

Méthodes et mesures d'intérêt pour l'extraction de règles d'exception

Béatrice DUVAL*, Ansaf SALLEB**, Christel VRAIN**

*LERIA UFR Sciences 2 Bd Lavoisier 49045 Angers Cedex 01

Beatrice.Duval@univ-angers.fr,

<http://www.info.univ-angers.fr/pub/bd/>

**LIFO Rue Léonard de Vinci BP 6759 45067 Orléans Cedex 02

Ansaf.Salleb, Christel.Vrain@lifo.univ-orleans.fr

<http://www.univ-orleans.fr/SCIENCES/LIFO>

◇BRGM 3 Avenue Claude Guillemin BP 6009 Orléans Cedex 02

Résumé. Les systèmes de génération de règles sont en général fondés sur des critères leur permettant de juger de la qualité des règles engendrées. On recherche souvent les règles solides, avec un support et une confiance suffisants, *i.e.*, concernant une partie importante de la population et vérifiées sur un grand nombre d'individus. Cependant, les experts sont souvent plutôt intéressés par des règles qui les surprennent, soit parce qu'elles sont peu fréquentes (de support faible) mais de confiance élevée, soit parce qu'elles représentent des exceptions aux règles solides. Une recherche exhaustive des règles d'exceptions n'est pas envisageable, les stratégies de recherche sont alors tout aussi importantes que les critères de qualité. C'est pourquoi dans ce papier, nous présentons un état de l'art, non seulement des mesures, mais aussi des méthodes pour l'extraction d'exceptions.

1 Introduction

L'extraction de connaissances dans les bases de données est devenue aujourd'hui un domaine de recherche prometteur et très actif, permettant de découvrir des connaissances dans un ensemble de données volumineux. Dans ce papier, nous nous intéressons principalement aux connaissances exprimées sous forme de règles. Ces règles doivent être non seulement intelligibles à l'utilisateur, mais également intéressantes, exploitables, c'est-à-dire qu'elles doivent aider l'expert dans son travail et notamment dans la prise de décision. Beaucoup de travaux ont été effectués sur l'extraction des règles d'association, comme introduites dans [Agrawal *et al.*, 1993], et exprimant des corrélations dans les données.

Par la suite, on considère une base de données à une seule table sur des attributs qualitatifs, éventuellement binaires. On appelle *item* l'instantiation d'une valeur à un attribut et *itemset* un ensemble d'items. Un tuple de la base de données est un itemset qu'on appelle aussi *transaction*.

On note $\mathcal{I} = \{x_1, x_2, \dots, x_n\}$, l'ensemble des items possibles et $\mathcal{T} = \{t_1, t_2, \dots, t_m\}$ un ensemble de transactions.

Extraire des règles d'association consiste à rechercher toutes les règles *solides* de la forme $\mathcal{X} \rightarrow \mathcal{Y}$, où \mathcal{X} et \mathcal{Y} sont des itemsets disjoints ; une telle règle exprime le