

Une approche de co-classification automatique à base des cartes topologiques

Kais Allab *

Khalid Benabdeslem **, Alexandre Aussem **

*allab.kais@gmail.com,

**{kbenabde, aaussem}@univ-lyon1.fr

Université de Lyon, F69622-Lyon, France

Université de Lyon1 - Villeurbanne

Résumé. Nous présentons dans ce papier une nouvelle approche de co-classification automatique sur les tableaux de données continues. Cette approche est basée sur les cartes topologiques de Kohonen que nous appelons Bi-SOM (*Bi-clustering based on one Self-Organizing Map*). En outre de la question principale de la co-classification automatique liée au traitement simultané des lignes et des colonnes d'une matrice de données, nous proposons dans cette approche de répondre à plusieurs problématiques liées à cette tâche, à savoir: (1) la visualisation topologiques de bi-clusters avec une notion de voisinage, (2) l'optimisation de ces dits bi-clusters dans des macro-blocs et (3) la réduction de dimension par élimination itérative de blocs de "bruit". Enfin, nous présentons des résultats issus des expérimentations faites sur plusieurs bases de données réelles et d'autres synthétiques pour valider notre approche en comparaison avec d'autres méthodes de co-classification automatique.

1 Introduction

La classification non-supervisée est une tâche primordiale en apprentissage automatique et data mining. Elle consiste à représenter la meilleure structure induite de la distribution d'un ensemble de données non étiquetées. Dans ce cadre, son principe revient à regrouper les données en classes (clusters) en respectant deux propriétés intéressantes, à savoir la cohésion et la séparation. Ce regroupement est fait seulement sur les lignes (individus) en fonction de toutes les colonnes (variables) de la matrice de données. Or, il est clair que dans la partition finale, chaque classe d'individus est caractérisée par un sous ensemble de variables qui participent le plus à sa construction. Il est donc plus approprié de procéder par une recherche simultanée des relations spécifiques qui peuvent exister entre les individus et les variables. Cette démarche s'inscrit dans le cadre de la classification par blocs (block clustering, bi-clustering, co-classification ou classification croisée) qui consiste à réorganiser la matrice de données en blocs homogènes, suivant une certaine mesure de similarité (Govaert et Nadif, 2009).

Depuis la présentation de *Block Clustering* (Hartigan, 1972), le premier algorithme de co-classification qui remonte à plus de trente huit ans, plusieurs autres algorithmes ont été proposés dans divers domaines d'application, particulièrement en traitement d'images (Shi et Malik,