

# Classification en référence à une matrice stochastique

Stéphane Verdun, Véronique Cariou, El Mostafa Qannari

Unité de Sensométrie et Chimiométrie  
ENITIAA, rue de la Géraudière - BP82 225  
44322 Nantes Cedex 3 - France  
{prénom.nom}@enitiaa-nantes.fr

**Résumé.** Étant donné un tableau de données portant sur un ensemble d'objets et une matrice stochastique qui peut être assimilée à une matrice de passage d'une marche aléatoire, nous proposons une méthode de partitionnement consistant à identifier les différents états stationnaires associés à la matrice stochastique. Les classes formant la partition sont déterminées à partir de ces états stationnaires. De manière pratique, la matrice stochastique peut être définie à partir d'une matrice de mesure de ressemblance entre les objets du tableau de données. Différentes mesures sont étudiées (plus proches voisins, noyaux de densité...). L'approche est présentée sur la base des propriétés des graphes et illustrée sur des données simulées. La méthode de classification proposée est également comparée à une autre approche de classification, la classification spectrale.

## 1 Introduction

Une méthode de classification basée sur une matrice stochastique est proposée. La matrice stochastique reflète une affinité entre les objets à classer et sera généralement déduite d'une matrice de mesure de ressemblance. L'intérêt de cette méthode de classification est sa grande flexibilité du fait de la prise en compte de différentes mesures de ressemblance (mesures de similarités, mesures de densité, ...). Il est ainsi possible de l'adapter à différents contextes (e.g. classification d'individus, classification de variables) et à plusieurs configurations (séparations linéaires ou non linéaires des classes). Cependant, cette méthode de classification exige que les classes soient bien séparées et qu'il n'y ait pas d'objets qui créent un effet de chaînage. C'est pourquoi une étape consistant à rechercher les individus isolés s'avère généralement nécessaire.

La démarche de classification proposée est intimement liée aux marches aléatoires sur un graphe (Lovasz, 1993) car les classes sont déterminées en considérant la matrice de convergence de la suite des puissances de la matrice stochastique. En effet, cette suite converge vers une matrice qui permet d'identifier les différentes classes. De plus, la méthode de classification permet d'associer à chaque objet une probabilité qui indique son degré de centralité dans la classe à laquelle il appartient. Il faut également souligner que tous les objets ne sont pas affectés de manière déterministe aux différentes classes. En effet, certains objets qui ne constituent pas les "noyaux forts" des classes sont caractérisés par un vecteur de probabilité qui