

Une nouvelle approche visuelle pour la classification hiérarchique et topologique

Hanane Azzag*, Mustapha Lebbah *

* Université Paris 13, LIPN UMR 7030
99, avenue Jean-Baptiste Clément
93430 Villetaneuse, France
prénom.nom@lipn.univ-paris13.fr

Résumé. Nous proposons dans cet article une nouvelle méthode de classification hiérarchique et topologique. Notre approche consiste à construire de manière auto-organisée une partition de données représentées par un ensemble "forêt" d'arbres répartis sur une grille 2D. Chaque cellule de la grille est modélisée par un arbre dont les noeuds représentent les données. La partition globale obtenue est visualisée à l'aide d'une carte de TreeMap dans laquelle chaque TreeMap représente un arbre de données. Nous évaluerons les capacités et les performances de notre approche sur des données aux difficultés variables. Des résultats numériques et visuels seront présentés et discutés.

1 Introduction

Actuellement, la quantité d'informations stockées dans les bases de données est en croissance rapide. A cela se rajoute la complexité de la structure des données à manipuler. En effet ces données ne sont pas uniquement composées d'attributs numériques ou symboliques, mais peuvent être enrichi, par exemple, par des sons, images, vidéos, textes ou des sites Web. Trouver l'information pertinente cachée dans ces ensembles de données structurées est une tâche difficile. Les méthodes de classification et de visualisation sont des techniques qui sont souvent utilisées pour l'analyse des données structurées. L'objectif principal de ce papier est de développer des méthodes qui permettent aux experts du domaine d'analyser de manière intuitive les ensembles de données structurées.

Souvent, l'analyse intuitive des données implique l'utilisation de méthodes de classification hiérarchique combinées à des techniques de visualisation Jain et Dubes (1988); Jain et al. (1999); Handl et al. (2003). Dans la littérature il existe deux types de méthodes de classification hiérarchique : divisives et agglomératives. Les méthodes divisives sont celles qui partent d'un ensemble d'individus et procèdent par divisions successives des classes jusqu'à l'obtention de la partition la plus fine. En revanche, les méthodes ascendantes partent des singletons et procèdent par agrégations successives. Dans cet article nous proposons une nouvelle approche de classification hiérarchique qui permet de construire des forêts d'arbres et de manière auto-organisée.

Les modèles des cartes auto-organisatrices (Self-Organizing-Map) sont souvent utilisées pour la visualisation et la classification topologique non supervisé Kohonen (2001). Des ex-