

Proposition d'une méthode de classification associative adaptative

Emna Bahri*, Stéphane Lallich*

*Laboratoire ERIC

5, avenue Pierre Mendès-France, 69500 Bron
emna.bahri | stephane.lallich@univ-lyon2.fr,
<http://eric.univ-lyon2.fr>

Résumé. La classification associative est une méthode de prédiction à base de règles issue de la fouille de règles d'association. Cette méthode est particulièrement intéressante car elle recherche de façon exhaustive les règles d'association pertinentes qu'elle filtre pour ne garder que les règles d'association de classe (celles admettant pour conséquent une modalité de classe), qui sont utilisées comme classifieur. Les connaissances produites sont ainsi directement interprétables. Des études antérieures montrent les inconvénients de cette approche, qu'il s'agisse de la génération massive de règles non utilisées ou de la mauvaise prédiction de la classe minoritaire lorsque les classes sont déséquilibrées. Nous proposons une approche originale du type boosting de règles d'association de classes qui utilise comme classifieur faible une base de règles significatives construites par un algorithme de génération d'itemsets fréquents qui se limite à l'extraction des seules règles de classe significatives et qui prend en compte le déséquilibre des données. Des comparaisons avec d'autres méthodes de classification associative montrent que notre approche améliore la précision et le rappel.

1 Introduction

La classification associative est une méthode d'apprentissage supervisé à base de règles introduite par (Liu et al. (1998)), puis améliorée par différents auteurs, ainsi (Yin et Han (2003)) et (Li et al. (2001)). De façon générique, cette méthode utilise pour la prédiction un ensemble de règles d'association dites règles d'association de classe (au sens où leur conséquent est une modalité de classe) qui sont produites à partir des algorithmes de recherche d'itemsets fréquents. Le principal intérêt de la classification associative est qu'il s'agit d'une méthode dont le classifieur est un ensemble de règles, ce qui permet à l'expert de comprendre le processus de classification et de justifier auprès d'un tiers les résultats de la prédiction (par exemple dans le cas du *scoring* bancaire ou dans le domaine médical). Par rapport aux arbres de décision (Quinlan (1993)), qui constituent l'une des principales méthodes d'apprentissage supervisé à base de règles, la classification associative présente deux avantages. En premier lieu, elle procède à une exploration exhaustive des règles, et non pas à une exploration gloutonne. En second lieu, elle