



# Réseau Bayésien Aplaté pour l'Inférence dans les HMM hiérarchiques factorisés et Apprentissage avec peu de données

Sylvain Gelly, Nicolas Bredeche, Michèle Sebag

► **To cite this version:**

Sylvain Gelly, Nicolas Bredeche, Michèle Sebag. Réseau Bayésien Aplaté pour l'Inférence dans les HMM hiérarchiques factorisés et Apprentissage avec peu de données. Conférence d'Apprentissage, Jun 2005, Nice, France. pp.143-144, 2005. <inria-00175528>

**HAL Id: inria-00175528**

**<https://hal.inria.fr/inria-00175528>**

Submitted on 28 Sep 2007

**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.

# Réseau Bayésien Aplati pour l'Inférence dans les HMM hiérarchiques factorisés et Apprentissage avec peu de données<sup>1</sup>

Sylvain Gelly\*, Nicolas Bredeche\*, Michèle Sebag\*

\*Équipe Inférence&Apprentissage - Projet TAO (INRIA futurs)  
LRI - Université Paris-Sud, 91405 Orsay Cedex  
(gelly,bredeche,sebag)@lri.fr - <http://tao.lri.fr>

Une limite essentielle des HMM, et plus généralement des modèles de Markov, concerne le passage à l'échelle, l'impossibilité de la prise en compte efficace de l'influence de phénomènes indépendants et la difficulté de généralisation.

Pour répondre à ces problèmes, plusieurs extensions existent. En particulier, nous nous intéresserons dans ce qui suit à la *hiérarchisation* (Shai Fine, 1998; Theodorou *et al.*, 2004) et à la *factorisation* (Ghahramani & Jordan, 1997).

La hiérarchisation permet de réduire le nombre de liens entre états nécessaires dans un HMM et par là même de réduire la complexité algorithmique de l'apprentissage ainsi que l'imprécision. Quant à la factorisation, le principe est d'expliquer les observations par plusieurs causes plutôt qu'une seule. C'est à dire qu'on remplace le  $P(Y|X)$  des HMM par  $P(Y|X^1, X^2, \dots, X^n)$ . Les  $X^i$  sont des variables cachées pouvant être gérées indépendamment. Les  $P(X_{t+1}^i | X_t^i)$  sont alors différents pour chaque  $i$ .

En pratique, on ne peut pas adapter directement les algorithmes d'inférence existants dans le cas des HMM factorisés, ou hiérarchiques. De plus, un aspect important du problème est que notre système apprend à partir de données éparses car nous faisons l'hypothèse que nous ne disposons que d'un petit nombre d'exemples pour apprendre. Ceci se justifie par notre domaine d'application (la robotique située), où le processus d'échantillonnage des données est contrôlé par un comportement dépendant entre autres de l'environnement et des capacités du robot qui ne permet pas d'obtenir beaucoup d'exemples. Par conséquent, nous souhaitons exprimer un compromis entre précision et vitesse de l'apprentissage.

L'approche que nous proposons consiste à changer de formalisme de représentation en transformant un graphe orienté et avec circuits (i.e. HMM hiérarchique factorisé dans lequel on fait abstraction du typage des dépendances) en un réseau Bayésien. Le formalisme des réseaux Bayésiens s'inscrit en effet dans un cadre théorique développé et bien connu, ce qui laisse espérer une résolution plus facile. Toutefois, nous identifions les deux problèmes suivants :

- L'existence de dépendances multiples dans les FHHMM entraîne à priori une explosion combinatoire du nombre de paramètres à apprendre, ce qui est d'autant plus problématique lorsque peu d'exemples sont à notre disposition ;
- La présence de circuits dans les dépendances conditionnelles entre les variables d'un FHHMM empêchent la modélisation directe par un réseau Bayésien. Il est à noter que ces dépendances ne concernent les variables qu'à un même pas de temps (synchrones).

---

<sup>1</sup>Acknowledgements : This work was supported in part by the PASCAL Network of Excellence.

FIG. 1 – Exemple de changement de représentation (RB  $\Rightarrow$  RBA). Le graphe de gauche donne la structure des dépendances conditionnelles entre les variables, et le graphe de droite montre l'expression dans le formalisme des réseaux Bayésien après transformation et ajout de variables additionnelles booléennes. Ces variables additionnelles permettent de forcer la dépendance entre deux variables du réseau Bayésien (elles sont systématiquement observées à *vrai*). Les probabilités conditionnelles de ces variables sont calculées de telle sorte que les probabilités jointes entre deux variables d'origine soient cohérentes avec les données.

En conséquence, nous avons mis au point un algorithme de changement de représentation permettant de construire ce que nous appelons des réseaux Bayésiens aplatis (RBA). Cette modélisation est basée sur un compromis entre précision et vitesse d'apprentissage et repose sur la prise en compte des dépendances multiples en les exprimant deux à deux seulement (cf. Fig. 1).

Les résultats obtenus sont prometteurs puisqu'ils montrent que les réseaux Bayésiens aplatis ont les propriétés suivantes :

- prise en compte et modélisation des circuits, ceux-ci étant fréquents dans les HMM considérés ;
- apprentissage plus rapide avec peu d'exemples, au prix, il est vrai, d'une perte de précision à long terme. En robotique, ce compromis est avantageux puisque l'on dispose souvent d'exemples peu nombreux.

Cependant, le formalisme final étant toujours celui des RB (avec des variables additionnelles et un calcul des paramètres satisfaisant des axiomes), nous pouvons très bien envisager des représentations "hybrides" dans lesquelles certaines dépendances (pour lesquelles peu de données sont disponibles) sont exprimées dans le cadre des RBA, tandis que d'autres sont exprimées de façon classique. Ainsi, cette méthode permettrait de tirer parti du compromis expressivité/apprentissage avec peu d'exemples des RBA tout en gardant une expressivité pouvant atteindre celle des RB lorsque le nombre d'exemples disponibles croît suffisamment.

## Références

- GHAHRAMANI Z. & JORDAN M. I. (1997). Factorial hidden markov models. *Machine Learning*, vol. 29. 1996, pages 245-273.
- SHAI FINE, YORAM SINGER N. T. (1998). The hierarchical hidden markov model : Analysis and applications. *Machine Learning*.
- THEOCHAROUS G., MURPHY K. & KAEHLING L. (2004). Representing hierarchical pomdps as dbns for multi-scale robot localization. *Proc. of the IEEE international Conference on Robotics and Automation (ICRA'04)*.