



Laboratoire d'Informatique de Paris 6

Université Pierre et Marie Curie - Paris VI - CNRS

**Rapport sur le mémoire d'habilitation présenté par Mr F. ROSSI,
intitulé
« Contribution à l'analyse des données complexes ».**

Fabrice Rossi est depuis 1997 CR à l'INRIA après avoir été Maître de Conférences à Paris Dauphine pendant 6 ans. Son domaine de recherche concerne l'apprentissage statistique. Les travaux qu'il présente dans son mémoire concernent l'étude de méthodes d'apprentissage pour le traitement de données non vectorielles et plus particulièrement de données fonctionnelles et de données représentées par des similarités.

Le mémoire est court et assez dense. Il donne une vision synthétique des principaux résultats récents de F. Rossi, il est complété par une série d'articles.

Il est composé de deux grandes parties. La première concerne l'analyse de données fonctionnelles. Les exemples classiques de l'apprentissage sont ici remplacés par une description des données sous la forme de fonctions. Ce cadre permet donc d'étudier des données bien plus complexes que les représentations attribut-valeur traditionnelles. Le travail de F. Rossi a consisté à étendre dans le cas de la régression, les modèles traditionnels comme les perceptrons multi-couches, les cartes topologiques et de façon plus limitée les machines à noyau, pour traiter des données fonctionnelles. C'est un domaine qui a été relativement exploré en apprentissage statistique et qui ouvre la porte à de nombreuses perspectives.

L'intérêt de l'étude de données fonctionnelles est assez bien motivé dans le manuscrit. La part la plus importante du travail présenté concerne sûrement les systèmes inspirés des perceptrons multicouches. Dans ces systèmes, un neurone fonctionnel prend en entrée une fonction et la transforme par une forme linéaire. Un réseau de neurone sera alors une combinaison linéaire de ces unités fonctionnelles. F. Rossi montre que ces modèles possèdent des capacités d'approximateurs universels. Il s'appuie pour cela sur une approximation des neurones fonctionnels par des approximateurs classiques (réseaux de neurones, machines à noyaux etc). Au-delà de ce résultat d'existence, il faut proposer des algorithmes d'apprentissage pour entraîner ces modèles. La forme théorique demande d'une part de connaître la forme fonctionnelle exacte des entrées et d'autre part de réaliser des calculs d'intégrale pour chacune de ces fonctions. En pratique, les

Surface address : University Paris VI, 8 rue du Capitaine Scott, 75015 Paris, France

Tel : (33-1) 44 27 73 70 - Fax (33-1) 44 27 70 00 - E-mail: Patrick.Gallinari@lip6.fr

fonctions ne seront connues que par un ensemble de réalisations (valeurs de ces fonctions sur un ensemble de points) et les intégrales ne seront pas calculables. F. Rossi propose une version empirique du neurone fonctionnel où les fonctions d'entrées sont remplacées par des valeurs issues d'un échantillon et les paramètres des neurones fonctionnels sont des fonctions paramétriques. L'intégrale définissant le neurone théorique est ainsi remplacée par une approximation qui est une fonction paramétrique des entrées discrètes. Ces approximations permettent de définir un processus d'apprentissage.

F. Rossi montre la convergence à la limite des paramètres ainsi obtenus vers les paramètres optimaux. Il montre ensuite comment l'algorithme peut être mis en œuvre de façon simple dans le cas d'une fonction paramétrique linéaire.

En continuant sur l'idée de ces modèles fonctionnels, F. Rossi a exploré une seconde voie qui consiste à remplacer les fonctions d'entrée par des projections sur un sous-espace fonctionnel, ce qui offre une représentation des données plus facile à manipuler. De même que précédemment, il définit un neurone fonctionnel sur ces espaces de projection et montre dans ce cas particulier les propriétés d'approximateur universel des réseaux obtenus. Il décrit également comment estimer les paramètres.

Cette idée de projection dans une base de fonctions est ensuite étendue au cas d'autres modèles comme les réseaux à fonctions radiales et les cartes auto-organisatrices. Pour ces méthodes qui utilisent des calculs de normes, F. Rossi propose des techniques opérationnelles pour apprendre les bonnes bases de projection.

L'ensemble des méthodes proposées est testé sur plusieurs types de données et montre de bonnes performances, malgré une complexité importante des algorithmes (par rapport aux méthodes standard qui travaillent en vectoriel).

Enfin, F. Rossi a également commencé à travailler sur les machines à noyau, cette fois-ci dans le cadre de la classification à 2 classes. Il propose en s'appuyant sur une formulation régularisée du problème d'apprentissage dans les machines à noyau de combiner des opérateurs de transformations fonctionnelles avec des noyaux. La mise en œuvre pratique est similaire à celle déjà introduite pour les modèles neuronaux.

La deuxième partie du manuscrit est consacrée à l'analyse de tableaux de dissimilarités. L'analyse de données complexes peut se ramener à l'utilisation de méthodes classiques si on est capable de définir des similarités sur ces données complexes. La littérature récente sur les noyaux est riche de tels exemples de similarités définies sur des séquences, des arbres, des graphes. Dans ce cadre, une première contribution de F. Rossi a consisté à proposer une extension des cartes auto-organisatrices au cas des tableaux de dissimilarités. Sa méthode repose sur la représentation d'un prototype par un ensemble de données d'apprentissage et l'utilisation d'une fonction de coût adéquate. Il propose une implantation très efficace de cet algorithme. Ces idées sont ensuite utilisées dans le cadre de l'analyse de traces utilisateurs pour des données web. A partir de la définition de dissimilarités adaptées à ces traces, il analyse le comportement utilisateurs sur plusieurs sites web.

Ces travaux ont été réalisés en collaboration avec des collègues et des étudiants en thèse.

Les travaux de F. Rossi sont originaux. Ils explorent un aspect encore relativement vierge de l'apprentissage statistique qui est l'analyse de données fonctionnelles. Ils devraient

Surface address : University Paris VI, 8 rue du Capitaine Scott, 75015 Paris, France

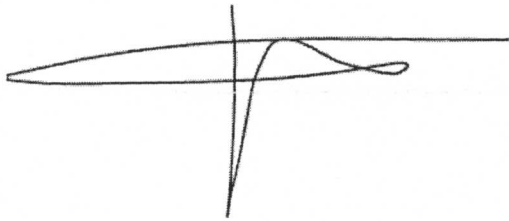
Tel : (33-1) 44 27 73 70 - Fax (33-1) 44 27 70 00 - E-mail: Patrick.Gallinari@lip6.fr

permettre d'apporter un nouveau regard sur des approches récentes dans le domaine de l'apprentissage statistique, qui s'intéressent à travailler de façon effective avec des données complexes. F. Rossi a encadré ou co-encadré 3 thèses. Les travaux ont donné lieu à de nombreuses publications (8 journaux internationaux, une vingtaine de conférences internationales et plusieurs publications en Français).
Compte tenu de la qualité des travaux présentés, je donne un avis favorable à leur présentation pour l'habilitation à diriger des recherches de l'université Paris Dauphine.

Paris le 11-11-2006

P. GALLINARI

Professeur

A handwritten signature in black ink, consisting of a horizontal line with a vertical stroke intersecting it, and a series of loops and curves extending downwards and to the right.

Surface address : University Paris VI, 8 rue du Capitaine Scott, 75015 Paris, France

Tel : (33-1) 44 27 73 70 - Fax (33-1) 44 27 70 00 - E-mail: Patrick.Gallinari@lip6.fr