

PROPOSITION DE SUJET DE THESE – Février 2010

Titre du sujet : Reconnaissance de documents multilingues par Réseaux Bayésiens Dynamiques et Réseaux de Neurons Récurrents

Comité d'encadrement :	Laurence Likforman-Sulem (Telecom ParisTech), directrice de thèse Tél. : 01 45 81 73 28 Courriel : laurence.likforman@telecom-paristech.fr 46 rue Barrault, 75013 Paris Emmanuèle Grosicki (DGA) co-directrice de thèse Courriel : emmanuele.grosicki@etca.fr Marc Sigelle (Telecom ParisTech) Courriel : marc.sigelle@telecom-paristech.fr
Etablissement d'inscription :	Telecom ParisTech (Ecole Doctorale EDITE)
Laboratoire :	Telecom ParisTech/Dept Traitement du Signal et des Images CNRS LTCI UMR

Contexte et objectifs

La reconnaissance de la parole et de l'écriture sont des domaines de recherche très actifs depuis une vingtaine d'années. Les recherches dans ces deux domaines ont fait émerger des modèles comme les modèles de Markov cachés (HMMs), les réseaux Bayésiens dynamiques et les réseaux de neurones récurrents. Les modèles HMMs notamment sont très présents dans les systèmes de reconnaissance uni

lingues (lecture d'adresses postales, de montants de chèques, de formulaires). Les modèles HMMs pour la reconnaissance de l'écriture manuscrite, sont très proches de ceux développés initialement pour la reconnaissance de la parole. Le passage du domaine de la parole au domaine de l'écrit nécessite cependant de prendre en compte les spécificités des deux domaines (signal mono-dimensionnel pour la parole, bi-dimensionnel pour l'écriture).

L'application d'une fenêtre d'analyse glissant le long de l'image du caractère ou du mot à reconnaître permet d'obtenir une séquence de signal mono-dimensionnelle. Cependant la notion de contexte spatial en dimension 2 est perdue le long de cette séquence. Le formalisme des réseaux Bayésiens permet de re-introduire la notion de contexte spatial. Un réseau Bayésien est un modèle graphique probabiliste dans lequel des liens sont introduits entre observations distantes. De tels modèles ont été proposés par (Likforman-Sulem et Sigelle, 2008) pour la reconnaissance de caractères dégradés. De même les modèles basés sur les réseaux de neurones récurrents étendent les réseaux de neurones classiques en introduisant la notion de contexte par des dépendances bidirectionnelles. En particulier le modèle LSTM (Long Short Term Memory) a été récemment appliqué à la reconnaissance d'écriture cursive en-ligne (Graves et al. 2009).

Dans la continuité des travaux de K. Hallouli [Hallouli, 2002], et de R. El-Hajj [El-Hajj, 2008], nous voulons faire progresser la reconnaissance de mots dégradés imprimés et manuscrits dans un contexte multilingue en développant les axes suivants :

1. fusion des modélisations HMM et réseaux Bayésiens dynamiques pour reconnaître des mots imprimés dégradés ou cursifs (français, anglais, arabe). La segmentation implicite

proposée par le modèle HMM permettra d'obtenir les observations bi-directionnelles propres à un caractère.

2. modélisation et paramétrisation d'une approche de type réseaux de neurones récurrents.

3. comparaison des deux approches contextuelles, réseaux Bayésiens dynamiques et réseaux récurrents. La comparaison portera sur les aspects théoriques liés à la notion de contexte, ainsi que sur la complexité algorithmique de chaque approche.

4. apprentissage de modèles de mots dans un contexte multilingue.

Evaluation

L'évaluation des techniques de reconnaissance développées se fera sur des bases publiques. Les expériences seront réalisées sur les mots isolés dans différents contextes de vocabulaire (vocabulaire d'une dizaine, centaine, milliers de mots). La principale base publique utilisée sera la base issue du projet TechoVision RIMES collectée en 2008 par la DGA et ses partenaires (Grosicki et al. 2009). Cette base est constituée de 3500 lettres manuscrites et de documents dégradés tels que les télécopies, pour lesquels une transcription manuelle est disponible. La base Rimes est la base la plus réaliste de documents manuscrits et de télécopies disponible en français et la plus grande base en Français actuellement disponible. Elle inclut des scénarios clients s'adressant à une entreprise ou une administration. Pour les langues autres que le français, les bases publiques en anglais (IAM) et en arabe (IFN/ENIT et Université de Maryland) seront utilisées.

La thèse se déroulera au sein de l'unité CNRS LTCI à Telecom ParisTech et sera financée par la DGA (Direction Générale de l'Armement).

Références

R. El-Hajj, L. Likforman-Sulem, C. Mokbel, Combining slanted-frame classifiers for improved HMM-based Arabic handwriting recognition, *IEEE PAMI*, 31(7): 1165-1177, 2009.

A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke, J. Schmidhuber. A Novel Connectionist System for Improved Unconstrained Handwriting Recognition. *IEEE PAMI*, vol. 31, no. 5, 2009.

E. Grosicki, M. Carré, J-M. Brodin, E. Geoffrois, Results of the RIMES Evaluation Campaign for Handwritten Mail Processing. ICDAR 2009: 941-945

K. Hallouli, L. Likforman-Sulem, M. Sigelle: A Comparative Study between Decision Fusion and Data Fusion in Markovian Printed Character Recognition. ICPR (3) 2002: 147-150.

L. Likforman-Sulem, M. Sigelle, Recognition of degraded characters using Dynamic Bayesian Networks, *Pattern Recognition*, Vol. 41, pp. 3092-3103, 2008,

PyBrain Machine Learning Library : <http://pybrain.org/>