

# Positionnement multidimensionnel et partitionnement pour la visualisation de données multivariées

Antoine Naud

Département d'Informatique Appliquée  
Université N. Copernic  
ul. Grudziadzka 5  
87-100 Toruń, Pologne  
naud@phys.uni.torun.pl  
<http://www.phys.uni.torun.pl/~naud/>

Grâce aux réseaux informatiques, de plus en plus de données sont réunies et accessibles dans de grandes bases où la quantité de données disponibles crée un besoin en outils de synthèse. La visualisation de données multivariées par réduction de dimensionnalité permet à l'observateur humain la perception directe de structures ou de régularités existantes dans les données. Diverses techniques de réduction de la dimensionnalité des données ont été proposées, entre autres les réseaux de neurones CCA et GTM, et comparés du point de vue de la classification des données [De Backer *et al.*, 1998].

Le positionnement multidimensionnel (MDS) est une méthode d'Analyse Exploratoire de Données qui produit des configurations reflétant les relations de similarité existantes entre les données, et peut donc être appliquée à la visualisation de données multivariées. La complexité algorithmique du MDS limite son application à la visualisation d'ensembles de taille inférieure quelques milliers d'objets. Afin de répondre au besoin de visualisation d'ensembles plus importants, nous proposons d'associer le MDS à un partitionnement préalable des données. La visualisation se fait en trois étapes: 1) Partitionnement de l'ensemble des données pour former une Base, 2) Réduction de dimension de la Base (formée des centres de partitionnement) par MDS, 3) Ajout des données par *MDS relatif* [Naud et Duch, 2000].

Trois algorithmes: les *k-moyennes*, les *dendrogrammes*, et la *quantification vectorielle* ont été employés et comparés du point de vue de la configuration et du *Stress* résultants. L'approche proposée, testée sur l'ensemble *Satimage* provenant du site UCI, indique que les *k-moyennes* sont mieux adaptées à cette tâche que leurs concurrentes.

## Références

- [De Backer *et al.*, 1998] S. De Backer, A. Naud, et P. Scheunders. Non-linear dimensionality reduction techniques for unsupervised feature extraction. *Pattern Recognition Letters*, 19:711–720, 1998.
- [Naud et Duch, 2000] Antoine Naud et Włodzisław Duch. Interactive data exploration using MDS mapping. In *Proceedings of the Fifth Conference "Neural Networks and Soft Computing"*, pages 255–260, Zakopane, Poland, 2000.