

# Méthode visuelle et interactive de partitionnement d'un ensemble de données à l'aide de graphes de voisinage construits par des fourmis artificielles

Julien Lavergne\*, Hanane Azzag\*\*  
Christiane Guinot\*,\*\*\*, Gilles Venturini\*

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

\*\*Laboratoire d'Informatique de l'Université Paris-Nord XIII  
CNRS - UMR 7030

99, avenue Jean-Baptiste Clément, 93430 Villetaneuse, France  
hanane.azzag@lipn.univ-paris13.fr,  
<http://www-lipn.univ-paris13.fr/A3/>

\*\*\*CE.R.I.E.S, 20 rue Victor Noir, 92521 Neuilly-Sur-Seine, France  
christiane.guinot@ceries-lab.com,  
<http://www.ceries.com>

**Résumé.** Nous présentons dans cet article une méthode de découverte visuelle et interactive d'un partitionnement de données qui s'appuie sur la visualisation d'un graphe de voisinage obtenu à l'aide de la méthode biomimétique AntGraph. Le but est de construire ce graphe en un temps de calcul très faible grâce aux principes de l'heuristique développée, puis de laisser l'expert du domaine procéder de manière interactive à la définition d'une classification sur l'ensemble de données considéré. Nous proposons une évaluation expérimentale de notre méthode interactive sur un panel d'utilisateurs experts et non-experts et comparons les résultats obtenus en terme de qualité avec une méthode de classification interactive visuelle à base de points d'intérêts, et deux méthodes automatiques de classification que sont la Classification Ascendante Hiérarchique et AntTree. Nous montrons finalement que l'utilisation d'une technique de visualisation et d'exploration interactive par des utilisateurs experts ou non sur des graphes de données construits par AntGraph permet de découvrir une classification ayant une qualité similaire à celles obtenues avec des méthodes interactives ou automatiques.

## 1 Introduction

La classification est un des domaines importants de la fouille de données (Jain et al., 1999). La majorité des méthodes de classification sont automatiques. Cependant, dans de nombreux cas, les résultats obtenus nécessitent une validation de la part de l'expert du domaine. D'autre part, les méthodes automatiques fournissent peu d'informations sur les classes (e.g. densité, forme, proximité). Les informations généralement retournées concernent le nombre de classes et l'appartenance des données à ces classes. Des actions supplémentaires de la part de l'expert du domaine peuvent être alors nécessaires pour améliorer les résultats. A ce titre, l'utilisation partielle ou exclusive de méthodes interactives de classification (donc non automatiques) permet à l'expert du domaine d'agir sur la classification. Les méthodes de visualisation associées permettent, conjointement avec des techniques d'interaction, d'explorer les ensembles de données et d'extraire plus facilement de nouvelles caractéristiques des classes et données. Nous proposons dans cet article une adaptation de notre méthode de construction de graphes de voisinage AntGraph (Lavergne, 2008) à la fouille visuelle de données pour des tâches de classification interactive. A ce titre, nous montrons que notre méthode offre une qualité satisfaisante de partitionnement d'un ensemble de données et qu'elle permet la découverte des propriétés topologiques d'un ensemble de données.