

Interprétation spectrale de la classification relationnelle

Lazhar Labiod, Younès Bennani

LIPN UMR 7030, Université Paris 13
99, avenue Jean-Baptiste Clément, 93430 Villetaneuse
Prénom.Nom@lipn.univ-paris13.fr,

Résumé. Ce papier présente une vue spectrale sur l'approche de l'analyse relationnelle pour la classification des données catégorielles. Il établit d'abord le lien théorique entre l'approche de l'analyse relationnelle et le problème de classification spectrale. En particulier, le problème de classification relationnelle est présenté comme un problème de maximisation de trace, ce problème est donc transformé par la relaxation spectrale en un problème d'optimisation sous contraintes qui peut être résolu par des multiplicateurs de Lagrange, la solution est donnée par un problème de valeurs propres.

1 Introduction

Le Clustering a reçu une attention accrue comme un problème important en apprentissage non supervisé avec de nombreuses applications, un certain nombre de différents algorithmes et des méthodes ont émergé au fil des années. La plupart des techniques utilisent une mesure de similarité par paire pour mesurer la distance entre deux points de données. Récemment, des recherches ont été élaborées sur les approches de clustering de données catégorielles (Li et al, 2004) (Huang, 1998), où les données catégorielles sont celles dont les valeurs d'attributs sont nominales et non ordonnées, par exemple, la couleur et le sexe. Il est à noter que les difficultés se posent en matière de clustering des données catégorielles en raison de l'absence de liens intrinsèquement ordonnés de données catégorielles. La plupart des techniques de clustering basées sur les mesures métriques de distance ne sont donc pas applicables à la classification des données catégorielles. D'autre part, les méthodes spectrales ont été utilisées efficacement pour résoudre un certain nombre de problèmes de partitionnement de graphes, les méthodes de coupure minimale Ratio-cut et N-cut ont été utiles dans de nombreux domaines, tels que l'agencement du circuit et la segmentation d'image (Shi et al, 2000). Étant donné un ensemble de données, la minimisation des fonctions objectives des méthodes N-cut et Ratio-cut est un problème d'optimisation NP-difficile. Ainsi, en général ce problème est transformé par la relaxation spectrale en un problème d'optimisation avec une contrainte qui peut être résolu par des multiplicateurs de Lagrange, et la solution est donnée par un problème de valeurs propres. L'approche de l'analyse relationnelle a été utilisée pour la classification de données catégorielles (Marcotorchino et al, 1978) (Marcotorchino, 2006). Dans ce papier, nous montrons que le problème de la classification relationnelle peut être formellement modélisé comme un problème de maximisation de trace. Nous avons également établi le lien entre la méthode de classification spectrale et l'Analyse relationnelle (AR) qui est fondée sur le critère