

# Amélioration du pouvoir discriminant des modèles de Markov cachés par modélisation de la durée d'état

Mohamed ali MAHJOUB<sup>1</sup>

<sup>1</sup>Laboratoire de Système et Traitement du Signal LSTS  
Ecole Nationale d'Ingénieurs de Tunis, Belvedere BP 37, 1002 Tunis - TUNISIE

MedAli.Mahjoub@enit.rnu.tn

**Résumé** - Cet article décrit l'apport de l'incorporation de la durée d'état dans le modèle de Markov caché. Nous proposons à cet effet, un HMM non stationnaire caractérisé par des paramètres de transition dynamiques  $A(t)=\{a_{ij}(t)\}$  dépendants du temps. Comparé aux modèles traditionnels, ce modèle peut être vu comme une généralisation du DHMM. Ce modèle a été appliqué à la reconnaissance en ligne chaînes de caractères arabes cursives appelées PAW (Piece of Arabic Word). La discrimination entre les modèles a été améliorée. De même, de bons taux de reconnaissance ont été enregistrés.

**Abstract** - This paper describes the utility of incorporation of state duration in HMM. Thus, we propose a nonstationary Hidden Markov Model (NSHMM) which is defined with a set of dynamic transitions probability parameters  $A(t)=\{a_{ij}(t)\}$  that depend on the time  $t$  already spent in state  $i$ . When compared to traditional models, this NSHMM is defined as generalisation of the DHMM. The model was applied to on line recognition of hand-written PAWs. The PAWs are represented by a radial sequence which is independent of translation, orientation and homothetic. Experiments have been conducted, and a good classification score has been obtained. The discrimination between characters models has been improved. Experiments showed that this implies a good recognition rates.

## 1. Introduction

Bien qu'il présente différents avantages, relatifs à la clarté, la rapidité et l'efficacité, le modèle de base des modèles de Markov cachés (HMM) reste relativement limité pour exprimer l'information concernant la variabilité de la durée d'état. Le modèle HMM classique ne dispose pas de modélisation explicite de la durée de l'état. En effet, la probabilité de durée d'un état d'un nombre entier de périodes consécutives  $p_i(d)$  peut être calculée comme la probabilité de  $(d-1)$  autotransitions suivies d'une transition vers un état suivant:  $p_i(d) = a_{ii}^{d-1} (1 - a_{ii})$ . Ce qui permet de déduire la relation :  $p_i(d+1) = p_i(d) * a_{ii}$

Ceci montre que la durée d'un HMM classique est un modèle de type sans mémoire à répartition temporelle exponentielle. Dans le cas de l'écriture, par exemple, ce modèle exponentiel de durée d'état n'est pas appropriée. Ainsi, il est préférable d'opérer une modélisation explicite de la durée. A cet effet, des travaux de modélisation de la durée dans les HMMs ont été appliqués en reconnaissance de la parole et en reconnaissance de caractères et ont permis de valider cette approche.

Le cadre de notre étude concerne la reconnaissance en ligne de l'écriture arabe manuscrite. Nous avons choisi le PAW comme unité de traitement (sans tenir compte des diacritiques), afin d'éviter le problème de segmentation, d'autant plus que l'écriture arabe est cursive et difficile à segmenter. La méthode utilisée pour représenter une forme 2-D de PAW est l'utilisation d'une séquence radiale  $r(k)$  représentant la distance du centre de gravité de la forme aux différents points de la forme (FIG. 1). Les observations sont générées après une phase d'échantillonnage, de normalisation, et de quantification. La séquence obtenue est

invariante par rapport à la rotation, la translation et l'homothétie.

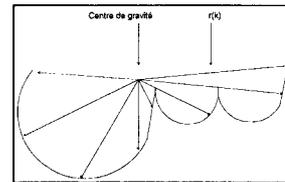


FIG. 1 : Représentation radiale

## 2. Le modèle de durée d'état

### 2.1. Contexte stationnaire

Une première forme de modélisation de la durée, consiste à incorporer de l'information sur la durée de l'unité de reconnaissance globale, c'est à dire, à modéliser le temps de séjour dans un modèle HMM global sans s'occuper des différents états du modèle [8]. Une autre solution est réalisée par un post-processeur. En effet, la probabilité de la durée d'état  $P_j(d)$ , est mesurée à partir des séquences d'apprentissage segmentées. Ainsi, l'estimation de  $P_j(d)$  est strictement heuristique (on utilise souvent une procédure d'alignement temporel de Viterbi).

Une autre solution pour représenter la durée, consiste en l'utilisation des modèles à états étendus, en identifiant chaque état dans un HMM standard, avec un sous-HMM ayant une fonction de densité de probabilité d'observations unique, ou en remplaçant chaque état boucle par un nombre de reproductions de lui-même [10].

Dans [9] la durée d'état est estimée à partir de la base d'apprentissage, pour tenir compte de l'ambiguïté de la segmentation dans le cas de caractères successifs en la reconnaissance de mot manuscrits à l'aide de HMM. Le

modèle est caractérisé par quatre paramètres  $\lambda = (\Pi, A, \Gamma, B, D)$  où  $D$  est la matrice de durée d'états et  $\Gamma$  représente la probabilité de l'état final. Le système correspondant utilise des primitives basées sur la morphologie mathématique où le nombre d'états est 26 (nombre de caractères de l'alphabet).

Dans [3], Furguson propose un modèle de HMM à durée d'état explicite appelé DHMM (Duration Hidden Markov Model). Selon ce modèle, les variables forward-backward sont définies comme suit:

$$\alpha_t(j) = P(o_1, \dots, o_t, S_j \text{ fini à } t \mid \lambda)$$

$$= \sum_{i=1}^N \sum_{d=1}^D \alpha_{t-d}(i) a_{ij} p_j(d) \prod_{s=t-d+1}^t b_j(O_s) \quad (1)$$

$$\beta_t(i) = P(o_{t+1}, \dots, o_T \mid S_i \text{ fini à } t, \lambda)$$

$$= \sum_{j=1}^N \sum_{d=1}^D \beta_{t+d}(j) a_{ji} p_j(d) \prod_{s=t+1}^{t+d} b_j(O_s)$$

L'émission des symboles par DHMM peut se résumer à la procédure suivante:

- 1- Choisir l'état initial  $q_1 = S_i$  suivant la distribution de l'état initial  $\pi_i$
- 2- choisir une durée  $d_1$  suivant la distribution de la durée de l'état  $q_1, p_{q_1}(d_1)$
- 3- choisir la séquence de symboles  $o_1, o_2, \dots, o_{d_1}$  suivant la densité de probabilité conjointe  $b_{q_1}(o_1, o_2, \dots, o_{d_1})$ .
- 4- choisir l'état suivant  $q_2 = S_j$  selon la matrice des transitions, sachant la contrainte  $a_{q_1 q_1} = 0$ . Le processus est répété de l'étape 2 à 4.

Bien que ce modèle, a apporté une amélioration dans les systèmes de reconnaissance, il est resté restreint pour représenter une certaine variabilité. Ainsi, la modélisation dans un processus intrinsèquement dynamique, ou non stationnaire de la transition peut apporter une meilleure description de la variabilité. En effet, dans le cas du DHMM, la transition vers un état est indépendante de la durée de l'état courant à cause de la stationnarité de la matrice des transitions.

## 2.2. Contexte non stationnaire

A ce jour, le caractère non stationnaire de la transition d'état dans un HMM reste encore peu développé. En 1988 Falashi [2] a proposé un modèle caractérisé par une évolution temporelle de la transition. Il a proposé une loi exponentielle, décroissante avec le temps  $d$ , de séjour dans un état  $j$ , pour la probabilité de transition-boucle:  $a_{jj}(d) = \exp(-\mu d)$ . La loi de probabilité de la durée résultante est alors:  $P_j(d) = (1 - \exp(-\mu d)) \prod_{\theta=0}^{d-1} \exp(-\mu \theta)$

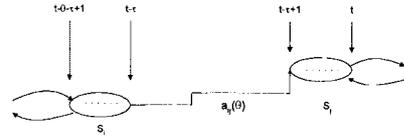
Dans [4], Ghitza propose un modèle de Markov caché appliqué à la reconnaissance de la parole. La méthodologie de reconnaissance proposée combine à la fois les techniques de la comparaison dynamique et les HMMs. Le modèle proposé par Ghitza est un modèle similaire à celui de Levinson [5], où la matrice des transitions est une matrice à diagonale nulle. La différence réside dans la définition des probabilités

d'émission des observations. Selon le modèle proposé, ces probabilités sont dépendantes du temps:  $P(O, T \mid S_q)$ . Il s'agit de la probabilité d'observer la tranche d'observations  $O$  de longueur  $T$  sachant l'état  $S_q$ . Dans [7], une méthode pour modéliser la durée de l'état dans les systèmes de reconnaissance de la parole par HMM est proposée, où les  $a_{ij}(d)$  sont dérivés de la fonction de densité cumulative de la durée d'état. Dans [1], Bong-kee propose un modèle de Markov caché non stationnaire où la matrice des transitions  $A$  est caractérisée par des paramètres dépendants du temps. Contrairement au DHMM, le modèle proposé est caractérisé par des auto-transitions. La loi de probabilité de la durée d'état est:  $P(\tau \mid S_i) = \sum_j a_{ij}(\tau) \prod_{d=1}^{\tau-1} a_{ii}(d)$  qui a un comportement

similaire à celui du DHMM.

## 3. Modèle de Markov caché non stationnaire

Pour mieux exprimer la nature dynamique du script arabe et mieux gérer les transitions entre les états, l'exploration des HMMs non stationnaires a été envisagée. Ainsi, nous proposons un modèle de Markov caché non stationnaire, noté NSHMM, dans lequel les éléments de la matrice de transition sont dépendants du temps. Il est caractérisé par un ensemble de paramètres de probabilité de transitions dynamiques  $A = \{a_{ij}(\tau)\}$ . L'émission des observations se fait comme suit [6]:



- 1- Débuter l'état  $i$  à l'instant  $t - \tau - \theta + 1$
- 2- choisir une durée  $\theta$  suivant la distribution de la durée de l'état  $q, p_q(\theta)$
- 3- choisir la séquence de symboles  $O_{t-\tau-\theta+1}, \dots, O_{t-\tau}$  suivant la densité de probabilité conjointe  $b_q(O_{t-\tau-\theta+1}, \dots, O_{t-\tau})$ . Sachant que les observations  $O_{t-\tau-\theta+1}, \dots, O_{t-\tau}$  sont supposées indépendantes, on a  $b_i(O_{t-\tau-\theta+1}, \dots, O_{t-\tau}) = \prod_{d=t-\tau-\theta+1}^{t-\tau} b_i(O_d)$
- 3- l'état suivant  $j$  est choisi suivant les probabilités de transition  $a_{ij}(\theta)$
- 4- choisir une durée  $\tau$  suivant la distribution de la durée de l'état  $q, p_q(\tau)$
- 5- choisir la séquence de symboles  $O_{t-\tau+1}, \dots, O_t$  suivant la densité de probabilité conjointe  $b_q(O_{t-\tau+1}, \dots, O_t)$ .

En incorporant la durée d'état dans la matrice des transitions, les variables forward-backward seront définies comme suit:

$$\alpha_{t,d}(j) = P(o_1, \dots, o_t, q_1 = S_j, t-d+1 \leq 1 \leq t \mid \lambda)$$

$$= \sum_{i=1}^N \sum_{d=1}^D \alpha_{t-d}(i) a_{ij}(\delta) \prod_{k=1}^{d-1} a_{jj}(k) \prod_{s=t-d+1}^t b_j(O_s) \quad (2)$$

$$\beta_{t,d}(i) = P(o_{t+1}, \dots, o_T \mid q_1 = S_i, t-d+1 \leq 1 \leq t, \lambda)$$

$$= \sum_{j=1}^N \sum_{d=1}^D a_{ij}(\delta) \beta_{t+d}(j) \prod_{k=1}^{d-1} a_{jj}(k) \prod_{s=t+1}^{t+d} b_j(O_s)$$

$$\text{avec } P(O|\lambda) = \sum_{i=1}^N \sum_{\tau=1}^D \alpha_{\tau,\tau}(i) = \sum_{i=1}^N \alpha_{\tau}(i)$$

L'analyse des formules associées au modèle non stationnaire proposé permet de constater que le NSHMM représente une généralisation du DHMM. En effet, dans le cas particulier :

$\tau_1 \neq \tau_2$  et  $a_{ij}(\tau_1) = a_{ij}(\tau_2)$ , on peut écrire:

$$a_{ij}(\tau) = a_{ij} \quad \text{et} \quad \alpha_{t-d}(i) = \sum_{\theta=1}^D \alpha_{t-d:\theta}(i)$$

Donc, l'équation (1) devient:

$$\alpha_{t,d}(j) = \sum_{i \neq j}^N \sum_{\theta=1}^D \alpha_{t-d:\theta}(i) a_{ij}(\theta) P_j(d) \prod_{s=t-d+1}^t b_j(O_s) = \sum_{i=1}^N \alpha_{t-d}(i) a_{ij} P_j(d) \prod_{s=t-d+1}^t b_j(O_s)$$

Ce qui permet de retrouver l'équation (1) correspondant au DHMM:

$$\alpha_{t,d}(j) = \sum_{i=1}^D \alpha_{t,d}(j) = \sum_{i=1}^N \sum_{d=1}^D \alpha_{t-d}(i) a_{ij} P_j(d) \prod_{s=t-d+1}^t b_j(O_s)$$

Un raisonnement similaire permet de déduire aussi  $\beta_i(i)$ . Pour plus de détails, les lecteurs peuvent consulter [6].

## 4. Application à la reconnaissance de PAWs

L'application du NSHMM pour la génération des modèles associés aux PAWs, a permis de justifier la modélisation de la durée d'états. En effet, pour la plupart des lettres, et en particulier celles présentant des structures dynamiques tels que ع, س, ص, ط, ش, ي, les résultats ont permis de vérifier la variation de la durée d'état. La figure 2 illustre la distribution de la durée d'états pour la lettre س, pour deux représentations différentes.

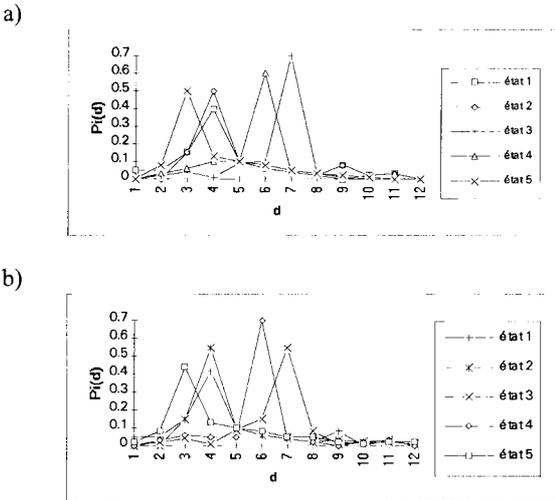


FIG. 2: Distribution de la durée d'états associée la lettre س  
a) Petite taille b) Grande taille

En revanche, pour les lettres de structures uniformes tels que « 1, 0 » il a été remarqué que les différents états présentent une distribution similaire de durée (FIG. 3), ce qui permet dans ce cas de remplacer la matrice de durée d'état par un vecteur.

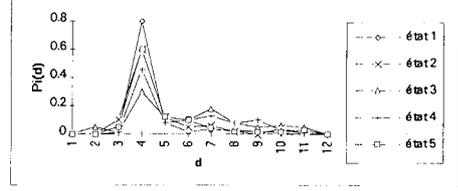


FIG. 3 :Distribution de la durée d'états associée à la lettre 1

Par ailleurs, il est à remarquer dans tous les cas que cette approche est indépendante de la taille des caractères. Les figures 2.a et 2.b montrent une nette similitude pour les deux représentations de la lettre س pour deux tailles différentes.

### 4.1. Discrimination inter-modèles

Pour mieux apprécier l'apport des NSHMMs dans la génération des modèles associés aux PAWs, nous avons procédé au calcul de la distance inter-modèles.

La distance  $D_s(d)$  (d est la durée maximale d'état associé à un modèle  $\lambda^d$ ) entre deux modèles  $\lambda_1$  et  $\lambda_2$  peut être exprimée comme suit:

$$D_s(d) = \frac{D(\lambda_1^d, \lambda_2^d) + D(\lambda_2^d, \lambda_1^d)}{2} \quad (3)$$

avec,

$$D(\lambda_1^d, \lambda_2^d) = \frac{1}{T_2} \left[ \log P(O^2 | \lambda_2^d) - \log P(O^2 | \lambda_1^d) \right] \quad (4)$$

$$D(\lambda_2^d, \lambda_1^d) = \frac{1}{T_1} \left[ \log P(O^1 | \lambda_1^d) - \log P(O^1 | \lambda_2^d) \right] \quad (5)$$

où,  $O^1$  est une séquence d'observations de longueur  $T_1$  générée par le modèle  $\lambda_1$  et  $O^2$  est une séquence d'observations de longueur  $T_2$  générée par le modèle  $\lambda_2$ .

$D_s$  est une mesure de similitude entre  $\lambda_1$  et  $\lambda_2$ , plus sa valeur est petite plus les deux modèles sont semblables, en revanche plus la valeur de  $D_s$  est grande plus la discrimination entre les modèles est importante.

Considérons la fonction  $f(d)$  définie par :

$$f(d) = \frac{D_s(d_{\max}) - D_s(d)}{D_s(d)} \quad (6)$$

où  $d_{\max}$  est une valeur limite de d.

En faisant varier d ( $d=1, \dots, d_{\max}$ ), la figure 4 montre une diminution de la valeur de  $f(d)$  pour des valeurs croissantes de d, pour un échantillon restreints de PAWs. Ceci signifie que  $D_s(d_{\max}) \geq D_s(d)$  pour  $d=1, \dots, d_{\max}$ . Ce qui permet de constater que la distance entre deux modèles est plus significative quand la valeur de d croit.

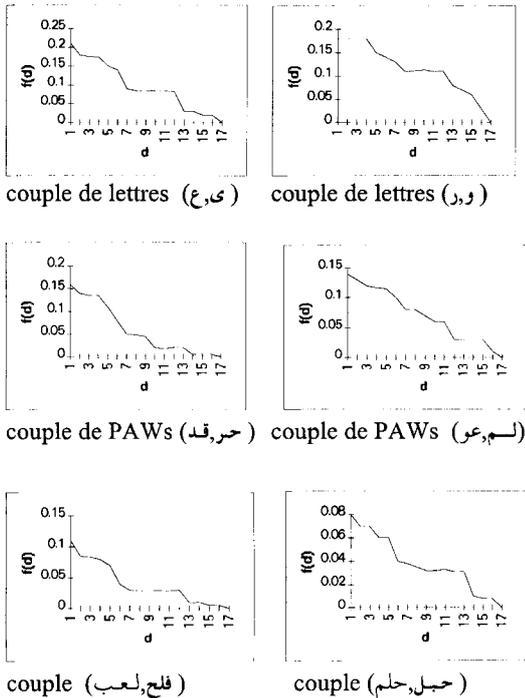


FIG. 4 : Amélioration de la discrimination inter-modèles

Signalons à cet effet, que pour les différents tests réalisés, nous avons considéré trois bases d'apprentissage (SB1,SB2,SB3) :

SB1={PAWs composés de 1 lettres} + {chiffres arabes}

SB2=SB1+{PAWs composés de 2 lettres}

SB3=SB2+{PAWs composés de 3 lettres}

Notre approche a permis de valider l'application des NSHMMs à la reconnaissance en ligne des PAWs, en effet le taux de reconnaissance a atteint 93.5% pour une base d'apprentissage comportant 5700 PAWs écrits par 3 scripteurs différents (TAB. 1). Pour mieux apprécier l'apport du NSHMM dans la discrimination entre les modèles, nous avons opéré la reconnaissance sur la base d'apprentissage en appliquant le HMM, le DHMM et le NSHMM. Nous avons reporté dans TAB. 2 les taux de reconnaissance par forme de PAW (échantillon restreint).

TAB. 1: Taux de reconnaissance, pour la base de tests (taille : 11400) en fonction de D (durée maximale d'état)

D	8	10	13	14	15
SB1	95.3%	96.3%	97.9%	97.9%	97.5%
SB2	90.5%	91.2%	93.2%	94.1%	93.6%
SB3	88.8%	88.9%	91.3%	91.9%	93.5%

TAB. 2 : Taux de reconnaissance (Base d'apprentissage) par forme de PAW  
Taille de la base d'apprentissage : 5700

Modèle	فلح	بنت	مصر	قفص	علم	معه	حجم	صلب	سجل	حلم	خير	سمع	لعب	حبل
HMM	92	92	93	93	91	93	96	93	98	91	90	89	90	88
DHMM	93	96	95	93	96	96	100	97	99	95	93	91	95	91
NSHMM	95	96	96	95	97	99	100	99	99	97	98	98	97	95

Les résultats montrent que les meilleurs taux de reconnaissance sont obtenus par application du NSHMM. Ce qui permet de conclure alors, que le NSHMM assure une meilleure discrimination inter-modèles comparé au HMM et au DHMM.

## 5. Conclusion

Nous avons décrit dans cet article un modèle de markov caché non stationnaire (NSHMM). Ce modèle est caractérisé par des transitions dynamiques et une modélisation explicite de la durée d'état. De ce fait, le NSHMM représente une généralisation du DHMM. En effet, le NSHMM permet une modélisation de la durée d'état à l'instar du DHMM tout en offrant une meilleure transition entre les états. L'application de ce modèle à la reconnaissance en ligne de PAWs, a permis d'apporter une amélioration dans la discrimination entre les modèles, ce qui représente une amélioration loin d'être négligeable. Ainsi, pour une valeur optimale de D, le taux de reconnaissance a atteint 93.5%.

## Références

- [1] Bong-Kee Sin and Jin H. Kim « *Nonstationary hidden Markov model* », Signal processing, 36(1995), pp. 31-46.
- [2] A. Falaschi « *Phonetic recognition of natural speech by non stationary Markov models* », Rapport technique, ENST, Dep. Signal, Paris 1988.
- [3] Furguson J. D. « *Variable duration models for speech* », proc. symp. on the application of hidden markov models to text and speech, Princeton, NJ, 1980, pp 143-179.
- [4] Ghitza Oded , M. M. Sondhi « *Hidden Markov models with templates as non-stationary states : an application to speech recognition* », Computer speech and language, pp 101-119.
- [5] S. E. Levinson « *Continuously variable duration hidden Markov models for automatic speech recognition* », comp. speech language, vol. 1, N° 1, March 1986, pp 29-45.
- [6] Mahjoub Mohamed Ali, Noureddine Ellouze « *Non stationary Hidden Markov model: Application to on-line character recognition* », CESA'98, Computational Engineering in Systems Applications, Hammamet-Tunisie, April 1-4, 1998.
- [7] Vaseghi S. « *State duration modelling in hidden Markov models* », Signal processing , 41, pp 31-41, 1995
- [8] Lawrence R. Rabiner, B. H. Juang « *Mixture autoregressive hidden Markov models for speech signals* » IEEE, Proceedings, vol. ASSP 33, N° 6, pp. 1404-1413, December 1985.
- [9] Mou-Yen Chen, Amlan Kundu and Sargur N. Srihari « *Handwritten word recognition using continuous density variable duration hidden Markov model* », IEEE, pp. 105-108, 1993.
- [10] M. Codogno, L. Fissore « *Duration modelling in finite state automata for speech recognition and fast speaker adaptation* », Proc. IEEE ICASSP, pp. 1269-1272, Dallas 1987.