Classification de fonctions continues à l'aide d'une distribution et d'une densité définies dans un espace de dimension infinie

Etienne Cuvelier*, Monique Noirhomme-Fraiture*

* Institut d'Informatique
Facultés Universitaires Notre-Dame de la Paix (FUNDP)
Namur, Belgique
{cuvelier.etienne,noirhomme.monique}@info.fundp.ac.be

Résumé. Il n'est pas rare que des données individu soient caractérisées par une distribution continue et non une seule valeur. Ces données fonctionnelles peuvent être utilisées pour classer les individus. Une solution élémentaire est de réduire les distributions à leurs moyennes et variances. Une solution plus riche a été proposée par Diday (2002) et mise en oeuvre par Vrac et al. (2001) et Cuvelier et Noirhomme-Fraiture (2005). Elle utilise des points de coupures dans les distributions et modélise ces valeurs conjointes par une distribution multidimensionnelle construite à l'aide d'une copule. Nous avons montré dans un précédent travail que, si cette technique apporte de bons résultats, la qualité de la classification dépend néanmoins du nombre et de l'emplacement des coupures. Les questions du choix du nombre et de l'emplacement des coupures restaient des questions ouvertes. Nous proposons une solution à ces questions, lorsque le nombre de coupures tend vers l'infini, en proposant une nouvelle distribution de probabilité adaptée à l'espace de dimension infinie que forment les données fonctionnelles. Nous proposons aussi une densité de probabilité adaptée à la nature de cette distribution en utilisant la dérivée directionnelle de Gâteaux. La direction choisie pour cette dérivée est celle de la dispersion des fonctions à classer. Les résultats sont encourageants et offrent des perspectives multiples dans tous les domaines où une distribution de données fonctionnelles est nécessaire.

1 Introduction

En analyse de données symbolique (voir Bock et Diday (2000)) une variable peut, entre autre être décrite par une distribution de probabilité continue. La classification en K groupes de ces données fonctionnelles peut être obtenue en utilisant une décomposition de mélange. Mais cette technique nécessite de pouvoir calculer la densité d'une distribution de fonction. Or l'espace des fonctions n'est pas un espace de dimension finie, tels que ceux où sont définies les distributions classiques. Projeter les fonctions dans un espace multidimensionnel par échantillonnage (voir Diday (2002)) permet de contourner ce problème, pour autant que l'on choisisse des distributions conjointes adéquates. Dans la section 2 de cet article nous rappel-