

Ciblage des règles d'association intéressantes guidé par les connaissances du décideur

Claudia Marinica* et Fabrice Guillet*

*LINA, Ecole polytechnique de l'Université de Nantes
Rue Christian Pauc, BP 50609, 44306 Nantes cedex 3
{Claudia.Marinica, Fabrice.Guillet}@univ-nantes.fr

Résumé. L'usage du modèle des règles d'association en fouille de données est limité par la quantité prohibitive de règles qu'il fournit et nécessite la mise en place d'une phase de post-traitement efficace afin de cibler les règles les plus utiles. Cet article propose une nouvelle approche intégrant explicitement les connaissances du décideur afin de filtrer et cibler les règles intéressantes.

1 Présentation de l'approche

La technique d'extraction de règles d'association (Agrawal et al., 1993) a pour but de découvrir des tendances implicatives parmi les items d'une base de données transactionnelle. La force de cette technique réside dans sa capacité d'extraire toutes les associations intéressantes existant dans les données. Malheureusement, le grand nombre de règles produites rend très difficile, voire impossible, la sélection des règles intéressantes par le décideur. Par conséquent, il est essentiel d'aider le décideur lors d'une phase de post-traitement permettant une réduction efficace du nombre de règles.

A cette fin, plusieurs méthodes de post-traitement peuvent être utilisées, comme l'élagage, le résumé, le groupement ou la visualisation (Baesens et al., 2000). La phase d'*élagage* consiste dans l'élimination des règles redondantes ou inintéressantes, et un *résumé* réunit plusieurs règles plus spécifiques. Des groupes de règles sont générés par la phase de *groupement* et la phase de *visualisation* permet une meilleure présentation des résultats.

Cependant, la plupart des méthodes de post-traitement existantes sont basées uniquement sur des informations statistiques sur les données. Toutefois, l'intérêt d'une règle dépend fortement des connaissances et des attentes du décideur. Par exemple, si l'utilisateur cherche des règles inattendues, toutes les règles déjà connues doivent être élaguées. Ou encore, si le décideur souhaite cibler une famille de règles spécifique, le sous-ensemble correspondant doit être sélectionné (Padmanabhan et Tuzhuilin, 1999).

Cet article propose une nouvelle approche décrivant un nouvel environnement formel pour élaguer et grouper les associations en intégrant les connaissances du décideur dans le processus spécifique d'extraction de règles. L'approche est conçue autour de trois éléments principaux. Dans un premier lieu, un processus élémentaire de découverte de règles est appliqué sur les données générant l'ensemble total de règles d'association. Dans deuxième lieu, la base de connaissances offre un formalisme pour les connaissances et les attentes du décideur. Les connaissances du domaine permettent d'avoir une vision générale sur les connaissances du décideur dans le domaine de la base de données, et ses attentes expriment des associations que le décideur détient déjà sur les items. Finalement, la phase de post-traitement