

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

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

consiste dans l'application de plusieurs opérateurs (par exemple : élagage) sur les attentes du décideur de manière à extraire que les règles intéressantes.

En conséquence, les ontologies du domaine étendent la notion de Règles d'Association Généralisées (Srikant and Agrawal, 1995) basé sur les taxonomies, comme un résultat de la généralisation de la relation de subsomption par l'ensemble R de relations dans l'ontologie. En outre, les ontologies sont utilisées telles que des filtres sur les items, générant des familles d'items.

Pour améliorer la sélection de règles d'association, nous proposons un modèle de filtrage de règles, nommé Schémas de Règles, en généralisant les Impressions Générales (Liu et al., 1999). En d'autres termes, un schéma de règles décrit, à travers un formalisme à base de règles, les attentes du décideur relatives aux règles intéressantes/triviales. Par conséquent, les Schémas de Règles agissent de manière à grouper les règles, en définissant des familles de règles.

La phase de post-traitement développée est basée sur un ensemble d'opérateurs appliqués sur les schémas de règles permettant au décideur d'effectuer plusieurs actions sur les règles découvertes. Nous proposons deux opérateurs principaux : d'élagage et de filtrage de règles d'association. L'opérateur de filtrage est composé de deux autres opérateurs qui ciblent mieux les règles intéressantes : l'opérateur de conformité et l'opérateur de surprise.

Une plateforme interactive et itérative a été développée afin d'assister le décideur lors de la phase d'analyse. L'outil a été utilisé pour analyser une base de questionnaires fournie par Nantes Habitat¹, portant sur la satisfaction des clients concernant le logement. L'étude a été guidée par un expert du domaine et les résultats montrent l'efficacité de notre approche en termes de forte réduction du nombre de règles.

Nous envisageons de poursuivre cette approche en l'améliorant selon deux directions : l'enrichissement des formalismes de schémas de règles et l'intégration de cette approche dans l'algorithme de découverte de règles.

Références

- Agrawal, R., T. Imielinski, and A. Swami (1993). Mining Association Rules between Sets of Items in Large Databases. *Proceedings of the 12th ACM SIGMOD International Conference on Management of Data*, 207 - 216.
- Baesens, B., Viaene, S. and Vanthienen, J. (2000). Post-Processing of Association Rules. *The Sixth International Conference on Knowledge Discovery and Data Mining*, pages 2–8.
- Liu, B., W. Hsu, K. Wang and S. Chen (1999). Visually Aided Exploration of Interesting Association Rules. *Proceedings of the Third Pacific-Asia Conference on Methodologies for Knowledge Discovery and Data Mining*, Vol. 1574, Springer-Verlag, 26 – 28.
- Padmanabhan, B. and A. Tuzhuilin (1999). Unexpectedness as a Measure of Interestingness in Knowledge Discovery. *Decision Support Systems*, Volume 27, 303-318.
- Srikant, R. and R. Agrawal (1995). Mining Generalized Association Rules. In U. Dayal, P.M.D. Gray, and S. Nishio, eds, *Proceedings of the 21st International Conference on Very Large Databases*, 407 – 419.

¹ Nous remercions Nantes Habitat, l'Office Public d'Habitations à Bon Marché de Nantes, France, et plus particulièrement Mme. Christelle Le Bouter pour avoir soutenu cette étude.