

## **Apprentissages discriminants en reconnaissance de mots cursifs en-ligne.**

Emilie CAILLAULT, Christian VIARD-GAUDIN  
IRCCyN – UMR CNRS 6597, Image et Video Communications  
Ecole polytechnique de l'université de Nantes  
Rue Christian Pauc, BP 50609, FR-44306 NANTES cedex 3  
Mel : {emilie.caillault,christian.viard-gaudin}@univ-nantes.fr  
tel. : 02 40 68 32 36 – fax : 02 40 68 32 32

### ***Auteur principal :***

Emilie CAILLAULT (POISSON)

### ***Thème(s) choisi(s) :***

- 4. Décision et Interprétation
- 4.7 Réseaux de neurones, techniques floues, théories non probabilistes
- 4.4 Reconnaissance des formes

### ***Problème traité :***

Nous présentons un système de reconnaissance de mots cursifs en-ligne basé sur une architecture hybride TDNN-HMM.

### ***Originalité du travail :***

Nous avons construit une fonction de coût générique qui permet un apprentissage global du système, de type discriminant ou non, au niveau mot et au niveau caractère.

### ***Résultats nouveaux :***

Nous présentons la comparaison des différentes combinaisons possibles à partir de la fonction générique sur la base IRONOFF.

### ***Mots-clés :***

Hybride TDNN-HMM, Reconnaissance en-ligne, apprentissage global discriminant, critères MLE, MMI.

## 1. Introduction

Nos travaux s'intègrent dans le contexte de la reconnaissance de l'écriture en ligne destiné aux systèmes mobiles communicants (assistant numérique personnel, ardoise électronique, smart-phone). Dans ce domaine, les systèmes de reconnaissance mot les plus contraints imposent une écriture de type script [Oud 04][Rag 03] facilitant ainsi le problème de segmentation. Les systèmes supportant une écriture cursive sans contrainte conduisent à une architecture nettement plus complexe, c'est précisément le cas du système présenté ici. Nous avons opté pour un système hybride neuromarkovien avec une architecture de type TDNN (Time Delay Neural Network) pour le réseau de neurones. Si de tels systèmes ont déjà été abordés dans la littérature [Sch 95] [Jae 00] [Wim 00], leurs apprentissages posent encore de nombreux problèmes. Nous proposons ici une méthode d'apprentissage global au niveau mot basée sur une fonction de coût générique qui permet de mixer apprentissage de type discriminant et apprentissage de type modèle générateur. La solution proposée représente une simplification intéressante par rapport aux approches nécessitant plusieurs étapes, quelques fois alternées, pour mener à bien l'apprentissage du niveau graphème ou lettre au niveau mot.

## 2. Architecture globale du système TDNN+HMM

La figure 1 illustre le système de reconnaissance global [Poi 04]. Il est basé sur une approche analytique avec segmentation implicite et un apprentissage global au niveau mot.

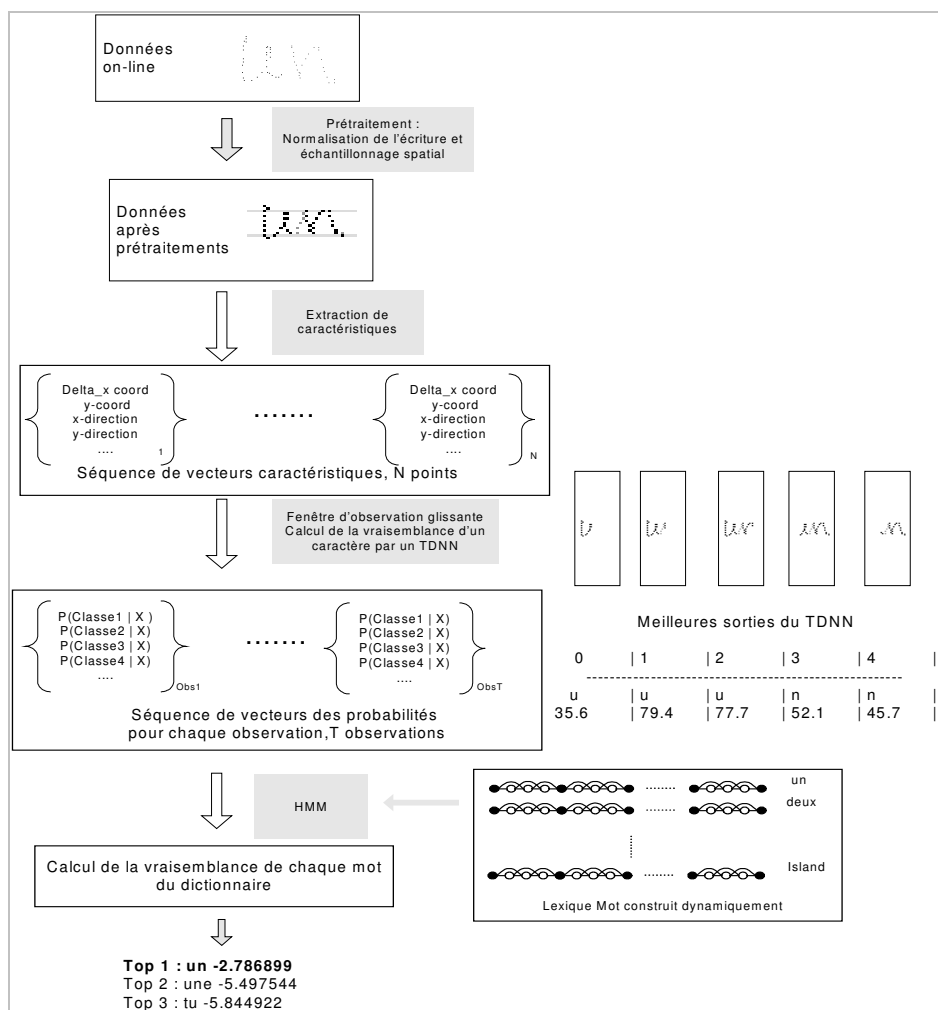


Figure 1 : Processus de reconnaissance du système ThDNN+HMM (cas du mot « un »)

De plus, il permet de manipuler un lexique de taille variable et ne requiert aucun apprentissage additionnel pour ajouter de nouvelles entrées au dictionnaire mot.

La première étape consiste à normaliser le signal d'écriture, principalement en taille, orientation des lignes de base et s'affranchir de la vitesse d'écriture. Une fois le signal normalisé, sont extrait sept caractéristiques standard pour chaque point qui seront présentées en entrée du système TDNN+HMM. Le TDNN balaie régulièrement la trajectoire du signal d'écriture avec un pas de segmentation implicite et une fenêtre d'une largeur donnée, elle-même balayée par une fenêtre de convolution de taille fixée à la hauteur du corps du mot. Pour chaque position du TDNN, celui-ci calcule un vecteur  $O_t$  correspondant aux probabilités a posteriori des classes associées à ses différentes sorties (66 classes : 2\*26 lettres + 14 extensions), elles représentent les probabilités d'observations du HMM. Le HMM effectue la tâche de reconnaissance au niveau mot en trouvant le meilleur alignement temporel (programmation dynamique) sur les différents modèles mots du lexique.

Avec une telle approche, une segmentation du mot en lettres est obtenue à l'issue de la reconnaissance, elle correspond aux changements de modèles-lettres sur le meilleur chemin du treillis de reconnaissance du modèle-mot considéré (Viterbi). Le mécanisme d'apprentissage associé à ce système tire profit de l'étiquetage des observations issues du TDNN résultant de cette segmentation implicite. Celui-ci est décrit dans la section suivante.

### 3. Apprentissages discriminants au niveau mot

L'objectif ici est de rétro-propager directement l'erreur au niveau mot dans le réseau de neurones pour mettre à jour ses paramètres en se basant sur l'étiquetage implicite. Pour cela, nous avons construit une fonction de coût générique  $L_G$  combinant à la fois un critère ML (Maximum Likelihood) et un critère MMI (Maximum Mutual Information) permettant de discriminer le modèle mot HMM reconnu à partir d'un lexique et/ou les sorties du TDNN mal classées.

$$L_G = (1 + \varepsilon) \log P(O | \lambda_{HMMvrai}) - \beta \times \left[ (1 - \alpha) \log P(O | \lambda_{HMMreconnu}) + \alpha \log P(O | \lambda_{TDNNreconnu}) \right] \quad (1)$$

avec  $HMMreconnu = \arg \max_i P(O | \lambda_i)$

Les paramètres  $\alpha$ ,  $\beta$ , et  $\varepsilon$  sont compris entre 0 et 1. Avec  $\varepsilon = \beta = 0$ , nous obtenons le critère classique du maximum de vraisemblance (ML). Avec  $\beta = 1$ , on introduit un apprentissage discriminant (MMI simplifié) qui prend en compte uniquement le mot reconnu en 1ère position par l'algorithme de Viterbi ( $\lambda_{HMMreconnu}$ ) si  $\alpha = 0$ , et dans le cas où  $\alpha = 1$  seulement les classes reconnues en 1<sup>ère</sup> position en sortie du TDNN ( $\lambda_{TDNNreconnu}$ ). La valeur  $\alpha$  permet de pondérer ces deux derniers types de discrimination.

Par souci de place, nous ne détaillerons pas les calculs permettant d'aboutir au gradient de l'erreur  $L_G$  donné ci-dessous pour la couche de sortie du TDNN, les couches cachées suivant l'algorithme de rétro-propagation classique :

$$\frac{\partial L_G}{\partial w_{ji}} = \sum_t \delta_{j,t} \cdot x_i(O_t) \text{ avec } \delta_{j,t} = Grad_{j,t} - x_{j,t} \sum_k Grad_{k,t} \quad (2)$$

et  $Grad_{j,t} = \left( (1 + \varepsilon) \frac{P(O, q_t = j | \lambda_{HMMvrai})}{P(O, \lambda_{HMMvrai})} - \beta * \left[ (1 - \alpha) \frac{P(O, q_t = j | \lambda_{HMMreconnu})}{P(O, \lambda_{HMMreconnu})} + \alpha \frac{P(O, q_t = j | \lambda_{TDNNreconnu})}{P(O, \lambda_{TDNNreconnu})} \right] \right)$

Où  $j$  est la position du neurone considéré,  $i$  celle du neurone associé de la couche inférieure,  $t$  l'indice temporel de l'observation et  $x_j(O_t)$  la sortie du neurone  $j$  de la couche de sortie avec  $x_j(O_t) = b_j(O_t)$  selon la notation courante des HMMs :  $\lambda(A, B, \pi)$ .  $P(O, q_t = j | \lambda)$  est calculé par

programmation dynamique. Ainsi pour chaque observation  $O_t$ , un gradient positif est rétropropagé pour le vrai modèle HMM (dont une fraction  $\varepsilon$  correspond à un pur critère ML) et un gradient négatif pour le modèle HMM reconnu et /ou pour les classes reconnues par le TDNN.

#### 4. Résultats et conclusion

Nous avons testé différentes combinaisons du critère précédent sur la base IRONOFF [Via 99]. Un sous-ensemble de 20 898 mots, représentant 197 labels différents, est utilisé pour l'apprentissage et un ensemble séparé de 10 448 mots est testé en généralisation. Le tableau suivant illustre les résultats obtenus selon le critère et la pondération de discrimination choisie.

Critère	MMI_HMM $\varepsilon=0 \beta=1 \alpha=0$	ML + MMI_HMM $\varepsilon=1 \beta=1 \alpha=0$	ML + MMI_TDNN $\varepsilon=1 \beta=1 \alpha=1$	Mixte $\varepsilon=1 \beta=1 \alpha=0.5$
<b>Taux en apprentissage</b>	83.92	83.82	79.73	87.09
<b>Taux en généralisation</b>	78.09	81.30	77.36	83.42

Tableau 1 : Comparaison des taux de reconnaissance sur la base IRONOFF.

Ce tableau montre l'intérêt de renforcer le modèle vrai à chaque itération par rapport à un critère de discrimination simple où le gradient est nul si le modèle vrai correspond au modèle reconnu. Par ailleurs un point important apparaît : une discrimination au niveau caractère ( $\alpha=0.5$ ) en plus d'une discrimination au niveau mot accroît les performances du reconnaiseur mot. Il est intéressant de souligner la convergence constatée expérimentalement de ce système d'apprentissage en dehors de toute initialisation spécifique du TDNN et de toute labellisation explicite au niveau caractère.

#### Références

- [Jae 00] S. Jaeger, S. Manke, J. Reichert, A. Waibel, "On-Line Handwriting Recognition: The NPen++ Recognizer", *International Journal on Document Analysis and Recognition, IJDAR'00*, volume 3, p. 169-180, 2000.
- [Oud 04] L. Oudot, L. Prevost L. & M. Milgram, *Un modèle d'activation vérification pour la lecture de textes manuscrits dynamique*, CIFED'04 (Confédération sur l'écrit et le document), La Rochelle, France, 2004.
- [Poi 04] E. Poisson, C. Viard-Gaudin, P.-M. Lallican. « Système TDNN/HMM de reconnaissance de mots cursifs en ligne à apprentissage simplifié », CIFED'04 - Colloque International Francophone sur l'écrit et le Document, La Rochelle, Juin 2004.
- [Rag 03] N. Ragot, E. Anquetil, "A generic hybrid classifier based on hierarchical fuzzy modeling experiments on on-line handwritten character recognition", *Seventh International Conference on Document Analysis and Recognition (ICDAR)*, UK, vol. 2, p. 963-967, 2003.
- [Sch 95] M. Schenkel, I. Guyon, D. Henderson. *On-line cursive script recognition using Time Delay Neural Networks and Hidden Markov Models*. Machine Vision and Applications, special issue on Cursive Script Recognition, (8):215--223, 1995.
- [Via 99] C. Viard-Gaudin, P.M. Lallican, S. Knerr, P. Binter, «The IRONOFF Dual Handwriting Database», *ICDAR'99*, pp. 455-458. Bangalore, Sept. 1999.
- [Wim 00] Z. Wimmer, S. Garcia-Salicetti, A. Lifchitz, B. Dorizzi, P. Gallinari, T. Artières, « REMUS », <http://www-connex.lip6.fr/~lifchitz/Remus/> .