

OSOM : un algorithme de construction de cartes topologiques recouvrantes

Guillaume Cleuziou

Laboratoire d'Informatique Fondamentale d'Orléans (LIFO)
Université d'Orléans
Rue Léonard de Vinci - 45067 Orléans Cedex 2
Guillaume.Cleuziou@univ-orleans.fr

Résumé. Les modèles de classification recouvrante ont montré leur capacité à générer une organisation plus fidèle aux données tout en conservant la simplification attendue par une structuration en classes strictes. Par ailleurs les modèles neuronaux non-supervisés sont plébiscités lorsqu'il s'agit de visualiser la structure de classes.

Nous proposons dans cette étude d'étendre les cartes auto-organisatrices traditionnelles aux cartes auto-organisatrices recouvrantes. Nous montrons que cette nouvelle structure apporte des solutions à certaines problématiques spécifiques en classification recouvrante (nombre de classes, complexité, cohérence des recouvrements).

L'algorithme OSOM s'inspire de la version recouvrante des nuées dynamiques et de l'approche de Kohonen pour générer de telles cartes recouvrantes. Nous discutons du modèle proposé d'un point de vue théorique (fonction d'énergie associée, complexité, ...). Enfin nous présentons un cadre d'évaluation générale que nous utilisons pour valider les résultats obtenus sur des données réelles.

1 Introduction

La problématique de la classification recouvrante s'inscrit dans le processus général d'extraction de connaissances à partir de données ; elle consiste à faire émerger une organisation synthétique d'un ensemble d'individus à l'aide d'une structure de classes (hiérarchique ou non) dans laquelle chaque individu peut appartenir à plusieurs classes. Souvent plus riches et plus adaptés que leurs analogues non recouvrants, les modèles recouvrants accompagnent les avancées réalisées dans le domaine de la classification non-supervisée en général ; à titre d'exemples on notera les adaptations suivantes : les pyramides (Diday, 1984) généralisent les hiérarchies, OKM (Cleuziou, 2008) généralise l'algorithme des k -moyennes et MOC (Banerjee et al., 2005) correspond à une variante recouvrante des modèles de mélanges.

Outre les problématiques récurrentes inhérentes aux méthodes de classification (paramétrage, évaluation, passage à l'échelle, etc.) la classification recouvrante présente des problématiques propres telles que :

- **le choix du nombre de classes** : dans le cas des modèles recouvrants non hiérarchiques, l'ampleur des recouvrements entre classes est une caractéristique incontournable voire