

Analyse comparative de classifications : apport des règles d'association floues.

Pascal Cuxac*, Martine Cadot**, Claire François*

* INIST-CNRS, 2 allée du Parc de Brabois, 54 154-Vandoeuvre-lès-Nancy Cedex
pascal.cuxac@inist.fr ; claire.francois@inist.fr

** UHP/LORIA, Département d'informatique, BP 239, 54 506- Vandoeuvre-lès-Nancy
martine.cadot@loria.fr

Résumé. Notre travail s'appuie sur l'analyse d'un corpus bibliographique dans le domaine de la géotechnique à l'aide de cartes réalisées avec la plateforme Stanalyst®. Celui-ci intègre un algorithme de classification automatique non hiérarchique (les K-means axiaux) donnant des résultats dépendant du nombre de classes demandé. Cette instabilité rend difficile toute comparaison entre classifications, et laisse un doute quant au choix du nombre de classes nécessaire pour représenter correctement un domaine.

Nous comparons les résultats de classifications selon 3 protocoles : (1) analyse des intitulés des classes ; (2) relations entre les classes à partir des membres communs ; (3) règles d'association floues.

Les graphes obtenus présentant des similitudes remarquables, nous privilégions les règles d'association floues : elles sont extraites automatiquement et se basent sur la description des classes et non des membres. Ceci nous permet donc d'analyser des classifications issues de corpus différents.

1 Introduction

Les méthodes de cartographie de l'information sont devenues indispensables pour analyser de gros corpus de données bibliographiques dans le cadre de besoins de veille scientifique ou d'analyses stratégiques de la recherche. Elles s'appuient sur la combinaison de techniques de classification automatique et de cartographie. Pour réaliser ces analyses, nous utilisons la plateforme Stanalyst® (Polanco et al. 2001) qui permet de traiter des corpus bibliographiques et inclut la méthode des K-means Axiaux comme méthode de classification.

Cette méthode des K-means axiaux (Lelu 1993, Lelu et François 1992, Polanco et François 2000) est basée sur le principe de classification par centres mobiles, plus connue sous le nom de K-means, voir Forgy (1965) pour sa variante non adaptative et MacQueen (1967) pour sa variante adaptative. Comme ces méthodes, la méthode des K-means axiaux forme des groupements d'individus en affectant les éléments à des classes provisoires, puis en recentrant ces classes, et en recommençant ces deux phases de façon itérative. Cependant, cette méthode réalise une analyse factorielle sphérique sur chaque classe, les classes sont donc matérialisées par des demi-axes représentatifs des éléments. L'utilisation de cet axe permet de quantifier l'appartenance d'un élément à une classe (typicité). De plus, au lieu d'affecter l'élément à la seule classe où sa valeur est la plus grande, on l'affecte également aux classes pour lesquelles cette valeur dépasse un certain seuil. Il est alors possible