

# Un cadre graphique pour la visualisation et la caractérisation de classes en mode non-supervisé

Khalid Benabdeslem\*, Haytham Elghazel\*  
Rakia Jaziri\*\*

\*Université Lyon1, LIESP EA 4125  
43, Bd du 11 Novembre 1918 Villeurbanne Cedex  
{kbenabde, elghazel}@bat710.univ-lyon1.fr  
\*\*Université Paris13, LIPN, UMR CNRS 7030  
99, Av Jean Baptiste Clément 93430 Villetaneuse  
rakia.jaziri@lipn.univ-paris13.fr

**Résumé.** Dans ce papier, nous proposons un système d'extraction de connaissances à partir de données pour la visualisation et l'interprétation de profils issus d'une classification automatique en mode non supervisé. Une méthode efficace utilisée pour la classification est la carte auto-organisatrice ou Self-Organizing Map (SOM). Dans cette méthode, la détection des regroupements est en général obtenue par d'autres techniques de classification (par partitionnement ou hiérarchisation). Dans le cadre de ce travail, nous explorons une autre voie pour la segmentation de la carte par une approche basée sur la théorie des graphes (la coloration minimale). Enfin, pour caractériser les classes issues de cette segmentation, nous proposons une solution basée sur un test statistique et un arbre couvrant maximum pour la sélection de variables locales à chaque classe permettant ainsi sa caractérisation de manière automatique. Des expérimentations seront présentées sur plusieurs bases de données pour valider l'approche proposée.

## 1 Introduction

La classification automatique (Clustering) est une étape primordiale dans les problèmes liés à l'apprentissage non supervisé. Elle consiste à détecter la meilleure structure à partir d'une collection de données non étiquetées. Dans ce cadre, son principe revient à organiser les données en groupes (clusters) dits homogènes. Les algorithmes de classification peuvent être regroupés en cinq familles : par partitionnement, hiérarchiques, basés sur la densité, basés sur les grilles et ceux basés sur les modèles (Jain et Murty, 1999). Dans ce papier, nous nous basons sur la classification par modèles et plus particulièrement celui des cartes auto-organisatrices de Kohonen (Kohonen, 2001). Connue sous le nom de SOM (Self Organizing Map) ce modèle repose sur un algorithme neuro-inspiré qui, par un processus non-supervisé compétitif, est capable de projeter des données de grandes dimensions dans un espace à faible dimension (en général deux). Ce modèle est populaire et très utilisé pour la visualisation des données dans plusieurs domaines d'application.