C. R. Acad. Sci. Paris, Ser. I 335 (2002) 275–278

Statistique/*Statistics*
(Probabilités/*Probability Theory*)

Chaînes de Markov Triplet

Wojciech Pieczynski

Institut National des Télécommunications, Département CITI, 9, rue Charles Fourier, 91000 Evry, France

Reçu le 9 avril 2002 ; accepté le 31 mai 2002

Note présentée par Paul Deheuvels.

Résumé Les Chaînes de Markov Cachées (CMC), qui permettent l'estimation des variables d'intérêt dans les cas des quantités importantes des données, sont largement utilisées dans les problèmes le plus divers. Ces modèles ont été récemment généralisés aux Chaînes de Markov Couple (CMCouple), qui permettent des modélisations plus complètes des liens entre les processus caché et observé. Nous proposons dans cette Note de généraliser ces derniers aux modèles « Chaînes de Markov Triplet » (CMT) dans lesquels la loi du couple (processus caché, processus observé) est la loi marginale d'un triplet Markovien. Nous montrons la calculabilité des estimations Bayésiennes du processus caché et présentons une CNS pour qu'une CMT soit une CMCouple, montrant en particulier que les modèles CMT sont strictement plus généraux que les modèles CMCouple. Nous précisons également un lien avec la fusion de Dempstert–Shafer. ***Pour citer cet article : W. Pieczynski, C. R. Acad. Sci. Paris, Ser. I 335 (2002) 275–278.*** © 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

Triplet Markov Chains

Abstract The Hidden Markov Chains (HMC) are widely applied in various problems. This succes is mainly due to the fact that the hidden process can be recovered even in the case of very large set of data. These models have been recetly generalized to 'Pairwise Markov Chains' (PMC) model, which admit the same processing power and a better modeling one. The aim of this note is to propose further generalization called Triplet Markov Chains (TMC), in which the distribution of the couple (hidden process, observed process) is the marginal distribution of a Markov chain. Similarly to HMC, we show that posterior marginals are still calculable in Triplets Markov Chains. We provide a necessary and sufficient condition that a TMC is a PMC, which shows that the new model is strictly more general. Furthermore, a link with the Dempster–Shafer fusion is specified. ***To cite this article: W. Pieczynski, C. R. Acad. Sci. Paris, Ser. I 335 (2002) 275–278.*** © 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

1. Introduction

La modélisation par processus de Markov cachés consiste à considérer, pour l'ensemble d'indices S, de deux processus stochastiques : le processus d'intérêt inobservable, ou « caché », $X = (X_s)_{s \in S}$, et le processus observable $Y = (Y_s)_{s \in S}$. La loi du couple (X, Y) est donnée par la loi de X, qui est supposée Markovienne, et les lois de Y conditionnelles à X. Lorsque ces dernières sont suffisamment simples, les lois de X conditionnelles à Y sont également Markoviennes, ce qui permet divers traitements ayant pour

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

S1631-073X(02)02462-7/FLA

objectif d'estimer le processus d'intérêt X à partir du processus observé Y. La loi de X est parfois dite « loi a priori », et sa loi conditionnelle à $Y = y$ « loi a posteriori ». Le succès des Modèles de Markov Cachés (MMC) est essentiellement dû au fait que ces traitements sont possibles pour un cardinal élevé de S. Cette possibilité est impliquée par la calculabilité, ou l'estimation par des méthodes de type Monte Carlo par Chaînes de Markov (MCMC), des diverses lois a posteriori marginales. En particulier, les chaînes de Markov cachées [2,4] et les champs de Markov cachés [5–7] jouent un rôle important en traitement du signal et des images.

Malgré la bonne robustesse de ces modèles, la simplicité des lois de Y conditionnelles à X généralement considérées dans les MMC (on suppose fréquemment que les $(Y_s)_{s\in S}$ sont indépendantes conditionnellement à X) est parfois difficile à justifier. De plus, cette simplicité peut rendre impossibles la modélisation correcte de certaines situations réelles. A titre d'exemple, la présence des classes texturées dans une image que l'on souhaite segmenter ne peut être prise en compte dans le cadre strict d'un champ de Markov caché et l'on est obligé de faire appel à des approximations du modèle. Afin de pallier à cette limitation, les champs, chaînes, et arbres de Markov cachés (MMC) ont été récemment généralisés aux champs, chaînes, et arbres de Markov Couple (MMCouple) [8–11], dans lesquels on considère directement la Markovianité du couple (X, Y). Cela implique la Markovianité des lois de X conditionnelles à Y, d'une part, et la Markovianité des lois de Y conditionnelles à X, d'autre part. La première propriété montre que ces modèles offrent le même pouvoir de traitements que les MMC, et la deuxième montre que leur pouvoir modélisant est supérieur à celui des MMC. En effet, la Markovianité des lois de Y conditionnelles à X ne peut être considérée dans le cadre des MMC car elle n'implique pas, si elle est jointe à la Markovianité de X, la Markovianité des lois de X conditionnelles à Y. En d'autres termes, les MMCouple sont plus généraux que les MMC car dans un MMCouple la loi de X n'est plus nécessairement de Markov, alors qu'un MMC est toujours un MMCouple.

Nous proposons dans cette Note une extension des Chaînes de Markov Couple (CMCouple) aux Chaînes de Markov Triplets (CMT), dans lesquelles la loi de (X, Y) est la loi marginale d'une chaîne de Markov (X, U, Y). Notons que nous recherchons toujours X à partir de Y, le processus U servant uniquement comme outil de calcul. Nous montrons que le modèle CMT est strictement plus général que le modèle CMCouple tout en offrant les mêmes possibiltés de traitements. Nous donnons également un exemple d'application des CMT dans la fusion de Dempster–Shafer [1,3,12,13].

2. Chaînes de Markov Triplets

Soit $S = N$ l'ensemble de nombres naturels privé de zero et $X = (X_n)_{n\in N}$, $Y = (Y_n)_{n\in N}$, $U = (U_n)_{n\in N}$ trois processus, où les X_n, Y_n, et U_n sont, respectivement, à valeurs dans $\Omega = \{\omega_1, \ldots, \omega_k\}$, R, et $\Lambda = \{\lambda_1, \ldots, \lambda_m\}$. Afin de simplifier les notations, nous posons $T_n = (X_n, U_n, Y_n)$, $Z_n = (X_n, Y_n)$, $V_n = (X_n, U_n)$, et T, Z, V les processus correspondants. Nous dirons que Z est une chaîne de Markov triplet (CMT) s'il existe un processus U tel que T est une chaîne de Markov. Dans ce cas la chaîne (V, Y) est une chaîne de Markov Couple ; nous pouvons donc écrire classiquement les lois de (V_i, Y) comme $p(v_i, y) = \alpha^i(v_i)\beta^i(v_i)$, avec $\alpha^i(v_i)$ (probabilités « Forward ») et $\beta^i(v_i)$ (probabilités « Backward ») définies respectivement par $\alpha^i(v_i) = p(y_1, \ldots, y_{i-1}, y_i, v_i)$ et $\beta^i(v_i) = p(y_{i+1}, \ldots, y_n|v_i, y_i)$ (les CMCouple permettent des calculs analogues aux CMC [10]). L'important est que $\alpha^i(v_i)$ et $\beta^i(v_i)$ sont calculables par les récursions directes et rétrogrades suivantes :

$$\alpha^1(v_1) = p(y_1, v_1), \quad \text{et} \quad \alpha^{i+1}(v_{i+1}) = \sum_{v_i\in\Omega\times\Lambda} \alpha^i(v_i)p(y_{i+1}, v_{i+1}|y_i, v_i) \quad \text{pour } 1 \leqslant i \leqslant n-1, \tag{2.1}$$

$$\beta^n(v_n) = 1, \quad \text{et} \quad \beta^i(v_i) = \sum_{v_{i+1}\in\Omega\times\Lambda} \beta^{i+1}(v_{i+1})p(y_{i+1}, v_{i+1}|y_i, v_i) \quad \text{pour } 1 \leqslant i \leqslant n-1. \tag{2.2}$$

Sachant que $p(x_i, y) = \sum_{u_i \in \Lambda} p(x_i, u_i, y) = \sum_{u_i \in \Lambda} p(v_i, y)$, il en résulte que $p(x_i|y)$ sont calculables. En particulier, la restauration Bayésienne Maximum des Marginales a Posteriori (MPM) donnée par

$$\left[\hat{x} = \hat{s}(y)\right] \quad \Leftrightarrow \quad \left[\hat{x}_i = \arg\max_{\omega \in \Omega} p(x_i = \omega|y)\right] \tag{2.3}$$

est calculable.

Nous montrons dans la Proposition 2.1 ci-après que le modèle CMT est strictement plus général et précisons une CNS pour qu'une CMT soit une CMCouple.

PROPOSITION 2.1. – *Soit $T = (X, U, Y)$ une CMT vérifiant*
(H) *Pour tout $i \in N$, $z_i = z_{i+2}$ implique $p(u_{i+1}|z_i, z_{i+1}) = p(u_{i+1}|z_{i+1}, z_{i+2})$.*
Alors $Z = (X, Y)$ est une CMCouple si et seulement si pour tout $i \in N$, $p(u_{i+1}|z_i, z_{i+1}) = p(u_{i+1}|z_{i+1})$.

La preuve, proche des preuves des références [10,11], est disponible sur demande auprés de l'Académie des sciences.

Exemple 2.1. – Soit U une chaîne de Markov. Supposons les variables (Z_i) indépendantes conditionnellement à U, avec $p(z_i|u) = p(z_i|u_i)$. On vérifie alors que T est une chaîne de Markov avec $p(t_1) = p(u_1)p(z_1|u_1)$ et $p(t_{i+1}|t_i) = p(u_{i+1}|u_i)p(z_{i+1}|u_{i+1})$. Nous pouvons donc calculer, en vertu de ce qui précède, les marginales $p(x_i|y)$. Cependant, $Z = (X, Y)$ n'est pas nécessairement une chaîne de Markov. En effet, $p(u_2|z_1, z_2) = p(u_2|z_2)$ équivaut à $p(z_1|u_2, z_2) = p(z_1|z_2)$. Or, on montre immédiatement que $p(z_1|u_2, z_2) = p(z_1|u_2)$. Donc (X, Y) est une chaîne de Markov si et seulement si $p(z_1|u_2) = p(z_1|z_2)$, ce qui est manifestement faux dans le cas général. Nous avons ainsi une CMT simple et maniable, sans qu'elle soit une CMC, ni même une CMCouple.

3. Fusion de Dempster–Shafer

Considérons l'ensemble $\Omega = \{\omega_1, \ldots, \omega_k\}$, notons $\mathrm{P}(\Omega) = \{A_1, \ldots, A_q\}$, avec $q = 2^n$, l'ensemble des sous-ensembles de Ω. Une fonction M de $\mathrm{P}(\Omega)$ dans $[0, 1]$ est dite « masse » si $M(\emptyset) = 0$ et $\sum_{A \in \mathrm{P}(\Omega)} M(A) = 1$. A une masse M sur $\mathrm{P}(\Omega)$ on peut alors associer la fonction de « plausibilité » Pl définie sur $\mathrm{P}(\Omega)$ par $Pl(A) = \sum_{A \cap B \neq \emptyset} M(B)$ et la fonction de « crédibilités » Cr définie sur $\mathrm{P}(\Omega)$ par $Cr(A) = \sum_{B \subset A} M(B)$.

Définition 3.1. – Une masse M définie sur $\mathrm{P}(\Omega^n)$ sera dite « chaîne de Markov évidentielle » (CME) si elle est nulle en dehors de $[\mathrm{P}(\Omega)]^n$ et si, en tant que probabilité sur l'ensemble disctret $[\mathrm{P}(\Omega)]^n$, M est une chaîne de Markov.

PROPOSITION 3.1. – *Soit M^0 une CME définie sur $[\mathrm{P}(\Omega)]^n$, et M^1 la probabilité sur $\mathrm{P}(\Omega^n)$ définie classiquement à partir de la réalisation $y \in R^n$ du processus stochastique des observations par $M^1(x_1, \ldots, x_n) \propto p(y_1|x_1) \cdots p(y_n|x_n)$.*

Alors la probabilité $M = M^0 \oplus M^1$ est une CMT, avec $\Lambda = \mathrm{P}(\Omega)$ et T la chaîne de Markov définie par $p(t_1) \propto 1_{[x_1 \in u_1]} M^0(u_1) p(y_1|x_1)$ et $p(t_{i+1}|t_i) \propto 1_{[x_i \in u_i]} 1_{[x_{i+1} \in u_{i+1}]} M^0(u_{i+1}|u_i) p(y_{i+1}|x_{i+1})$.

Notons que M^0 nulle en dehors des suites des singletons est une chaîne de Markov classique et alors le résultat de la fusion de Dempster–Shafer n'est autre que la loi a posteriori d'une CMC classique ; la probabilité donnée par la fusion peut ainsi être considérée comme une généralisation de cette dernière.

Démonstration. – Soit $x = (x_1, \ldots, x_n) \in \Omega^n$. Le résultat de la fusion de Dempster–Shafer s'écrit (les sommes sont prises sur les $u = (u_1, \ldots, u_n) \in [\mathrm{P}(\Omega)]^n$ tels que $x_1 \in u_1, \ldots, x_n \in u_n$) :

$$\left(M^0 \oplus M^1\right)(x) \propto \sum_{x \in u} M^0(u_1, \ldots, u_n) M^1(x) \propto \sum_{x \in u} M^0(u_1) M^0(u_2|u_1) \cdots M^0(u_n|u_{n-1}) M^1(x)$$

$$= \sum_{u \in \mathrm{P}(\Omega)^n} 1_{[x_1 \in u_1]} \ldots 1_{[x_n \in u_n]} M^0(u_1) M^0(u_2|u_1) \cdots M^0(u_n|u_{n-1}) M^1(x)$$
$$= \sum_{u \in \mathrm{P}(\Omega)^n} p(t_1) p(t_2|t_1) \cdots p(t_n|t_{n-1}) = \sum_{u \in \mathrm{P}(\Omega)^n} p(t_1, \ldots, t_n)$$

d'où le résultat.

Remarque 3.1. – Notons que la TMC ci-dessus, définie par $p(t_1) \propto 1_{[x_1 \in u_1]} M^0(u_1) p(y_1|x_1)$ et $p(t_{i+1}|t_i) \propto 1_{[x_i \in u_i]} 1_{[x_{i+1} \in u_{i+1}]} M^0(u_{i+1}|u_i) p(y_{i+1}|x_{i+1})$ un cas particulier du modèle de l'Exemple 2.1 ; en effet, U est une chaîne de Markov et $p(z_i|u_i) = p(x_i, y_i|u_i) \propto 1_{[x_i \in u_i]} M^0(u_i) p(y_i|x_i)$. Il en résulte que (X, Y) n'est pas une CMCouple. En d'autres termes, la fusion de Dempster–Shafer détruit la Markovianité ; cependant, son résultat est une CMT et les restaurations Bayésiennes sont envisageables.

La preuve de la Proposition 2.1 a été déposée aux Archives de l'Académie et sera conservée cinq ans ; une copie peut être obtenue sur demande.

Références bibliographiques

[1] A. Appriou, Probabilités et incertitude en fusion de données multisenseurs, Rev. Sci. Techn. Défense 11 (1991) 27–40.
[2] L.E. Baum, T. Petrie, G. Soules, N. Weiss, A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains, Ann. Math. Statist. 41 (1970) 164–171.
[3] 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.
[4] G.D. Fornay, The Viterbi algorithm, Proc. IEEE 61 (3) (1973) 267–277.
[5] S. Geman, D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, IEEE Trans. PAMI 6 (6) (1984) 721–741.
[6] J. Marroquin, S. Mitter, T. Poggio, Probabilistic solution of ill-posed problems in computational vision, J. Amer. Statist. Assoc. 82 (1987) 76–89.
[7] P. Pérez, Markov random fields and images, CWI Quart. 11 (4) (1998) 413–437.
[8] W. Pieczynski, A.-N. Tebbache, Pairwise Markov random fields and segmentation of textured images, Machine Graphics & Vision 9 (3) (2000) 705–718.
[9] W. Pieczynski, Pairwise Markov chains and Bayesian unsupervisedc fusion, in : Proc. 3rd International Conference on Information Fusion, FUSION 2000, Vol. 1, July 10–13, Paris, France, 2000, pp. MoD4-24–MoD4-31.
[10] W. Pieczynski, Pairwise Markov chains, IEEE Trans. PAMI, to appear.
[11] W. Pieczynski, Arbres de Markov couple, C. R. Acad. Sci. Paris, Ser. I 335 (2002) 79–82.
[12] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, 1976.
[13] P. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, Internat. J. Approximate Reasoning 9 (1993) 1–35.