HAL open science

Quantification de l'incertitude sur la structure latente dans des modèles de Markov cachés

Jean-Baptiste Durand, Yann Guédon

To cite this version:

Jean-Baptiste Durand, Yann Guédon. Quantification de l'incertitude sur la structure latente dans des modèles de Markov cachés. 46èmes Journées de Statistique, Société Française de Statistique (SFdS). FRA., Jun 2014, Rennes, France. hal-01058317

HAL Id: hal-01058317

https://inria.hal.science/hal-01058317

Submitted on 26 Aug 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Quantification de l'incertitude sur la structure latente dans des modèles de Markov cachés

Jean-Baptiste Durand [1] & Yann Guédon [2]

[1] *Univ. Grenoble Alpes, Laboratoire Jean Kuntzmann et Inria, Mistis*
51 rue des Mathématiques
B.P. 53, F-38041 Grenoble cedex 9, France ; jean-baptiste.durand@imag.fr
[2] *CIRAD, UMR AGAP et Inria, Virtual Plants*
F-34095 Montpellier, France ; guedon@cirad.fr

Résumé. Nous introduisons les modèles de Markov cachés graphiques, qui généralisent les chaînes et arbres de Markov cachés (CMCs et AMCs). Nous montrons comment l'incertitude globale sur le processus d'état caché peut être décomposée en une somme d'entropies conditionnelles, qui s'interprètent comme une contribution locale à l'incertitude globale. Nous donnons un algorithme efficace de calcul de ces entropies pour les CMCs et AMCs et montrons leur apport, en complément d'autres algorithmes de restauration des états, au diagnostic et à l'interprétation des états cachés. Nous montrons également que les profils classiques de probabilités lissées (loi marginale de l'état caché à chaque instant, sachant l'ensemble des observations), ne permet pas de conclure sur la contribution locale à l'incertitude globale.

Mots-clés. Modèles de Markov cachés, inférence des états, entropie conditionnelle.

1 Introduction

Les modèles de Markov cachés ont été fréquemment utilisés en analyse de séquences pour modéliser divers types de structure latente, telles que des zones homogènes ou des motifs bruités (Ephraim & Mehrav, 2002). Ils ont été étendus des séquences à des structures graphiques plus générales, et notamment des arborescences. On peut distinguer l'inférence sur les paramètres du modèle de celle sur le processus d'état à paramètres fixés ; c'est cette dernière qui est l'objet de ce résumé.

L'inférence du processus d'état est une étape clé de l'analyse quand les états ont une interprétation forte, ou pour le calcul de certaines fonctions des états (notamment à des fins de validation du modèle, par exemple pour superposer des histogrammes à des densités conditionnelles aux états). Il s'agit alors de restaurer ce processus d'état, ce qui n'a de sens que si l'incertitude sur sa valeur est modérée.

Pour quantifier cette incertitude, il n'est pas suffisant de considérer l'incertitude globale du processus : il est également essentiel de savoir comment cette incertitude globale se décompose localement le long de la structure.

Nous considérons ici un modèle de Markov caché dont les observations $\boldsymbol{X} = (X_v)_{v\in\mathcal{V}}$ sont indexées par un graphe orienté sans circuit (GOSC) fixé $\mathcal{G}$ de sommets $\mathcal{V}$ et d'arcs $\mathcal{A}$. On parle de modèle de Markov caché graphique (MMCG). Cette famille de modèles contient en particulier les chaînes et arbres de Markov cachés (CMCs et AMCs). Soit $\boldsymbol{S} = (S_v)_{v\in\mathcal{V}}$ le processus d'état caché associé, S_v étant à valeurs dans $\{0, \ldots, K-1\}$. Soit $\boldsymbol{x}$ une réalisation possible de $\boldsymbol{X}$: on note $\mathrm{pa}(v)$ le parent d'un sommet v, et pour un sous-ensemble U de $\mathcal{V}$, on note X_U (resp. $\boldsymbol{x}_U$) les variables aléatoires $(X_u)_{u\in U}$ (resp. les observations $(x_u)_{u\in U}$). Nous faisons les hypothèses que : $\boldsymbol{S}$ vérifie la propriété de factorisation markovienne associée au GOSC $\mathcal{G}$, en assimilant l'ensemble des sommets $\mathcal{V}$ à l'ensemble des variables aléatoires $(S_v)_{v\in\mathcal{V}}$ (Lauritzen, 1996) ; la loi de $\boldsymbol{S}$ est paramétrée par les probabilités de transition $p_{\boldsymbol{s}_{pa(v)},k} = P(S_v = k|\boldsymbol{S}_{pa(v)} = \boldsymbol{s}_{pa(v)})$ et, pour les sommets sources (sans parent) u de $\mathcal{G}$, par les probabilités initiales $(P(S_u = k))_k$; sachant $\boldsymbol{S}$, les $(X_v)_v$ sont indépendantes, et X_u est indépendante des $(S_v)_{v\neq u}$.

Classiquement, la quantification de l'incertitude sur la valeur des états repose sur les profils de probabilités lissées $P(S_v = k|\boldsymbol{X} = \boldsymbol{x})_{v\in\mathcal{V}}$ pour $k = 0, \ldots, K-1$. Cette approche comporte deux inconvénients majeurs : comme nous le montrons par la suite, l'incertitude sur les états associée à ces profils conduit à une perception surestimant l'incertitude globale sur $\boldsymbol{S}$ sachant $\boldsymbol{X} = \boldsymbol{x}$; de plus la visualisation de ces profils est problématique pour des graphes $\mathcal{G}$ quelconques, hormis le cas $K = 2$. Notre approche considère l'entropie H comme la mesure canonique de l'incertitude ; ainsi $H(\boldsymbol{S}|\boldsymbol{X} = \boldsymbol{x})$ quantifie l'incertitude sur les états cachés sachant les observations. Notre approche vise à décomposer cette quantité comme une somme d'entropies dont chaque terme correspond à un sommet de $\mathcal{V}$, et qui peut donc s'interpréter comme une contribution locale à l'incertitude globale. Du fait du caractère univarié de ces profils, ils peuvent être visualisés sur des graphes $\mathcal{G}$ quelconques.

Nous explicitons cette décomposition, puis donnons un algorithme de calcul efficace des termes de la décomposition dans le cas des AMCs et CMCs. Nous montrons à l'aide d'exemples issus de données réelles et synthétiques comment les profils d'entropie locale ainsi obtenus permettent le diagnostic de l'incertitude sur les états et leur interprétation, en complément de l'algorithme d'énumération des L restaurations les plus probables du processus d'état (algorithme de Viterbi généralisé), et de l'algorithme de calcul de profils d'alternatives locales à la restauration la plus probable (algorithme de Viterbi avant–arrière) qui résoud le problème de maximisation suivant

$$(\arg) \max_{(s_u)_{u\neq v}} P((S_v = s_v)_{u\neq v}, S_v = k|\boldsymbol{X} = \boldsymbol{x})$$

Nous montrons également en quoi les profils usuels de probabilités lissées ne permettent pas de quantifier l'incertitude globale sur les états, à cause de la marginalisation opérée.

2 Profils d'entropies conditionnelles

Soit $\boldsymbol{X}$ un MMCG tel que défini en partie 1, associé au processus d'états $\boldsymbol{S}$; $\boldsymbol{S}$ vérifie la factorisation associée à la propriété de Markov sur $\mathcal{G}$:

$$\forall \boldsymbol{s},\ P(\boldsymbol{S}=\boldsymbol{s}) = \prod_{v\in\mathcal{V}} P(S_v = s_v | \boldsymbol{S}_{pa(v)} = \boldsymbol{s}_{pa(v)}). \tag{1}$$

La décomposition de l'entropie $H(\boldsymbol{S}|\boldsymbol{X}=\boldsymbol{x})$ résulte du fait que conditionnellement à $\boldsymbol{X}=\boldsymbol{x}$, $\boldsymbol{S}$ vérifie également la propriété de factorisation sur $\mathcal{G}$:

$$P(\boldsymbol{S}=\boldsymbol{s}|\boldsymbol{X}=\boldsymbol{x}) = \prod_{v} P(S_v = s_v | \boldsymbol{S}_{\mathrm{pa}(v)} = \boldsymbol{s}_{pa(v)}, \boldsymbol{X}=\boldsymbol{x}),$$

où $P(S_v = s_v|\boldsymbol{S}_{pa(v)} = \boldsymbol{s}_{pa(v)}, \boldsymbol{X}=\boldsymbol{x})$ désigne $P(S_s = s_s|\boldsymbol{X}=\boldsymbol{x})$ si $\mathrm{pa}(v)=\emptyset$.

Démonstration. Cette propriété est montrée (de même que (1)) par récurrence sur les sommets de $\mathcal{G}$. Les variables $(\boldsymbol{S},\boldsymbol{X})$ suivent la propriété de Markov sur le graphe $\mathcal{G}'$ dont l'ensemble des arcs $\mathcal{A}'$ est défini par $a \in \mathcal{A}' \Leftrightarrow \{[a=(S_u,S_v) \text{ et } (u\in \mathrm{pa}(v))] \text{ ou } a = (S_u, X_u)\}$. On considère un sommet puits u de $\mathcal{G}$ (sans enfants) et en remarquant que S_u est séparé des $(S_v)_{v\neq u, v\notin pa(u)}$ par $\boldsymbol{S}_{pa(u)}$ dans le graphe moral de $\mathcal{G}'$, on obtient la factorisation

$$P(\boldsymbol{S}=\boldsymbol{s}|\boldsymbol{X}=\boldsymbol{x}) = P(S_u = s_u|\boldsymbol{S}_{pa(u)} = \boldsymbol{s}_{pa(u)}, \boldsymbol{X}=\boldsymbol{x}) P((S_v)_{v\neq u} = (s_v)_{v\neq u}|\boldsymbol{X}=\boldsymbol{x}).$$

□

En appliquant la règle de chaînage pour l'entropie (Cover & Thomas, 2006, chap. 2) on obtient alors la décomposition

$$H(\boldsymbol{S}|\boldsymbol{X}=\boldsymbol{x}) = \sum_{v} H(S_v|\boldsymbol{S}_{pa(v)}, \boldsymbol{X}=\boldsymbol{x}),$$

avec la même convention que ci-dessus si $\mathrm{pa}(v)=\emptyset$. Par conséquent, l'entropie globale du processus d'état se décompose comme une somme de termes correspondant au profil d'entropies conditionnelles $(H(S_v|\boldsymbol{S}_{pa(v)}, \boldsymbol{X}=\boldsymbol{x}))_{v\in\mathcal{V}}$. Ainsi, chaque terme s'interprète comme une incertitude locale qui contribue à l'incertitude globale de manière additive.

A contrario, les entropies marginales $(H(S_v|\boldsymbol{X}=\boldsymbol{x}))_{v\in\mathcal{V}}$ quantifient l'incertitude sur les probabilités lissées $\xi_v(k) = P(S_v = k|\boldsymbol{X}=\boldsymbol{x})$ pour $v\in\mathcal{V}$ et $0\leq k < K$. D'après Cover & Thomas (2006), chap. 2, ces entropies marginales majorent les entropies conditionnelles, si bien que

$$H(\boldsymbol{S}|\boldsymbol{X}=\boldsymbol{x}) \leq \sum_{v} H(S_v|\boldsymbol{X}=\boldsymbol{x}),$$

et que les profils de probabilités lissées ne représentent pas l'incertitude sur la valeur de $\boldsymbol{S}$.

Nous montrons ci-dessous que dans le cas des AMCs, le calcul des $H(S_v|\boldsymbol{S}_{pa(v)}, \boldsymbol{X} = \boldsymbol{x})$ peut se faire en intégrant une étape de complexité $\mathcal{O}(K^2)$ à l'algorithme usuel ascendant-descendant pour le calcul les probabilités lissées. Nous considérons un graphe $\mathcal{G} = \mathcal{T}$ arborescent, de sommets $\mathcal{V} = \{0, \ldots, n-1\}$, enraciné en $u = 0$. En supposant le modèle homogène, les probabilités de transition entre états se résument à la matrice $(p_{jk})_{j,k} = (P(S_v = k|S_{pa(v)} = j))_{j,k}$.

L'algorithme ascendant-descendant vise à calculer les probabilités lissées $(\xi_v(k))_{v\in\mathcal{V};0\leq k<K}$. Une version itérative numériquement stable a été proposée par Durand *et al.* (2004). Elle repose sur une récursion ascendante, initialisée en les feuilles de $\mathcal{T}$ et qui calcule les quantités $\beta_v(k) = P(S_v = k|\bar{\boldsymbol{X}}_v = \bar{\boldsymbol{x}}_v)$ et $\beta_{pa(v),v}(k) = P(\bar{\boldsymbol{X}}_v = \bar{\boldsymbol{x}}_v|S_{pa(v)} = k)/P(\bar{\boldsymbol{X}}_v = \bar{\boldsymbol{x}}_v)$, où $\bar{\boldsymbol{X}}_v$ désigne la sous-arborescence enracinée en v. Ces quantités sont calculées en fonction des β_u et $\beta_{pa(u),u}$ pour les enfants u de v. La complexité est en $\mathcal{O}(K^2)$ par itération. Les probabilités lissées sont calculées par une récursion descendante, initialisée en la racine de $\mathcal{T}$ et qui calcule $\xi_v(k)$ en fonction des $\xi_{pa(v)}$, des β_v et des $\beta_{pa(v),v}$. La complexité est aussi en $\mathcal{O}(K^2)$ par itération. En ajoutant le calcul de

$$H(S_v|S_{\rho(v)}, \boldsymbol{X} = \boldsymbol{x}) = -\sum_{i,j} P(S_v = j, S_{pa(v)} = i|\boldsymbol{X} = \boldsymbol{x}) \log P(S_v = j|S_{pa(v)} = i, \boldsymbol{X} = \boldsymbol{x})$$

à la récursion descendante, avec

$$P(S_v = j|S_{pa(v)} = i, \boldsymbol{X} = \boldsymbol{x}) = \beta_v(j)p_{ij}/\{P(S_v = j)\beta_{\rho(v),v}(i)\},$$

on y incorpore le calcul des entropies conditionnelles en lui conservant une complexité en $\mathcal{O}(K^2)$ par itération.

Ce résultat s'applique en particulier aux CMCs. On obtient alors

$$H(\boldsymbol{S}|\boldsymbol{X} = \boldsymbol{x}) = H(S_0|\boldsymbol{X} = \boldsymbol{x}) + \sum_{t=1}^{n-1} H(S_t|S_{t-1}, \boldsymbol{X} = \boldsymbol{x}),$$

avec

$$H(S_t|S_{t-1}, \boldsymbol{X} = \boldsymbol{x}) = -\sum_{i,j} P(S_t = j, S_{t-1} = i|\boldsymbol{X} = \boldsymbol{x}) \log P(S_t = j|S_{t-1} = i, \boldsymbol{X} = \boldsymbol{x}).$$

On obtient alors une alternative à l'algorithme de Hernando *et al.* (2005) pour les CMCs, qui permet de calculer $H(\boldsymbol{S}|\boldsymbol{X} = \boldsymbol{x})$ avec une même complexité en $\mathcal{O}(nK^2)$ mais avec l'avantage de fournir le profil d'entropies conditionnelles $H(S_t|S_{t-1}, \boldsymbol{X} = \boldsymbol{x})$ pour $0 \leq t < n$.

3 Applications

3.1 Données synthétiques

On considère une famille de CMCs à deux états cachés paramétrée par $\varepsilon = P(S_t = 1|S_{t-1} = 0) = P(S_t = 0|S_{t-1} = 1)$, $\varepsilon \in [0, 0.5]$, avec pour loi initiale $P(S_0 = 0) = 0.5$ Les observations sont à valeurs dans $\{0, 1, 2\}$ et $P(X_t = 0|S_t = 0) = P(X_t = 1|S_t = 0) = 0,5$;

$P(X_t = 1|S_t = 1) = P(X_t = 2|S_t = 1) = 0,5$. On considère la séquence observée $x_t = 1$ pour $t = 0, \ldots, n-1$. Pour $t = 0, \ldots, T-1$, les probabilités lissées sont $\xi_t(0) = \xi_t(1) = 0.5$. Ainsi, pour tout ε, l'entropie marginale est $\log 2$, et la somme de ces entropies $n \log 2$. L'entropie globale de la séquence d'états cachés est au contraire une fonction strictement croissante de ε, minimale et égale à $\log 2$ pour $\varepsilon = 0$; maximale, égale à $n \log 2$ pour $\varepsilon = 0.5$.

Les profils d'entropies marginale et conditionnelle sont présentés figure 3.2. Pour $\varepsilon = 0$, le profil d'entropies conditionnelles traduit le fait que l'incertitude globale est $\log 2$, correspondant à l'incertitude sur l'état initial, et qu'une fois cet état fixé, les observations suivantes sont déterministes et ne contribuent pas à l'incertitude globale. Le profil d'entropies marginales traduit le fait qu'à chaque instant, les deux états sont équiprobables sachant les observations ; on obtiendrait le même résultat sous une hypothèse de mélange indépendant pour X_t, ce qui reflète la prise en compte partielle de la dépendance markovienne par ce type de profil. De manière générale, l'entropie marginale résulte de l'incertitude sur l'état S_t, due à l'observation X_t, mais aussi de la propagation de l'incertitude sur les états précédents. Elle n'a donc pas d'interprétation possible en termes de contribution locale à l'incertitude globale. Au contraire, via le conditionnement, l'entropie conditionnelle supprime la part liée à cette propagation de l'incertitude.

3.2 Structure de branches de pin d'Alep

Cette étude vise à proposer un modèle du développement architectural du pin d'Alep. Les données comportent 7 branches de pin d'Alep, provenant d'invididus différents, et décrites à l'échelle des pousses annuelles v (portions mises en place au cours d'une année). Chaque branche est assimilée à une arborescence caractérisée, au sommet (pousse) v, par 5 variables composant l'observation X_v : nombre de cycles de croissance de la pousse pendant son année de mise en place (de 1 à 3), présence d'organes reproducteurs mâles (binaire), présence d'organes femelles (binaire), longueur en cm et nombre de branches portées. Les paramètres du modèle sont estimés par maximum de vraisemblance par l'algorithme EM et le nombre d'états est choisi par une procédure prenant en compte l'entropie des états sachant les observations (voir détails dans Durand & Guédon, 2012).

Dans un premier temps, les sites où l'entropie conditionnelle est la plus forte sont visualisés à l'échelle de niveaux de couleur (Figure 3.2). Cette étape met en évidence la localisation des états les moins ambigus le long des axes principaux de la branche. Puis des profils d'états sont visualisés le long de chemins allant de la racine aux feuilles de l'arborescence, en passant par des sommets d'entropie conditionnelle élevée. Ces profils proviennent d'une part de l'algorithme de Viterbi avant–arrière, qui permet d'identifier les alternatives les plus probables localement à l'arborescence d'états la plus probable. Cette approche est complétée par l'algorithme de Viterbi généralisé qui permet de voir comment des sommets voisins changent simultanément d'état dans ces configurations alternatives. Ces résultats mettent en évidence que les chemins les plus ambigus sont

constitués de succession de pousses non ramifiées, comportant un seul cycle de croissance, et sans organes mâle ni femelle.

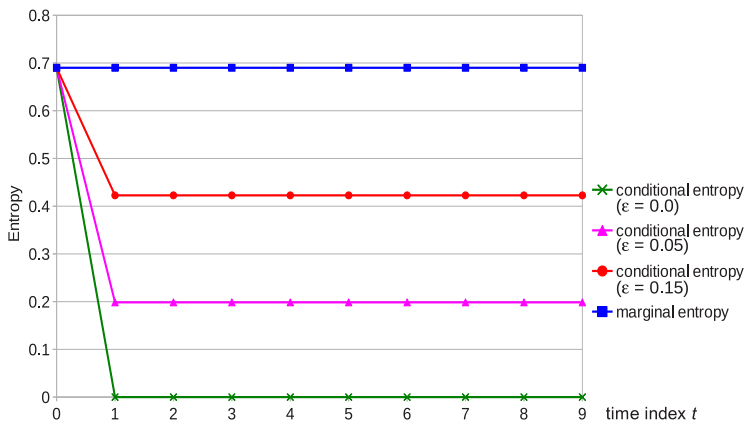

FIG 3.2 *Profils d'entropie marginale et conditionnelle pour une CMC à deux états avec probabilités de transition* $\varepsilon = 0.0$, $\varepsilon = 0.05$ *et* $\varepsilon = 0.15$.

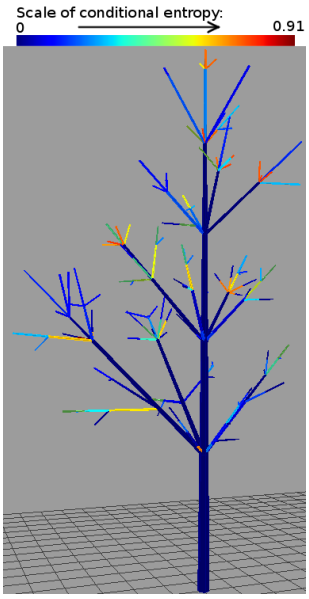

FIG 3.2 *Entropie conditionnelle* $H(S_u|S_{pa(u)}, \boldsymbol{X} = \boldsymbol{x})$ *pour les sommets* u *associés à une branche de pin d'Alep donnée. Les sommets bleus ont l'entropie la plus faible, les rouges la plus élevée.*

Références

[1] COVER, T., AND THOMAS, J. *Elements of Information Theory, 2nd edition.* Hoboken, NJ : Wiley, 2006.

[2] DURAND, J.-B., GONÇALVÈS, P., AND GUÉDON, Y. Computational Methods for Hidden Markov Tree Models – An Application to Wavelet Trees. *IEEE Transactions on Signal Processing 52*, 9 (Sept. 2004), 2551–2560.

[3] DURAND, J.-B., AND GUÉDON, Y. Localizing the Latent Structure Canonical Uncertainty : Entropy Profiles for Hidden Markov Models. Inria research report 7896, Inria, Feb. 2012.

[4] EPHRAIM, Y., AND MERHAV, N. Hidden Markov processes. *IEEE Transactions on Information Theory 48* (June 2002), 1518–1569.

[5] HERNANDO, D., CRESPI, V., AND CYBENKO, G. Efficient computation of the hidden Markov model entropy for a given observation sequence. *IEEE Transactions on Information Theory 51*, 7 (2005), 2681–2685.

[6] LAURITZEN, S. *Graphical Models.* Clarendon Press, Oxford, United Kingdom, 1996.