

Clustering multi-vues : une approche centralisée

Jacques-Henri Sublemontier, Guillaume Cleuziou,
Matthieu Exbrayat, Lionel Martin

Laboratoire d'Informatique Fondamentale d'Orléans
Université d'Orléans
B.P. 6759 - 45067 ORLEANS Cedex 2
{prenom.nom}@univ-orleans.fr

Résumé. Nous abordons dans ce papier le problème de la classification non-supervisée multi-vues, *i.e.* où les données peuvent être décrites par plusieurs ensembles de variables ou par plusieurs matrices de proximités. De nombreux domaines d'applications sont concernés, tels la Recherche d'Information, la Biologie, la Chimie et le Marketing. L'objet de cet axe de recherche est de proposer un cadre théorique et méthodologique permettant la découverte d'une classification réalisant un consensus entre les organisations émanant de toutes les vues. Il convient alors de combiner les informations de chacune des vues par l'intermédiaire d'un processus de fusion consistant à identifier l'accord entre les vues et à réduire le conflit. Plusieurs stratégies de fusion peuvent être appliquées, en amont, en aval, ou pendant le processus de classification. Nous présentons les différentes solutions de fusion envisageables suivant différents contextes applicatifs, puis nous nous focalisons sur des techniques dites centralisées. Nous proposons une approche de classification non supervisée floue qui généralise différentes solutions de fusion et nous présentons une extension à noyaux de cette approche, permettant le traitement de données hétérogènes. Nous montrons l'apport théorique et expérimental de cette approche sur des jeux de données *benchmarks* synthétiques et réels.

1 Introduction

La complexité toujours croissante des données constitue un défi pour la communauté fouille de données. Cette complexité peut concerner plusieurs aspects tels que la taille du jeu de donnée, la complexité des attributs, la temporalité ou plus généralement la multiplicité des données. Parmi les avancées récentes réalisées dans la communauté fouille de données, nous nous intéresserons dans ce travail à la fouille de données multi-vues. Pour de telles données, chaque individu peut-être décrit simultanément par plusieurs vues, et chaque vue apporte un éclairage différent sur l'organisation des individus.