

Qualité d'un ensemble de règles : élimination des règles redondantes

Rémi Lehn*, Fabrice Guillet*, Henri Briand*

*Institut de Recherche en Informatique de Nantes
Université de Nantes
bâtiment IRESTE
La Chantrierie
44300 Nantes
{lehn,guillet,briand}@irin.univ-nantes.fr

Résumé. La qualité d'un ensemble de règles d'association est souvent considérée, par un utilisateur, selon la compréhension qu'il obtient du domaine étudié, en interprétant les règles qui lui sont présentées. Pour rendre l'ensemble de règles plus lisible, et donc améliorer ce critère de qualité global, nous appliquons aux règles d'association une méthode de réduction initialement proposée pour l'élimination des dépendances fonctionnelles redondantes. Malgré les différences entre les propriétés de ces deux types de relations, cette méthode permet d'obtenir des représentations de règles très concises et facilement interprétables par l'utilisateur.

1 Introduction

La découverte de règles d'association [Agrawal *et al.*, 1996] est motivée par l'utilisation de bases de données opérationnelles –c'est-à-dire dont la vocation principale est autre que de servir à des tâches d'ECD– pour découvrir une connaissance a priori inconnue et exploitable par un utilisateur dans un processus d'analyse ou de prise de décision [Lehn *et al.*, 1999]. Plusieurs algorithmes de découverte de règles d'association performants ont été publiés [Hipp *et al.*, 2000]. Une des hypothèses fondamentale de la découverte de règles d'association est que l'utilisateur de la connaissance produite ne spécifie pas de but à la découverte de connaissances. De par la nature intrinsèquement combinatoire de la méthode de découverte de règles et l'absence de buts a priori, l'utilisation classique de ces algorithmes (enchaînement des phases de sélection de données, mise en forme de ces données, induction de règles d'association, présentation des règles découvertes) fournit généralement un grand nombre de règles, sans aucun ordre, qui va donc à l'encontre du principe d'intelligibilité de la connaissance découverte par l'utilisateur inclus dans le processus d'ECD, et donc, influe directement sur la qualité de la connaissance perçue et effectivement exploitable par l'utilisateur dans son processus de prise de décision. Des expérimentations passées sur l'utilisation brute d'algorithmes de découverte de règles d'association comme l'algorithme *A Priori*, nous ont conduit à mettre en évidence des milliers de règles, à partir de bases de données dont le volume était comparable à celui des règles produites. Quelle est alors la qualité de la vision que l'utilisateur obtient du domaine étudié, s'il doit explorer des milliers de règles ? Quelle est la qualité de la procédure d'induction elle-même, lorsque l'effort à mettre en œuvre