

Intégration interactive de contraintes pour la réduction de dimensions et la visualisation

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. Il existe aujourd'hui de nombreuses méthodes de réduction de dimensions, que ce soit dans un cadre supervisé ou non supervisé. L'un des intérêts de ces méthodes est de pouvoir visualiser les données, avec pour objectif que les objets qui apparaissent "visuellement" proches soient similaires, dans un sens qui correspond aux connaissances d'un expert du domaine ou qui soit conforme aux informations de supervision. Nous nous plaçons ici dans un contexte semi-supervisé où des connaissances sont ajoutées de façon interactive : ces informations seront apportées sous forme de contraintes exprimant les écarts entre la représentation observée et les connaissances d'un expert. Nous pourrions par exemple spécifier que deux objets proches dans l'espace d'observation sont en fait peu similaires, ou inversement. La méthode utilisée ici dérive de l'analyse en composantes principales (ACP), à laquelle nous proposons d'intégrer deux types de contraintes. Nous présentons une méthode de résolution qui a été implémentée dans un logiciel offrant une représentation 3D des données et grâce auquel l'utilisateur peut ajouter des contraintes de manière interactive, puis visualiser les modifications induites par ces contraintes. Deux types d'expérimentation sont présentés, 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 automatique 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.

Ce travail est soutenu par l'ANR "Masse de Données et Connaissances" Graphem (ANR-07-MDCO-006-05)