

Sujet de master recherche « Architectures logicielles distribuées » 2006- 2007

## Grammaire Hors- Contexte Probabiliste pour la Reconnaissance de Formules Mathématiques Manuscrites

Encadrant principal :

Christian VIARD- GAUDIN

Courriel : [Christian.Viard- Gaudin@univ-nantes.fr](mailto:Christian.Viard-Gaudin@univ-nantes.fr)

Tél : 02.40.68.30.40

Ecole Polytechnique de l'Université de Nantes

### Contexte du stage :

Les éditeurs d'équations mathématiques sont aujourd'hui systématiquement associés à une interface de saisie de type clavier/souris. Nous voulons étudier et proposer des voies originales pour permettre une saisie manuscrite des formules mathématiques.

Ce sujet de recherche rentre dans le cadre plus général de la reconnaissance de l'écriture manuscrite en- ligne, où les dispositifs de saisie sont les tablettes PC ou les papiers numériques couplés à l'utilisation de stylos digitaux.

### Travail à réaliser :

La problématique envisagée peut se définir selon trois niveaux :

- Une tâche de segmentation/reconnaissance des symboles élémentaires que contient la formule : il s'agit d'affecter un label symbolique aux différents éléments du tracé. Ceci nécessite la segmentation de ce tracé. Pour cela des outils de reconnaissance de formes (Réseaux de neurones, en particulier) seront utilisés.
- Une tâche d'analyse de la structure 2D des éléments précédemment extraits consistant à décrire les positions relatives des composantes.
- Une tâche d'analyse syntaxique permettant de reconstruire la formule mathématique correspondante. Cette tâche nécessite, à partir des labels symboliques et de leur disposition spatiale de construire un modèle, dans un cadre stochastique, qui permette de procéder à l'interprétation de la formule. Plusieurs formalismes seront étudiés pour cela, une attention particulière sera portée aux grammaires hors contextes probabilistes (SCFG : Stochastic Context- Free Grammar).

Mots- clés : *Ecriture manuscrite en- ligne, segmentation, reconnaissance des formes, apprentissage, réseaux bayésiens, parser.*

Pré- requis : bonne maîtrise des environnements de développement en Java ou C++.

Possibilité de thèse en continuité de ce travail.

## Références :

- K. F. Chan and D. Y. Yeung. *Mathematical expression recognition: A survey*. International Journal on Document Analysis and Recognition, 3(1):3-15, Aug. 2000. (<http://citeseer.ist.psu.edu/93100.html>)
- Trevor Hastie, Robert Tibshirani, Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics
- F. V. Jensen. "**Bayesian Networks and Decision Graphs**". Springer. 2001.
- R. Yamamoto, S. Sako, T. Nishimoto; S. Sagayama. On-Line Recognition of Handwritten Mathematical Expressions Based on Stroke-based Stochastic Context-Free Grammar. Tenth International Workshop on Frontiers in Handwriting Recognition (IWFHR 10), p. 249- 254. La Baule, 2006.