

# Une méthode de classification visuelle et interactive

David Da Costa<sup>\*,\*\*</sup>, Gilles Venturini<sup>\*</sup>

<sup>\*</sup>Laboratoire d'Informatique  
Ecole Polytechnique de l'Université de Tours  
64, Avenue Jean Portalis, 37200 Tours, France.  
{david.dacosta, venturini}@univ-tours.fr  
<http://www.antsearch.univ-tours.fr/webrtic>

<sup>\*\*</sup>COHESIUM  
71-73 rue de Saussure, 75017 Paris, France.  
ddacosta@cohesium.com  
<http://www.cohesium.com/>

**Résumé.** L'objectif de notre travail est de pouvoir représenter visuellement des ensembles de données et de laisser l'expert du domaine procéder interactivement à la définition d'une classification de ces données. Notre approche se base sur l'existence d'une fonction de similarité afin de pouvoir traiter des données de tous types (numériques, symboliques, images, textes, etc). Elle permet à l'expert du domaine, grâce à l'utilisation d'une visualisation à base de points d'intérêt (POIs), de définir des classes dans les données grâce à des opérations de sélections graphiques. Nous comparons notre approche interactive avec la classification ascendante hiérarchique (CAH) sur un ensemble de bases classiques. Nous montrons qu'un expert des données peut arriver à des performances similaires à celle d'un algorithme automatique, tout en bénéficiant d'informations complémentaires sur les classes.

## 1 Introduction

La classification est une des tâches importantes de la fouille de données et a pour but d'affecter à chaque donnée d'un ensemble un label de classe (Jain et Dubes, 1988). La grande majorité des méthodes dans ce domaine utilise des approches non interactives dans lesquelles un algorithme produit des résultats de manière entièrement automatisée. Cependant, on peut noter que l'expert final est le seul à pouvoir valider les résultats, et il arrive même que certaines méthodes (telles que la CAH) nécessitent l'intervention de l'utilisateur pour définir de manière fiable un "bon" nombre de classes. Par ailleurs, les méthodes automatiques ne donnent pas systématiquement des informations directes sur la densité des classes, leurs formes, les relations de proximité entre classes : il s'agit souvent d'un travail supplémentaire d'interprétation de la part de l'expert.

Dans ce travail, nous proposons une méthode qui se place dans la lignée des méthodes visuelles et interactives qui vont solliciter l'expert du domaine et obtenir ainsi un résultat directement validé par l'utilisateur et des informations sur les classes obtenues. Notre objectif est