

Réduction d'un jeu de règles d'association par des méta-règles issues de la logique du "sens commun"

Martine Cadot*, Joseph di Martino*, Amedeo Napoli*

* LORIA, B.P. 239
F-54506 Vandoeuvre-lès-Nancy France
{cadot, jdm, napoli}@loria.fr

Le processus d'extraction de règles d'association à partir d'une base de données génère un nombre important de règles [Han et Kamber, 2001]. Il existe des méthodes pour réduire la taille d'un jeu de règles d'association. Elles utilisent des seuils d'indices de qualité des règles [Kodratoff, 2001]. Mais ces techniques ne sont pas faciles à mettre en oeuvre car le choix des seuils appropriés nécessiterait des informations sur la distribution statistique des données [Cadot et Napoli, 2003]. Le but de cet article est de proposer une méthodologie de réduction à partir de méta-règles de nettoyage. Ces méta-règles ont pour rôle d'éliminer les contradictions et redondances qui existent dans les jeux de règles d'après la "logique du sens commun". Par exemple notre première méta-règle fait ressortir la contradiction de cette paire de règles $\{r1, r2\}$: $r1 =$ "les filles de la ville réussissent et s'inscrivent", $r2 =$ "les filles de la ville échouent et s'inscrivent", et une des deux ou les deux seront éliminées selon des paramètres choisis par l'expert. Nous exposons ici trois méta-règles.

Notre méthodologie permet de réduire de façon importante un jeu de règles complet, mais elle peut également nettoyer de ses incohérences un jeu de règles déjà réduit par les techniques habituelles de seuillage. Par exemple nous réduisons encore de 50% le jeu de règles de [Stumme *et al.*, 2001]. Nous étudions actuellement la possibilité de fournir à l'expert un ensemble plus riche de méta-règles et de techniques de nettoyage qu'il puisse paramétrer afin de rendre le jeu de règles optimal selon ses propres critères.

Mots-clés : Fouille de données, extraction de règles d'association, nettoyage d'ensemble de règles d'association, méta-règles d'élagage.

Références

- [Cadot et Napoli, 2003] M. Cadot et A. Napoli. Une optimisation de l'extraction d'un jeu de règles s'appuyant sur les caractéristiques statistiques des données. *RSTI-RIA-ECA*, 16 :631–656, 2003.
- [Han et Kamber, 2001] J. Han et M. Kamber. *Data Mining : Concepts and Techniques*. Morgan Kaufmann Publishers, San Francisco, 2001.
- [Kodratoff, 2001] Y. Kodratoff. Rating the interest of rules induced from data and within texts. In *Proc. DEXA 2001 IEEE eds*, pages 265–269, 2001.
- [Stumme *et al.*, 2001] G. Stumme, R. Taouil, Y. Bastide, N. Pasquier, et L. Lakhal. Intelligent structuring and reducing of association rules with formal concept analysis. In *Proc. KI 2001. LNAI 2174, Springer*, pages 335–350, 2001.