

Critères d'évaluation des mesures de qualité des règles d'association

Philippe Lenca*, Patrick Meyer**, Philippe Picouet*
Benoît Vaillant*, Stéphane Lallich***

*Département IASC - GET ENST Bretagne
Technopôle de Brest Iroise - CS 83818, 29238 Brest Cedex, France
{*prenom.nom*}@enst-bretagne.fr,
<http://www-iasc.enst-bretagne.fr/~nom/>

**Département de Mathématique - Université de Liège
Bât. B37 Grande Traverse, 12, B-4000 Liège, Belgique
patrick.meyer@internet.lu,
<http://patrickmeyer.tripod.com>

***Laboratoire ERIC - Université Lumière - Lyon 2
5 avenue Pierre Mendès-France, 69676 Bron Cedex, France
lallich@univ-lyon2.fr,
<http://eric.univ-lyon2.fr/~lallich/>

Résumé. Les algorithmes de fouille de données, en particulier dans le cadre de l'apprentissage non supervisé, génèrent un grand nombre de règles. Il est donc concrètement impossible de procéder à une validation de ces règles en les présentant à un expert du domaine. Afin d'assister ce traitement, de nombreuses mesures de qualité de règles ont été proposées pour sélectionner et ordonner automatiquement les règles. Se pose alors le problème du choix de la mesure adaptée aux données et aux besoins de l'expert, celui-ci n'étant pas - *a priori* - expert en fouille de données. Dans le contexte des règles d'association, nous proposons une caractérisation des mesures en fonction de propriétés ayant une sémantique intuitive nous permettant de mettre en œuvre un processus d'aide à la décision afin d'assister l'expert dans son choix d'une mesure adaptée à ses besoins et ainsi l'aider à sélectionner les *meilleures* règles.

1 Introduction

L'un des objectifs essentiels de l'extraction de connaissances à partir de données (ECD) est de produire des règles intéressantes du point de vue de l'utilisateur (expert métier). Il est aujourd'hui admis que l'intérêt d'une règle est difficile à évaluer de façon objective, tant son appréciation dépend des experts capables de l'analyser. Idéalement, une règle doit être valide, nouvelle, compréhensible [Fayyad et al., 1996] mais ces termes, très génériques, couvrent des réalités très variées lorsqu'une règle est examinée dans un contexte précis : maximiser le nombre de données concernées, minimiser le nombre de contre-exemples, etc.

Les mesures d'évaluation de la qualité des règles jouent alors un rôle essentiel dans les processus d'ECD afin de préfiltrer les règles qui seront effectivement présentées à