

# Représentation et reconnaissance de caractères manuscrits par Réseaux Bayésiens Dynamiques

Laurence Likforman-Sulem, Marc Sigelle

GET-ENST/ Traitement du Signal et des Images et CNRS-LTCI (UMR 5141)  
46, rue Barrault, 75013 Paris  
{likforman|sigelle}@tsi.enst.fr

## 1 Introduction

Les approches stochastiques, tels que les modèles de Markov cachés (HMM), sont largement utilisées pour la reconnaissance de la parole et de l'écrit (Elms et al. 1998 ; Hallouli et al. 2002) pour leur capacité à s'adapter aux distorsions élastiques temporelles et spatiales. Cependant ces modèles sont mono-dimensionnels. Une adaptation doit donc être réalisée pour les images, par nature bi-dimensionnelles : celles ci sont converties en séquences 1D d'observations le long d'une direction. Une séquence admissible d'observations est par exemple la suite des colonnes de pixels en balayant l'image de gauche à droite. D'autres séquences sont possibles : vecteurs de caractéristiques sur des fenêtres glissantes, lignes de texte...

Les HMM font l'hypothèse que les observations sont indépendantes conditionnellement aux états cachés, ce qui n'est pas toujours réaliste pour les images. Des extensions des HMM permettant de mieux prendre en compte l'aspect bi-dimensionnel des images ont ainsi été proposées avec les modèles pseudo-2D (ou planar HMM) (Gilloux 1994). Plus récemment, des modèles 2D à base de champs de Markov ont été développés (Park et Lee 1998 ; Saon et Belaid 1999 ; Chevalier et al. 2003). En faisant apparaître les dépendances entre variables d'états ou observations, une modélisation plus fine de phénomènes peut être obtenue. Dans cette optique, des modèles probabilistes s'appuyant sur les réseaux bayésiens statiques sont apparus dans le domaine de la reconnaissance de l'écriture en-ligne (Cho et Kim 2003), l'analyse de documents (Souafi 2002) et l'authentification de signatures (Xiao et Leedham, 2002). Les réseaux bayésiens dynamiques sont une extension des réseaux statiques qui prennent en compte des séquences variables d'observations. On note  $\xi_t$ , l'ensemble des variables d'états et d'observations au temps  $t$ . Un réseau bayésien dynamique à deux pas de temps (2TBN) est défini par

- un réseau initial  $B_1$  qui spécifie la distribution initiale des états et les distributions conditionnelles des états et des observations à  $t=1$
- un réseau de transition  $B_t$  qui spécifie les distributions  $P(\xi_{t+1} | \xi_t)$ . Ces distributions sont supposées stationnaires, i.e. indépendantes de  $t$ .

Cette étude expérimente des modèles simples mono-flux de type HMM et des modèles couplés. Les structures couplées sont toutes construites par la mise en correspondance de deux réseaux simples mono-flux (ajouts de liens dans la structure graphique). Dans notre application, les états cachés sont des variables discrètes et les observations sont continues. Les observations sont soit les lignes, soit les colonnes normalisées de pixels d'un caractère, obtenues par balayage séquentiel, soit les deux à la fois. L'évaluation de ces modèles a été réalisée sur la base de chiffres MNIST (LeCun 1998).

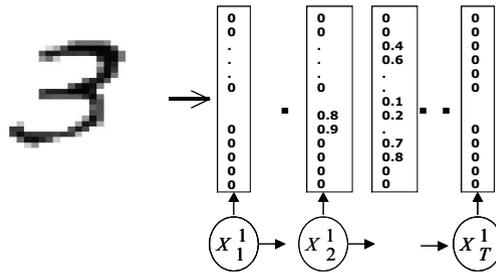


FIG. 1 – Séquence d’observations colonnes issues d’une image de caractère. HMM vertical observant ces colonnes et représenté sous forme de Réseau Bayésien Dynamique.

## 2 Modèles simples et couplés

Les modèles simples et couplés sont définis par les probabilités initiales des états, les CPD (*Conditional Probability Distributions*) et les CPT (*Conditional Probability Tables*). Les CPD modélisent la distribution des observations conditionnellement aux états et sont supposées gaussiennes. Les probabilités de transitions entre états définissent les CPT. La boîte à outils BayesNet (Murphy, 2003) permet l’apprentissage et l’inférence de ces modèles. L’inférence utilise l’algorithme de l’arbre de jonction et l’apprentissage des paramètres du réseau utilise l’algorithme EM.

Les modèles simples HMM sont des cas particuliers de RBD dans lesquels une variable (ou vecteur) observé a pour parent une seule variable d’état discrète. Le HMM vertical (resp. horizontal) a pour séquence d’entrée les colonnes (resp. les lignes) de pixels du caractère. Les matrices de transition entre états sont de type gauche-droite pour les modèles mono-flux.

Les modèles couplés associent les modèles simples mono-flux. Les modèles sont créés sur 2 pas de temps en ajoutant des liens orientés entre les deux structures de réseau bayésien de type HMM. Des liens sont créés entre états et observations des 2 structures, issus du même pas de temps et/ou de pas de temps consécutifs. Nous présentons 3 modèles couplés en Fig. 2. Les deux premiers modèles (ST\_CPL1 et ST\_CPL2) couplent les états des deux chaînes : les liens ajoutés vont dans le sens de la chaîne verticale vers la chaîne horizontale. Ceci donne plus de poids à la chaîne verticale qui contrôle ainsi la chaîne horizontale. En effet, des expériences précédentes (Hallouli et al. 2002) ont montré que le HMM vertical était plus performant que le HMM horizontal pour les caractères latins et les chiffres car la direction dominante des traits est celle de la verticale. Le troisième modèle est un modèle général qui couple, de plus, les états de la chaîne horizontale aux observations de la chaîne verticale. Dans ces 3 modèles, on observe conjointement la colonne et la ligne de même indice à un instant donné, ce qui correspond à un parcours de l’image selon la diagonale secondaire. Ce parcours semble a priori assez pertinent.

Pour limiter le nombre de paramètres et synchroniser les deux chaînes, les CPT ont été contraintes à être de type gauche-droite pour les transitions d’états provenant du même flux, et stochastiques pour les transitions d’états émanant de flux différents.

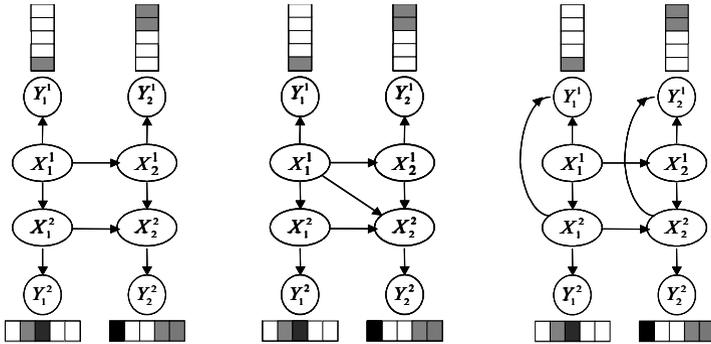


FIG. 2 – De gauche à droite : modèles couplés *ST\_CPL1*, *ST\_CPL2* et *GN\_CPL* sur deux pas de temps.

### 3 Expériences et conclusion

La base MNIST contient des chiffres manuscrits isolés issus d'une campagne de recensement américaine. Nous n'avons utilisé qu'une partie de la base d'apprentissage disponible (60 000 chiffres). 4000 chiffres ont été utilisés pour l'apprentissage (base d'entraînement) et 1000 pour la validation. Celle-ci sert à régler certains paramètres comme le nombre d'états optimal pour chaque modèle (entre 10 et 14 états). La base de test contient 10 000 chiffres. Les images ont été préalablement filtrées (filtrage gaussien) et les valeurs obtenues ont été normalisées entre 0 et 1. On construit un modèle de chiffre par classe. Lors de la reconnaissance, la classe estimée est celle obtenue au maximum de vraisemblance par rapport au modèle.

| HMM vertical | HMM horizontal | ST-CPL1     | ST-CPL2     | GN-CPL    |
|--------------|----------------|-------------|-------------|-----------|
| 90.2         | 87.4           | 92.4        | 92.2        | 91.49     |
| [89.6 90.8]  | [86.7 88]      | [91.8 92.9] | [91.6 92.7] | [90.9 92] |

TAB 1- Résultats expérimentaux des 2 modèles simples et des 3 modèles couplés.

Les modèles couplés sont plus performants que les modèles simples. Les corrélations entre lignes et colonnes de caractères sont en effet mieux prises en compte. La Table 1 donne les taux de reconnaissance pour les modèles simples et couplés, ainsi que les intervalles de confiance à 95% en supposant que le taux de reconnaissance suit une loi de Bernoulli. Bien que le modèle HMM-horizontale simple soit moins performant que le modèle vertical, son association avec celui-ci offre une meilleure représentation des caractères, et ce pour tous les modèles couplés (*ST\_CPL1* et 2, *GN\_CPL*). Le modèle général *GN\_CPL* couplant les observations d'une chaîne avec les états de l'autre chaîne donne encore plus d'importance aux observations de type 'colonnes'. De même que pour les HMM classiques, les paramètres liés aux observations (CPD) sont prépondérants par rapport aux matrices de transitions entre états.

Les méthodes d'apprentissage discriminatif (réseaux neuronaux, SVM) sont connues pour leurs remarquables performances sur la reconnaissance de caractères isolés (entre 1.4 et 1.9 % d'erreur sur la base MNIST) alors que l'apprentissage par réseaux bayésiens est génératif.

L'objectif n'est pas de présenter ici un résultat dépassant l'état de l'art sur la base MNIST mais d'introduire une nouvelle modélisation d'images de caractères, et de montrer l'apport du couplage. Nous avons cependant comparé ces modèles à la méthode standard des k-ppv (k=1) pour laquelle nous avons obtenu 92.6 % de bonne reconnaissance en utilisant la même base d'entraînement. Un résultat similaire est obtenu avec le modèle couplé ST-CPL1.

Les structures présentées ont été obtenues de manière empirique. D'autres structures sont possibles mais le nombre de paramètres augmente rapidement avec le nombre de variables et le nombre de liens. Des algorithmes de recherche de structure optimale peuvent aussi être envisagés. L'initialisation des paramètres avant apprentissage est très importante car l'algorithme EM trouve pour ces paramètres un optimum local de la vraisemblance. Le choix d'une initialisation convenable est encore en suspens et sa résolution devrait entraîner des améliorations importantes. L'implémentation de BayesNet en MATLAB rend l'apprentissage des modèles et la reconnaissance très lents (plusieurs heures pour chaque modèle de caractère). Des progrès sont attendus sur ce point avec la transposition de BayesNet en langage C, et l'utilisation d'algorithmes d'inférence optimisés.

**Remerciements** : nous tenons à remercier Chafic Mokbel (Professeur à l'Université de Balamand, Liban) pour les fructueuses discussions ainsi que pour ses conseils avisés d'implémentation.

## Références

- Chevalier S., Geoffrois E., Preteux F. (2003). A 2D Dynamic Programming Approach for Markov Random Field-based Handwritten Character Recognition, *ICIP 03*, vol. 2, pp 616-629.
- Cho S. J., Kim J. H. (2003). Bayesian Network Modeling of Hangul characters for on-line handwriting recognition, *7th ICDAR*, pp 207-211.
- Elms A.J., Procter S., Illingworth J. (1998). The advantage of using an HMM based approach for faxed word recognition. *IJDAR*, 1, pp 18-36.
- Gilloux M. (1994), Reconnaissance de chiffres manuscrits par modèles de Markov pseudo-2D, *Conférence Nationale sur l'Écrit et le Document CNED*, pp. 11-17.
- Hallouli K., Likforman-Sulem L., Sigelle M. (2002), A comparative study between decision fusion and data fusion in Markovian printed character recognition, *ICPR 02*, pp. 147-150.
- LeCun Y. (1998). The MNIST handwritten digit database: <http://www.research.att.com/yann/ocr/mnist>.
- Murphy K. (2002), *Dynamic Bayesian Networks : Representation, Inference and Learning*, Ph.D. thesis, University of California, Berkeley.
- Murphy K. (2003). BayesNet Toolbox for Matlab. <http://www.ai.mit.edu/murphyk/Bayes/bnintro.html>.
- Park H., Lee S. (1998) A truly 2-D Hidden Markov Model for Off-Line Handwritten Character Recognition. *Pattern Recognition*, 31, pp 1849-1864.
- Pearl J. (1988), *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference*, Morgan Kaufman.
- Saon G., Belaid A. (1999). Off-line handwritten word recognition using a mixed HMM-MRF approach. *4th ICDAR*, pp 118-122.
- Souafi S. (2002), *Contribution à la reconnaissance des structures de documents écrits : approche probabiliste*, Thèse de Doctorat, INSA de Lyon.
- Xiao X., Leedham G. (2002) Signature verification using a modified Bayesian network, *Pattern Recognition*, 35, 983-995.