

Des Règles d'Association Ordinales aux Règles d'Association "Classiques"

Sylvie Guillaume

Laboratoire LIMOS, UMR 6158 CNRS, Université Blaise Pascal
Complexe scientifique des Cézeaux, 63177 AUBIERE Cedex - France
sylvie.guillaume@isima.fr

Résumé. L'intensité d'inclination est une mesure qui permet d'extraire des règles directement sur les données ordinales sans avoir à les transformer par un codage disjonctif complet précédé d'une discrétisation pour les variables quantitatives. Ce nouveau type de règles, les règles d'association ordinales, dégagent les comportements généraux de la population et il est essentiel d'aller au plus près des individus en découvrant des règles spécifiques c'est-à-dire des règles vérifiées par des sous-ensembles d'individus. Cet article présente donc la technique pour extraire des règles d'association en partant des règles d'association ordinales, se libérant ainsi de l'étape de transformation des données et surtout en obtenant une discrétisation des variables quantitatives fonction du contexte c'est-à-dire fonction des variables auxquelles elles sont associées. L'étude se termine par une évaluation sur une base de données bancaires.

1 Introduction

Les règles d'association (Agrawal et al. 1993) permettent de détecter des implications intéressantes entre les variables binaires d'une base de données. Cependant, les algorithmes d'extraction (Mannila et al. 1994 ; Agrawal et al. 1996) de ces règles reposent sur des fréquences (*taux de couverture et probabilité conditionnelle*) et par conséquent, sont limités pour l'étude de variables ordinales (*variables qualitatives ordinales et variables quantitatives*). Srikant et Agrawal (Srikant et al. 1996) présentent des techniques pour discrétiser automatiquement les variables quantitatives afin d'utiliser les algorithmes précédents. C'est également une préoccupation de Miller et Yang (Miller 1997), que de déterminer les intervalles des variables quantitatives. Aumann et Lindel (Auman et al. 1999) ainsi que G.I. Webb (Webb 2001) proposent un nouveau type de règles, les règles d'impact, qui détectent des interactions intéressantes entre des combinaisons de variables binaires en prémisse et une variable quantitative en conclusion. Fukuda et al. (Fukuda et al. 1996) ainsi que Yoda et al. (Yoda et al. 1997) proposent des règles composées de deux variables quantitatives en prémisse et d'une variable binaire en conclusion. Pour finir, Guillaume (Guillaume 2002) propose également un nouveau type de règles, les règles d'association ordinales, qui détectent des implications entre des combinaisons de tout type de variables. Ces règles reposent sur une mesure d'intérêt, l'intensité d'inclination, qui évalue la petitesse du nombre d'individus qui violent la règle. Cette mesure évite l'étape de transformation des données ordinales, c'est-à-dire l'étape de codage disjonctif complet précédée de l'étape de discrétisation pour les variables quantitatives, évitant ainsi l'obtention d'un nombre important