

Une nouvelle approche de la programmation DC et DCA pour la classification floue

Le Thi Hoai An*, Le Hoai Minh**, Pham Dinh Tao***

*LITA, UFR MIM, Université Paul Verlaine - Metz, Ile du Saulcy, 57045 Metz Cedex, France
lethi@univ-metz.fr,

<http://www.lita;sciences.univ-metz.fr/lethi/>

**LITA, UFR MIM, Université Paul Verlaine - Metz, Ile du Saulcy, 57045 Metz Cedex, France
lehoai@univ-metz.fr

***LMI, INSA de Rouen, BP 08, Place Emile Blondel, 76131 Mont Saint Aignan Cedex, France
pham@insa-rouen.fr

Résumé. Dans cet article, nous nous intéressons à Fuzzy C-Means (FCM), une technique très connue pour la classification floue. Nous proposons un algorithme efficace basé sur la programmation DC (Difference of Convex functions) et DCA (DC Algorithm) pour résoudre ce problème. Les expériences numériques comparatives avec l'algorithme standard FCM sur les données réelles montrent la robustesse, la performance de cet nouvel algorithme DCA et sa supériorité par rapport à FCM.

1 Introduction

Le problème de classification automatique (clustering) est considéré comme une des problématiques majeures en extraction des connaissances à partir de données. Parmi les techniques de classification, la classification floue (fuzzy) via Fuzzy C-Means (FCM) est très connue. FCM a été introduite par Jim Bezdek en 1981 (Bezdek (1981)) comme une amélioration des méthodes clustering précédentes, et a été beaucoup développée dans les années 90. Cette approche a été appliquée avec succès dans plusieurs problèmes (diagnostic médical (Whitwell (2005)), classification de textes (Rodrigues and Sacks (2004))), et est de plus en plus utilisé dans le domaine du data mining.

Dans un travail récent (Le Thi et al. 3 (2006)) nous avons formulé le modèle de FCM pour la classification floue sous la forme d'un programme DC (Difference of Convex functions) et développé un schéma de DCA (DC Algorithm) pour sa résolution numérique. La programmation DC et DCA ont été introduits par Pham Dinh Tao en 1985 et intensivement développés par Le Thi Hoai An et Pham Dinh Tao depuis 1994 (voir (Le Thi Hoai An (1997)) - (Le Thi et al. 2 (2006)), (Pham Dinh Tao and Le Thi Hoai An (1997)), (Pham Dinh Tao and Le Thi Hoai An (1998)) et leurs références) pour devenir maintenant classiques et de plus en plus populaire. Ils ont été appliqués avec succès à nombreux problèmes d'optimisation non convexe différentiable ou non de grande dimension dans différents domaines des sciences appliquées, en particulier aux problèmes du data mining (voir par exemple (Le Thi et al. 1 (2006)), (Le Thi et al. 2 (2006)), (Liu et al (2003)), (Neumann et al. (2004)), (Weber et al. (2005))). Les résultats numériques présentés dans (Le Thi et al. 3 (2006)) montrent que, comme pour les autres problèmes déjà traités en data mining, DCA est efficace pour FCM. Ils prouvent également la supériorité de DCA par rapport à K-means. Cet algorithme est itératif et consiste en la résolution d'un programme convexe à chaque itération. Le temps de calculs de DCA est donc proportionnel à celui de la méthode utilisée pour résoudre les programmes convexes générés. Dans (Le Thi et al. 3 (2006)) l'algorithme du gradient pro-