

# Méta-règles pour la génération de règles négatives

Sylvie Guillaume\* et Pierre-Antoine Papon\*\*

\*Clermont Université, Univ. Auvergne, LIMOS, BP 10448, F-63000 Clermont-Ferrand  
guillaum@isima.fr,

\*\*Clermont Université, Blaise Pascal, LIMOS, BP 10448, F-63000 Clermont-Ferrand  
papon@isima.fr

**Résumé.** La littérature s'est beaucoup intéressée à l'extraction de règles classiques (*ou positives*) et peu à l'extraction des règles négatives en raison essentiellement d'une part, du coût de calculs et d'autre part, du nombre prohibitif de règles redondantes et inintéressantes extraites. La démarche que nous avons retenue est de dégager les règles négatives lors de l'extraction des règles positives, et pour cela, nous recherchons les règles négatives que l'on peut inférer ou pas à partir de la pertinence d'une règle positive. Ces différentes inférences vont être formalisées par un ensemble de méta-règles.

## 1 Introduction

L'extraction de règles d'association est une tâche importante en fouille de données. Ces algorithmes d'extraction ont essentiellement deux limites : les variables doivent être binaires et seules les règles positives sont extraites. L'importance de l'extraction des règles négatives fut mise en évidence par Brin et al. (1997). L'extraction de telles règles est un défi car l'absence de variables binaires pour un individu dans une base de données est en général plus importante que la présence de ces mêmes variables. Pour finir, beaucoup de règles redondantes et inintéressantes sont extraites. Plusieurs techniques ont été proposées.

Brin et al. (1997) utilisent le test du  $\chi^2$  pour déterminer la dépendance entre deux motifs et ensuite une mesure de corrélation afin de trouver la nature de cette dépendance (*positive ou négative*). Savasere et al. (1998) combinent les motifs fréquents<sup>1</sup> positifs avec la connaissance du domaine afin de détecter les dissociations (*ou associations négatives*). Boulicaut et al. (2000) recherchent les règles négatives du type  $XY \rightarrow \bar{Z}$  ou  $\bar{X}Y \rightarrow Z$  en proposant une approche basée sur les contraintes. Wu et al. (2004) et Antonie et Zaïane (2004) utilisent une mesure supplémentaire pour générer les règles positives et négatives. Dans Missaoui et al. (2008), les auteurs proposent la génération de règles négatives à partir de règles positives mais dans un contexte bien particulier : celui des implications logiques<sup>2</sup>. Dans la lignée de cette dernière approche, nous souhaitons inférer les règles négatives à partir des règles positives mais dans un contexte plus large : celui des règles admettant des contre-exemples<sup>3</sup>. Ces différentes inférences vont être formalisées par un ensemble de méta-règles.

---

1. Un motif  $X$  est dit fréquent si sa probabilité d'apparition  $P(X)$  est supérieure à un seuil fixé par l'utilisateur.

2. Une implication logique est une règle d'association avec une confiance de 100%.

3. Un contre-exemple est un individu qui vérifie la prémisse  $X$  de la règle mais qui ne vérifie pas la conclusion  $Y$ .