_{n-1},y_n|A_{n-1},A_n)]/p(y_2|A_2)\cdots p(y_{n-1}|A_{n-1})$. Alors la fusion DS $p\oplus Q$ est la loi de X conditionnelle à $Y=y$, où (X,Y) est une CMT $T=(X,U,Y)$, avec U_i à valeurs dans $P(\Omega)$ et dont la loi est définie par $p(t_1)\propto 1_{[x_1\in u_1]}p(x_1)Q(u_1)$, et $p(t_{i+1}|t_i)\propto 1_{[x_i\in u_i]}1_{[x_{i+1}\in u_{i+1}]}p(x_{i+1}|x_i)Q(u_{i+1}|u_i)$.*

Enfin, dans le cas des r capteurs indépendants les résultats des deux propositions précédentes se généralisent à

Proposition 2.4. *Soit $M=Q^0$ une CME définie sur $[P(\Omega)]^n$ comme dans la Proposition 2.2, supposons que chaque $Y_i=(Y_i^1,\dots,Y_i^r)$ est à valeurs dans R^r, et notons Q^j la CME définie par $Y^j=(Y_1^j,\dots,Y_n^j)$ comme dans la Proposition 2.3. Supposons que l'une au mois (Q^0 par exemple) des EMC $Q^0,Q^1,\dots,Q^r$ est une CM. Alors la fusion DS $Q^0\oplus Q^1\oplus\cdots\oplus Q^r$ est la loi de X conditionnelle à $Y=y$, où (X,Y) est une CMT $T=(X,U,Y)$ avec $U_i=(U_i^1,\dots,U_i^r)$ à valeurs dans $[P(\Omega)]^r$ et dont la loi est définie par*

$$p(t_1)\propto 1_{[x_1\in u_1^1\cap\cdots u_1^r]}Q^0(x_1)Q^1\big(u_1^1\big)\cdots Q^r\big(u_1^r\big),\quad \textit{et}$$
$$p(t_{i+1}|t_i)\propto 1_{[x_i\in u_i^1\cap\cdots\cap u_i^r]}1_{[x_{i+1}\in u_{i+1}^1\cap\cdots\cap u_{i+1}^r]}Q^0(x_{i+1}|x_i)Q^1\big(u_{i+1}^1|u_i^1\big)\cdots Q^r\big(u_{i+1}^r|u_i^r\big).$$

3. Arbres de Markov

Soit S un ensemble fini d'indices, et $X=(X_s)_{s\in S}$, $Y=(Y_s)_{s\in S}$, $U=(U_s)_{s\in S}$ trois processus, avec chaque $T_s=(X_s,U_s,Y_s)$ à valeurs dans $\Omega\times\Lambda\times R^r$. On note $Z_s=(X_s,Y_s)$, $V_s=(X_s,U_s)$, et Z,V les processus correspondants. Soit $S^1, \ldots, S^n$ une partition de S. Posons $N = \mathrm{Card}(S)$ et, pour $1 \leqslant i \leqslant n$, $N^i = \mathrm{Card}(S^i)$. Supposons que $N^1 \leqslant N^2 \leqslant \cdots \leqslant N^n$ (avec $N^1 = 1$) et pour tout $1 \leqslant i \leqslant n-1$, associons à chaque $s \in S^i$ un sous-ensemble s^+ de S^{i+1}, appelé « enfants de s », de manière à ce que $(s^+)_{s \in S^i}$ soit une partition de S^{i+1} (l'unique élément de N^1, noté s_1, est appelé « racine »). Pour tout $s \in S - S^1$, son unique « père » sera noté s^-. Le processus T est appelé Arbre de Markov Triplet (AMT) si sa loi s'écrit $p(t) = p(t_{s_1}) \prod_{i=2}^n \prod_{s \in S^i} p(t_s|t_{s^-})$. C'est alors un Arbre de Markov Couple (AMCouple) $T = (V, Y)$, ce qui permet divers traitements [7]. En particulier, $p(v_s|y)$ (donc également $p(x_s|y)$) et $p(v_s|v_{s^-}, y)$ sont calculables, ce qui permet le calcul de la restauration bayésienne MPM et la simulations de V (donc de X) selon $p(v|y)$. Par ailleurs, les AMT et AMCouple généralisent les CMT et CMCouple, ces dernières étant obtenues lorsque tous les s^+ sont des singletons.

Les résultats de la Proposition 2.4 se généralisent alors de la façon suivante.

De manière analogue à (1) on peut écrire

$$p(z) = p(z_{s_1}) \prod_{i=2}^n \prod_{s \in S^i} p(z_s|z_{s^-}) = \underbrace{\left[p(x_{s_1}) \prod_{i=2}^n \prod_{s \in S^i} p(x_s|x_{s^-})\right]}_{a(x)} \underbrace{\left[p(y_{s_1}|x_{s_1}) \prod_{i=2}^n \prod_{s \in S^i} \frac{p(y_{s^-}, y_s|x_{s^-}, x_s)}{p(y_{s^-}|x_{s^-})}\right]}_{b(x)}. \quad (2)$$

Par une démarche analogue à celle suivie ci-dessus dans le cadre des chaînes de Markov, on peut étendre l'AM de loi $a(x)$ à un AM évidentiel (AME) Q^0 défini sur $[P(\Omega)]^N$. De même, $b(x)$, qui est un produit (dû à l'indépendance des senseurs) de r AM, devient, lorsque les r AM deviennent des AME $Q^1, \ldots, Q^r$, le résultat de la fusion DS $Q^1 \oplus \cdots \oplus Q^r$. La Proposition 2.4 admet alors l'extension aux AME suivante :

Proposition 3.1. *Soit Q^0 un AME définie sur $[P(\Omega)]^N$ et $Q^1, \ldots, Q^r$ les r AME définis à partir des observations $Y^1 = y^1, \ldots, Y^r = y^r$ par $Q^j(A) \propto p(y^j_{s_1}|A_{s_1}) \prod_{i=2}^n \prod_{s \in S^i} \frac{p(y^j_{s^-}, y^j_s|A_{s^-}, A_s)}{p(y^j_{s^-}|A_{s^-})}$. Supposons que l'un au mois (Q^0 par exemple) des AME $Q^0, Q^1, \ldots, Q^r$ est un AM. Alors la fusion DS $Q^0 \oplus Q^1 \oplus \cdots \oplus Q^r$ est la loi de X conditionnelle à $Y = y$, où (X, Y) est un AMT $T = (X, U, Y)$ avec $U_i = (U_i^1, \ldots, U_i^r)$ à valeurs dans $[P(\Omega)]^r$ et dont la loi est définie par $p(t_1) \propto 1_{[x_1 \in u_1^1 \cap \cdots \cap u_1^r]} Q^0(x_1) Q^1(u_1^1) \cdots Q^r(u_1^r)$, et*

$$p(t_s|t_{s^-}) \propto 1_{[x_{s^-} \in u^1_{s^-} \cap \cdots \cap u^r_{s^-}]} 1_{[x_s \in u_s^1 \cap \cdots \cap u_s^r]} Q^0(x_s|x_{s^-}) Q^1\left(u_s^1|u_{s^-}^1\right) \cdots Q^r\left(u_s^r|u_{s^-}^r\right).$$

Remarque 2. Les observations y_s peuvent ne pas exister pour certains s, auquel cas on pose $p(y_s|A_s) = 1$ pour tout A_s. Par exemple, les observations peuvent concerner uniquement certains S^i, avec applications en traitements d'images multirésolution (voir [5] pour l'utilisation des AMC-BI classiques).

L'extension des Réseaux Bayésiens cachés [4] à des Réseaux Bayésiens « Triplet » constitue une perspective naturelle pour la poursuite du travail présenté.

Références

[1] A. Appriou, Probabilités et incertitude en fusion de données multisenseurs, Rev. Sci. Techn. Défense 11 (1991) 27–40.

[2] A. Bendjebbour, Y. Delignon, L. Fouque, V. Samson, W. Pieczynski, Multisensor images segmentation using Dempster–Shafer fusion in Markov fields context, IEEE Trans. GRS 39 (8) (2001) 1789–1798.

[3] S. Derrode, W. Pieczynski, SAR image segmentation using generalized Pairwise Markov Chains, in: SPIE's International Symposium on Remote Sensing, September 22–27, Crete, Greece, 2002.

[4] F.V. Jensen, An Introduction to Bayesian Networks, UCL Press, 2000.

[5] J.-M. Laferté, P. Pérez, F. Heitz, Discrete Markov image modeling and inference on the quadtree, IEEE Trans. Image Processing 9 (3) (2000) 390–404.

[6] W. Pieczynski, Pairwise Markov chains, IEEE Trans. PAMI 25 (5) 2003, to appear.

[7] W. Pieczynski, Arbres de Markov couple, C. R. Acad. Sci. Paris, Ser. I 335 (2002) 79–82.

[8] W. Pieczynski, Chaînes de Markov triplet, C. R. Acad. Sci. Paris, Ser. I 335 (2002) 275–278.

[9] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.

[10] P. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, Internat. J. Approximate Reasoning 9 (1993) 1–35.