

Prise en compte de connaissances pour la visualisation par l'intégration interactive de contraintes

Lionel Martin*, Matthieu Exbrayat*, Guillaume Cleuziou*, Frédéric Moal*

* Laboratoire d'Informatique Fondamentale d'Orléans
Université d'Orléans, BP 6759 F-45067 Orléans Cedex 2
Prenom.Nom@univ-orleans.fr
<http://www.univ-orleans.fr/lifo>

Résumé. La projection et la visualisation d'objets dans un espace à deux ou trois dimensions constituent une tâche courante de l'analyse de données. Cette opération induit des pertes dans le sens où certains objets peuvent se retrouver très proches alors qu'ils sont à l'origine assez éloignés. A partir de cette visualisation, il semble intéressant d'offrir à l'utilisateur la possibilité d'apporter de la connaissance sous forme de contraintes spécifiant les similarités attendues entre divers objets, lorsque ceux ci sont, dans l'espace d'observation, visuellement trop proches, ou au contraire trop éloignés.

Nous proposons ici trois types de contraintes et présentons une méthode de résolution de celles-ci dérivée de l'analyse en composantes principales (ACP). Deux types d'expérimentation sont présentées, reposant respectivement sur un jeu de données synthétique et sur des jeux standards. Ces tests montrent qu'une représentation de bonne qualité peut être obtenue avec un nombre limité de contraintes ajoutées.

1 Introduction

Les mécanismes d'apprentissage automatiques consistent en général à établir ce qui est important, vis-à-vis d'un objectif donné, au sein d'une masse d'information disponible. C'est notamment le cas dans le cadre de la classification automatique, qu'elle soit supervisée ou non. Les objets que l'on souhaite classer sont fréquemment définis par un ensemble de descripteurs, potentiellement nombreux, qui peuvent s'avérer redondants, bruités, ou tout simplement sans objet vis-à-vis de la classification recherchée. Par ailleurs, l'abondance de descripteurs peut rendre peu lisible le mécanisme de classification.

Deux approches sont envisageables afin de réduire le nombre de descripteurs. La première consiste à ne retenir que les descripteurs les plus significatifs (sélection d'attributs, ou *feature selection*, on pourra notamment consulter la synthèse de Guyon et Elisseeff (2003)) et la seconde à produire un ensemble restreint de descripteurs synthétiques, reflétant au mieux la répartition des objets dans l'espace d'origine (réduction de dimension, ou *dimensionality reduction*). Notons que la réduction de dimension est parfois considérée (e.g. dans Blum et Langley (1997)) comme appartenant à une catégorie de méthodes de sélection d'attributs appelée approche par filtre.