Modèle de markov à terminaisons cachées

Bruno Taconet, Abderrazak Zahour, Saïd Ramdane

Modèle de Markov à terminaisons cachées

Bruno Taconet, Abderrazak Zahour, Saïd Ramdane

Equipe de Gestion Electronique de Documents
Université du Havre, IUT
Place R. Schuman 76610 Le Havre

e-mail : taconet[zahour][ramdane]@iut.univ-lehavre.fr

Résumé. *Nous présentons un nouveau modèle stochastique de Markov caché différent du modèle classique états/observations, et plus simple que le modèle stochastique utilisé en reconnaissance génomique. La chaîne des observations est vue comme la partie apparente d'une chaîne cachée, elle-même composée d'observations (visibles) et de terminaisons (cachées) d'états. A un instant donné, l'examen de la paire formée par le symbole courant et le symbole précédent de la chaîne cachée caractérise une des 4 opérations élémentaires : substitution, suppression, insertion, terminaison normale. Les méthodes apprentissages de Baum-Welch et de Viterbi sont transposées à ce modèle, les formules détaillées sont fournies. Une comparaison avec le modèle classique est effectuée.*

Mots-clés : modèle de Markov caché, terminaison cachée, distance d'édition, apprentissage

1 Introduction

Le modèle de Markov caché 1-D orienté dans le sens de l'écriture est une représentation très souvent utilisée en reconnaissance de l'Ecrit, car il prend en compte une double variabilité : la présence/absence d'une observation, et la mesure d'une observation lorsqu'elle est présente. Dans le cas discret, deux matrices et un vecteur servent à définir quantitativement le modèle : la matrice A des transitions entre états, et la matrice B des observations dans les états, le vecteur Π de l'état initial. Ces grandeurs sont les invariants probabilistes du modèle. Ce modèle classique est pratiqué usuellement en reconnaissance de l'écriture et de la parole.

Dans le sillage des recherches intensives sur le décodage du génome, a surgi au milieu des années 90 un autre modèle de Markov caché, s'appuyant sur l'existence d'une chaîne de référence, et sur la détermination des probabilités des transformations élémentaires : suppression, substitution, insertion. Sa complexité de mise en œuvre, notamment le difficile alignement des séquences, explique sans doute que ce modèle reste actuellement cantonné au domaine de la bio-informatique.

En analysant certaines imperfections du modèle classique, nous étions parvenus à la conclusion que les invariants simples qui reflètent le mieux la réalité de l'Ecrit sont liés aux probabilités des transformations élémentaires des chaînes de symboles : suppression, substitution, insertion. Ces transformations élémentaires sont aussi celles de la distance d'édition [LEP 96]. Confrontés à l'utilisation de différentes méthodes, nous demeurons convaincus que les méthodes de nature stochastique ont une faculté inégalée d'adaptation automatique aux divers cas d'apprentissage. Enfin, garantir l'automaticité de l'apprentissage nous paraît indispensable.

C'est pourquoi depuis plusieurs années, nous travaillons sur la construction d'un modèle de description qui possède ces trois propriétés : markovien caché orienté, basé sur des invariants probabilistes reliés aux transformations élémentaires, à apprentissage automatique. Ce modèle est aujourd'hui achevé et en phase de test

2 L'objectif du nouveau modèle

2.1 La dépendance probabiliste substitution/insertion

Le modèle classique appliqué à l'Ecrit est calqué sur celui de la reconnaissance de la parole. Il est orienté dans le sens de l'écriture. C'est une double chaîne stochastique. Les sauts d'état (orientés) et les bouclages d'état sont autorisés dans la chaîne cachée. Un saut d'état modélise la suppression d'un emplacement logique d'observation, un bouclage d'état autorise l'insertion d'un emplacement logique d'observation [RAB 89][DUG 96]. Cependant, il est impossible de régler indépendamment, pour une observation donnée, la probabilité de substitution et celle d'insertion de chacune des N observations. Dans un état donné, pour N symboles d'observations, il faudrait 2N invariants indépendants, alors que le modèle en fournit seulement N + 1 : N composantes d'une ligne de la matrice des observations et le contenu d'une seule cellule de la matrice des transitions. Ce phénomène pourrait être évité par le dédoublage des états. Chaque macro-état se

dédoublerait alors en un état de substitution et un état d'insertion. Le nombre d'invariants du nouveau modèle serait alors approximativement multiplié par 2.

2.2 La complexité du modèle génomique

Le modèle génomique ou modèle profilé (Profile-hmm) repose sur l'existence d'une chaîne de référence. A chaque observation de référence on fait correspondre un état. Les opérations de substitution, de suppression et d'insertion sont directement modélisées à l'intérieur de l'état par des "sous-états" M : matching, seule une substitution avec l'observation de référence est possible ; I : insertion d'une observation ; D : deletion, suppression de l'observation de référence (cf. figure 1).

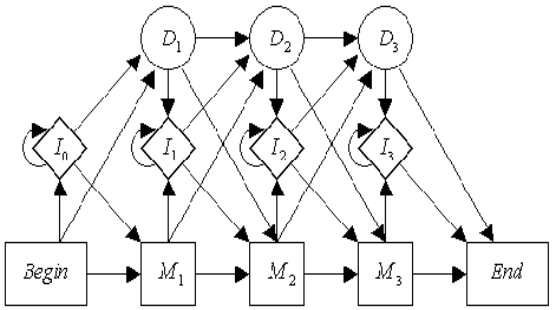

FIGURE 1 : Un HMM profilé à 3 états

L'algorithme de Baum-Welch met en jeu trois probabilités forward et trois probabilités backward comportant une somme de trois produits [EDD 98] [DUR 98].

2.3 Emission des symboles dans le modèle à terminaisons cachées

Le nouveau modèle doit être plus simple que le modèle génomique et plus précis que le modèle classique. Il repose sur l'existence d'une chaîne stochastisque cachée composée de symboles d'observation et de terminaisons cachées d'états. La figure 2 montre une cellule de la machine à émettre les symboles aléatoirement. La machine comprend autant de modules que d'états. L'horloge cadence la machine séquentielle : à un instant discret, seul un des deux moteurs d'émission dans l'état courant fonctionne. Le moteur de substitution/terminaison (substitution généralisée) est activé au premier top d'horloge qui affecte l'état courant, puis le moteur d'insertion/terminaison (insertion généralisée) est activé tant que l'on reste dans l'état courant. Un symbole émis est soit une observation soit un symbole caché de terminaison λ, lequel produit automatiquement le passage à l'état suivant. La chaîne d'observation est visible. La chaîne des symboles cachée est complète : elle permet de trouver les états et de ranger les observations dans les états. L'état initial (noté 0) ne contient qu'un générateur, celui d'insertion/terminaison.

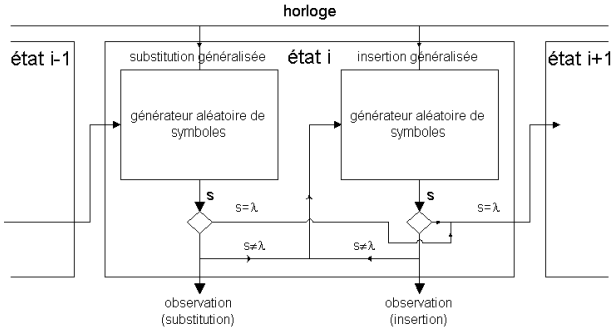

FIGURE 2 : Structure d'une cellule de la machine

3 Définition formelle du modèle de Markov à terminaisons cachées

3.1 Notations

$o_1, o_2, \ldots, o_j, \ldots o_N$: symboles d'observation
λ : symbole de terminaison cachée
$o_1, o_2, \ldots, o_j, \ldots o_N, \lambda$: symboles de la chaîne cachée
1,2, …i … n : états (cachés)
1,2, … t .. T : instants (cachés) d'émission observations généralisées ou symboles
1,2 … θ … Θ : indices des observations dans une séquence, rangs d'observation
q_t = état à l'instant t
s_t = symbole émis à l'instant t
o_θ = observation de rang θ
$O_{1\theta} = o_1, \ldots o_\theta$: séquence d'observations partielle de rang 1 à θ
$O_{\theta\Theta} = o_\theta, \ldots, o_\Theta$: séquence d'observations partielle de rang θ à Θ
$O = o_1, \ldots o_\Theta$: séquence d'observations complète de 1 à Θ
S = séquence complète de symboles : $s_0\ s_1\ s_2 \ldots s_t \ldots s_T$

3.2 Caractérisation

Nous définissons un modèle de Markov gauche/droite à terminaisons cachées par :
Un ensemble de symboles d'observation, numérotés de 1 à N
Un symbole de terminaison λ
Un ensemble d'états principaux $\{q_i\}$, numérotés de 0 à n, 0 est l'état de départ
deux règles d'évolution d'état
La règle d'initialisation impose que l'état initial vaut 0:
$q_{t=0} = 0$
La règle d'actualisation (déterministe) de l'état s'écrit en termes probabilistes:

$$p(q_t = q_{t-1} + 1 \mid s_{t-1} = \lambda) = 1$$

$$p(q_t = q_{t-1} \mid s_{t-1} \neq \lambda) = 1$$

Une matrice de substitution *(matching)*

$$p(s_t = o_j \mid q_t = i, q_{t-1} = i-1) = a_{ij} \qquad i = 0..n, j = 1..N$$

Un vecteur de destruction (terminaison de sous-chaîne vide) *(deletion)*

$$p(s_t = \lambda \mid q_t = i, q_{t-1} = i) = a_i \qquad i = 0..n$$

Une matrice d'insertion *(insertion)*

$$p(s_t = o_j \mid q_t = i, q_{t-1} = i) = b_{ij} \qquad i = 0..n,\ j = 1..N$$

Un vecteur de terminaison normale *(normal ending)*

$$p(s_t = \lambda \mid q_t = i, q_{t-1} = i) = b_i \qquad i = 0..n$$

Les relations liant les probabilités sont :

$$\sum_{j=1}^{N} b_{ij} + b_i = 1, \forall i = 0..n \text{ , et : } \sum_{j=1}^{N} a_{ij} + a_i = 1, \forall i = 0..n$$

Ces 4 invariants peuvent être groupés deux à deux pour former deux matrices :

1. Matrice des substitutions généralisées A*

On $a_i = a_{i0} = p(s_t = \lambda \mid q_t = i, q_{t-1} = i-1)$
La destruction est assimilée à la substitution par le symbole λ (d'indice 0).
La matrice de substitution généralisée s'écrit alors :
$A^* = \{a_{ij}\}\ i = 1..n,\ j = 0..K$

2. Matrice des insertions généralisées B*

On pose $b_i = b_{i0} = p(s_t = \lambda \mid q_t = i, q_{t-1} = i)$
La terminaison normale est assimilée à l'insertion du symbole λ (d'indice 0).
La matrice de substitution généralisée s'écrit alors :
$B^* = \{b_{ij}\}\ i = 0,1..n,\ j = 0..N$

La matrice A* (respectivement : B*) englobe la matrice A (respectivement : B) et le vecteur de destruction (respectivement : le vecteur de terminaison normale).
Les relations caractéristiques des probabilités s'écrivent alors plus simplement :

$$\sum_{j=0}^{N} a^*_{ij} = 1, \forall i = 0...n, \text{ et : } \sum_{j=0}^{N} b^*_{ij} = 1, \forall i = 0..n.$$

Dans la suite, pour plus de clarté, le modèle sera composé des vecteurs de suppression et de terminaison séparément des matrices de substitution et d'insertion.

3.3 Particularités du modèle

3.3.1 Les deux horloges et la représentation totale

Il est important de distinguer le temps "interne" t, appelé instant d'émission, temps d'émission des symboles, du temps "externe" θ, appelé rang des observations, qui est le temps observable rythmé par la chaîne d'observations (cf. figure 3). Pour une observation o_θ, la relation entre t (instant d'émission), θ (rang)et i (état) est : t=θ+i-1.
A un symbole de terminaison d'état λ émis à l'instant t, on convient d'associer le rang de l'observation la plus récente. La relation entre t, θ et i devient : t=θ+i.

Symboles	λ	A	λ	λ	B	B	λ
Instants t	0	1	2	3	4	5	6
Rang θ	0	1	1	1	2	3	3
Etat	0	1	1	2	3	3	3

FIGURE 3 : Représentation totale associée à la chaîne de symbole λAλλBBλ

3.3.2 Les deux représentations complètes

La relation précédente implique que la représentation par chaîne de symbole équivaut à la représentation classique observation/état (cf. figure 4). Chacune des deux représentations est complète : on déduit de chacune la représentation totale.

Symboles	λ	A	λ	λ	B	B	λ

Observations	A	B	B
Etat	1	3	3

FIGURE 4 : Les deux représentations complètes

3.3.3 Les quatre configurations de base

La suite de deux symboles de la chaîne cachée du modèle de Markov d'ordre 1 engendre 4 configurations élémentaires (cf. figure 5).

	Substitution		Destruction		Insertion		Terminaison	
Instants	t-1	**t**	t-1	**t**	t-1	**t**	t-1	**t**
Symboles	λ	$\mathbf{o_\theta}$	λ	**λ**	$o_{\theta-1}$	$\mathbf{o_\theta}$	$o_{\theta-1}$	**λ**
Etats	i-1	**i**	i-1	**i**	i	**i**	i	**i**

FIGURE 5 : les 4 configurations de base

4 Apprentissage par la méthode de Baum-Welch

4.1 probabilités forward/backward

Les probabilités backward et forward prennent deux formes, selon que l'observation o_θ , ou le symbole de terminaison d'état λ, est émis à l'instant t.

4.1.1 Probabilités forward

$$\alpha_\theta(i) = p(o_1,...,o_\theta, s_t = o_\theta = o_j, q_t = i) \quad t = \theta + i - 1$$

$$\gamma_\theta(i) = p(o_1,...,o_\theta, s_t = \lambda, q_t = i) \quad t = \theta + i$$

initialisations

$$\alpha_1(0) = b_0(o_1 = o_j) \quad \forall i \neq 0\ ,$$

$$\gamma_1(0) = b_0(o_1 = o_j).b_0(\lambda)$$

relations de récurrence

$$\alpha_\theta(i) = \alpha_{\theta-1}(i).b_i(o_\theta = o_j) + \gamma_\theta(i).a_i(o_\theta = o_j)$$

$$\gamma_\theta(i) = \alpha_\theta(i).b_i(\lambda) + \gamma_\theta(i-1).a_i(\lambda)$$

4.1.2 Probabilités backward

$\beta_\theta(i) = p(s_t \neq \lambda, q_t = i, o_{\vartheta+1}, ..., o_\Theta) \quad t = \theta + i - 1$

$\delta_\theta(i) = p(s_t = \lambda, q_t = i, o_{\vartheta+1}, ..., o_\Theta) \quad t = \theta + i$

initialisations

$\beta_\Theta(n) = b_n(\lambda)\ ,\ \ \delta_\Theta(n) = 1$

relations de récurrence

$\beta_\theta(i) = \beta_{\theta+1}(i).\, b_i(o_{\theta+1} = o_j) + \delta_\theta(i).b_i(\lambda)$

$\delta_\theta(i) = \beta_{\vartheta+1}(i+1).a_{i+1}(o_{\theta+1} = o_j) + \delta_\theta(i+1).a_{i+1}(\lambda)$

4.2 Formules de ré-estimation du modèle

Matrice de substitution

$$\overline{a_{ij}} = \frac{\sum_{\theta=1,\theta=o_j}^{\Theta} \gamma_{\theta-1}(i-1).\beta_\theta(i)}{\sum_{\theta=1}^{\Theta} \gamma_{\theta-1}(i-1).\beta_\theta(i) + \sum_{\theta=0}^{\Theta} \gamma_\theta(i-1).\delta_\theta(i)}$$

Vecteur de destruction

$$\overline{a_i} = \frac{\sum_{\theta=0}^{\Theta} \gamma_\theta(i-1).\delta_\theta(i)}{\sum_{\theta=1}^{\Theta} \gamma_{\theta-1}(i-1).\beta_\theta(i) + \sum_{\theta=0}^{\Theta} \gamma_\theta(i-1).\delta_\theta(i)}$$

Matrice d'insertion

$$\overline{b_{ij}} = \frac{\sum_{\theta=1,o_\theta=o_j}^{\Theta} \alpha_{\theta-1}(i).\beta_\theta(i)}{\sum_{\theta=1}^{\Theta} \alpha_{\theta-1}(i).\beta_\theta(i) + \sum_{\theta=0}^{\Theta} \alpha_\theta(i).\delta_\theta(i)}$$

Vecteur de terminaison

$$\overline{b_i} = \frac{\sum_{\theta=0}^{\Theta} \alpha_\theta(i).\delta_\theta(i)}{\sum_{\theta=1}^{\Theta} \alpha_{\theta-1}(i).\beta_\theta(i) + \sum_{\theta=0}^{\Theta} \alpha_\theta(i).\delta_\theta(i)}$$

5 Décodage

5.1 Transposition de l'algorithme de Viterbi

Classiquement, décoder c'est trouver la chaîne d'état cachée Q_C associée à l'observation O donnant la probabilité maximum : $Q_c = \arg\max_Q p(O, Q | O, \Lambda)$.

Dans le présent modèle, décoder équivaut aussi à déterminer la chaîne cachée de probabilité maximum $S_c = \arg\max_S p(S | O, \Lambda)$.

Ceci est fait par une transposition de l'algorithme de Viterbi au présent modèle. En réalité, nous avons appliqué un algorithme équivalent de programmation dynamique de la distance d'édition avec les coûts d'édition déduits des invariants

5.2 Coûts d'édition

Comme la probabilité estimée d'une séquence est obtenue par un produit de probablités (invariants du modèle), le logarithme est additif ; il permet donc d'exprime un coût d'édition par rapport à une chaîne de référence.

Coût d'insertion de o_j dans l'état i: :

$c(o_j, i) = -\log(b_{ij})$ i=0..n

Coût de substitution de o_k par o_j::

$c(o_k, i) = -\log(a_{ik}) + \log(a_{i\ ref})$ i=1..n

Coût de destruction de o_j dans l'état i :

$c(\lambda, i) = -\log(a_i) + \log(a_{ij\ ref})$ i=1..n

La chaîne de référence S_{ref} du modèle (sans insertion, ni suppression, de probabilité de substitution locale maximum) a ainsi un coût de transformation nul.

6 Comparaisons et tests

6.1 Nombre et répartition des invariants

Un invariant est un paramètre numérique probabiliste à apprendre.

Les deux matrices généralisées A* et B* comprennent nx(N+1) et (n+1)x(N+1) éléments, soit, en tenant compte de la somme des probabilités unitaires, (2n+1)xN invariants probabilistes indépendants, ainsi répartis : nxN pour la substitution, (n+1)xN pour l'insertion. Au contraire, le modèle classique (orienté) comprend ((n+1)x(n/2) invariants indépendants pour la matrice de transitions entre états, et nxN invariants pour la matrice d'observations dans les états. Le nouveau modèle est plus fin pour traiter les observations présentes (2nxN invariants au lieu de nxN) et moins fin pour traiter les suppressions (n invariants au lieu de (n-1)(n)/2). Ainsi, le présent modèle prend moins bien en compte les suppressions multiples d'états adjacents, puisqu'il les prend comme deux suppressions statistiquement indépendantes, mais il peut apprendre sans aucun mélange substitutions et insertions.

6.2 Un exemple simple avec calcul manuel

L'ensemble des caractères d'observation est {A, B,C}

L'ensemble d'apprentissage est composé de 20 mots :
ABC(18), ABCA(1), ABCB(1).

Cet ensemble peut s'interprété ainsi :

La séquence de référence est ABC;

Un caractère A ou B a été inséré 1 fois après le troisième caractère du modèle (C).

Il est donc naturel de fixer le nombre d'états du modèle à 3, c'est-à-dire le nombre d'observations de la séquence de référence.

Pour le modèle classique, les séquences cachées d'états se déduisent facilement de cette interprétation pour chaque configuration de l'ensemble d'apprentissage :

États	123	1233	1233
Observations	ABC	ABCA	ABCB

Il y a 20 transitions début->1, 20 transitions 1->2, 20 transitions 2->3, 2 transitions 3->3 et 20 transitions 3->fin. On en déduit la matrice de transition de la méthode d'apprentissage de Viterbi. (figure 6.a)

Il y a 20 observations A dans l'état 1, 20 observations B dans l'état 2, 20 observations C dans l'état 3, 1 observation A dans l'état 3, 1 observation B dans l'état 3. On en déduit la matrice d'observations. de la méthode d'apprentissage de Viterbi. (figure 6.b)

	0	1	2	3
1	1	0	0	0
2	0	1	0	0
3	0	0	1	2/20

a. Matrice des transitions

	A	B	C
1	1	0	0
2	0	1	0
3	1/22	1/22	20/22

b. Matrice des observations

FIGURE 6 : Modèle classique avec sauts d'état (Apprentissage manuel)

Pour le modèle à terminaison cachées, les matrices de subtitution et d'insertion découlent directement de l'interprétation.

ABC **λAλBλCλ**
ABCA **λAλBλC**Aλ
ABCB **λAλBλC**Bλ

Il n'y a pas de saut d'état, donc les probabilités de suppression sont nulles. (symbole λ dans la matrice A* de substitution généralisée)
Le caractère A apparaît 20 fois sur 20 comme caractère de substitution dans l'état 1
Le caractère B apparaît 20 fois sur 20 comme caractère de substitution dans l'état 2
Le caractère C apparaît 20 fois sur 20 comme caractère de substitution dans l'état 3

Il n'y a pas d'insertion dans les états 0, 1 2. Donc la terminaison normale λ a la probablité maximum dans la matrice B* dinsertion génralisée)
Dans l'état 3, A et B sont insérés une fois, et il y a 20 terminaisons normales.(symbole λ dans la matrice B*).

	λ	A	B	C
1	0	1	0	0
2	0	0	1	0
3	0	0	0	1

Matrice de substitution généralisée

	λ	A	B	C
0	1	0	0	0
1	1	0	0	0
2	1	0	0	0
3	20/22	1/22	1/22	0

Matrice d'insertion généralisée

FIGURE 7 : Modèle à terminaisons cachées (Apprentissage manuel)

Les probabilités estimées par comptage sur le modèle et les probabilités calculées par les deux modèles de Markov définis plus haut sont rassemblées dans le tableau de la figure 8.

	ABC	ABCA	ABCB	ABCC
Classique	18/20	2/440	2/440	2/22
T. Cachées	20/22	1/22	1/22	0
Ens.. d'App.	18/20	1/20	1/20	0

FIGURE 8 : Probabilités des séquences d'apprentissage

On constate un apprentissage assez précis du modèle à terminaisons cachées. Au contraire, le modèle classique fait apparaître un biais qui surestime l'insertion par répétition, et sous-estime les probabilités des autres insertions.

6.3 Mise en œuvre programmée et tests de validité

6.3.1 Programmation des modèles

Nous avons programmé le modèle à terminaisons cachées en langage C++, comme le modèle classique, sous forme d'une classe unique dont les données-membres sont protégées et la plupart des méthodes sont publiques. On peut choisir dans la ligne de commande le mode d'apprentissage (Baum-Welch ou Viterbi), et la distribution initiale : équiprobable, aléatoire, ou déjà apprise. La validité du programme a été testé sur plusieurs ensembles d'apprentissage simples, apprenables manuellement, comme celui du paragraphe 6.2. Les deux matrices de la figure 9 montrent le résultat de l'apprentissage automatique par le programme.

	λ	A	B	C
1	.0001	.9997	.0001	.0001
2	.0001	.0001	.9997	.0001
3	.0001	.0001	.0001	.9997

Matrice de substitution généralisée

	λ	A	B	C
0	.9997	.0001	.0001	.0001
1	.9997	.0001	.0001	.0001
2	.9997	.0001	.0001	.0001
3	.9090	.0455	.0455	.0001

Matrice d'insertion généralisée

FIGURE 9 : Modèle à terminaisons cachées (Apprentissage automatique)

6.3.2 Test sur des ensembles d'apprentissage prédéfinis

Une première série de tests a consisté à apprendre le modèle (10 passes Baum-welch + 1 passe Viterbi) à partir d'ensembles d'apprentissage représentatifs, obtenus par transformations élémentaires d'une séquence de référence d'une seule sorte, (suppression, seule, substitution seule, insertion seule), puis mixte. Les 10 passes de la méthode de méthode de Baum-welch suffisent pour assurer la convergence ; la passe de Viterbi, qui s'appuie sur l'apprentissage précédent, permet de finaliser l'apprentissage en éliminant les ambiguïtés propres à la méthode de Baum-Welch, ce que confirme le calcul de la distance entropique (définie ultérieurement).

Les ensembles d'apprentissage sont construits en privilégiant la séquence de référence, plus précisément en lui donnant un poids égal à la réunion de toutes les autres séquences.

Par exemple, pour l'insertion seule et pour 3 états, l'ensemble de test s'écrit :

ABC(9), BABC(1), CABC(1), AABC(1), ABBC(1), ACBC(1), ABAC(1), ABCC(1), ABCA(1), ABCB(1).

Pour mesurer la pertinence d'un modèle M, calculé à partir d'un ensemble d'apprentissage E, nous avons choisi de calculer la distance entropique entre la distribution de probabilité p_i des séquences comptées sur l'ensemble d'apprentissage et la distribution de probabilités estimée par apprentissage, après normalisation : $q'_i = qi/\sum q_i$

$$d_E(M) = -\sum_i p_i . \log \frac{p_i}{q'_i}$$

La figure 9 montre la distance entropique normalisée entre la distribution de départ et le modèle, rapportée au plus proche et moyennée lorsque la séquence de référence comprend de 3 à 8 symboles d'observation distincts (ABC, ABCD, …, ABCDEFGH)

	Suppression	Substitution	Insertion	Mixte
Classique	1.0	1.4	4.2	2.7
T. Cachées	1.2	1.0	1.0	1.0

FIGURE 10 : Distance entropique normalisée rapportée au plus proche

Ces résultats confirment la supériorité du nouveau modèle pour distinguer la substitution de l'insertion, et celle du modèle classique pour prendre en compte les sauts multiples d'états.

6.3.3 Premiers tests sur la base MNIST

Les tout premiers tests de reconnaissance élémentaires de caractères manuscrits effectués sur la base de chiffres manuscrits MNIST [LEC 98] ont montré une diminution globale du taux de confusion de 7% en faveur du nouveau modèle.

7 Conclusion

Nous avons présenté un modèle stochastique complet conçu pour reconnaître l'Ecrit., qui permet d'apprendre les coûts de la distance d'édition. Ce modèle vaut plus généralement pour toute forme qui peut être décrite par une chaîne d'observations, lorsque les probabilités des trois transformations élémentaires (substitution, suppression, insertion) sont des invariants probabilistes.

8. Bibliographie

[DUG 96] R. DUGAD, U. B. DESAI, "A Tutorial on Hidden Markov Models*", Technical Report No. SPANN-96.1*, May 1996.

[DUR 98] R. DURBIN, S. EDDY, A. KROGH, and G. MITCHISON. *Biological sequence analysis: Probabilistic models of proteins and nucleic acid*s. Cambridge University Press, 1998.

[EDD 98] S. R. EDDY. Profile Hidden Markov Models, *Bioinformatic*s, 14(9):755–763, 1998.

[LEP 96] LEPY N. "Intégration des aspect linguistiques en lecture automatique de l'écriture cursive manuscrite", *Rapport de DEA*, IRISA, Rennes, Septembre 1996.

[RAB 89] RABINER L. R. A tutorial on hidden markov models and selected applications in speech recognition. In Proc. IEEE, volume 77, pages 257-286, February 1989.

[LEC 98] Le Cun Y. MNIST Internet database *http://yann.lecun.com/exdb/mnist/*