

# Étude expérimentale de mesures de qualité de règles d'association

Benoît Vaillant\*, Philippe Lenca\*  
Stéphane Lallich\*\*

\*GET ENST Bretagne / Département IASC – CNRS TAMCIC  
Technopôle de Brest Iroise - CS 83818, 29238 Brest Cedex, France  
{*prenom.nom*}@enst-bretagne.fr,

<http://www-iasc.enst-bretagne.fr/~nom/>

\*\*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.** La validation des connaissances extraites d'un processus d'ECD par un expert métier nécessite de filtrer ces connaissances. Pour ce faire, de nombreuses mesures ont été proposées, chacune répondant à des besoins spécifiques. Ces mesures présentent des caractéristiques variées et parfois contradictoires qu'il convient alors d'examiner. Arguant du fait que la sélection des bonnes connaissances passe aussi par l'utilisation d'un ensemble de mesures adaptées au contexte, nous présentons dans cet article une étude expérimentale de différentes mesures. Cette étude est mise en regard d'une étude formelle synthétisant les qualités des mesures.

**Mots-clés :** mesures d'intérêt, règles d'association, validation expérimentale, aide à la décision.

## 1 Introduction

Un des principaux objectifs de l'Extraction de Connaissances à partir des Données (ECD) est la production de règles intéressantes du point de vue des utilisateurs, ces derniers étant des experts métier, spécialistes du domaine sur lequel portent les données.

Une règle pour être intéressante doit être "*valide, nouvelle et compréhensible*" [Fayyad *et al.*, 1996]. Ces termes génériques, appliqués à des contextes précis, couvrent des réalités très différentes et l'intérêt d'une règle est une information particulièrement difficile à évaluer objectivement. De plus, le nombre particulièrement important de règles générées par les algorithmes utilisés en ECD ne permet pas aux utilisateurs-experts de faire eux-mêmes la sélection des règles pertinentes sans assistance. Le pré-filtrage automatique est un moyen pour réduire considérablement le nombre de règles qui devront être effectivement présentées au jugement de l'expert.

En permettant d'extraire un sous-ensemble des "meilleures règles" d'un ensemble trop volumineux, les mesures de qualité évaluant l'intérêt des règles jouent alors un rôle essentiel en ECD. La sélection des meilleures règles passe par l'utilisation de mesures adaptées au contexte et les mesures de qualité doivent être conçues et manipulées dans une perspective d'aide à la décision [Lenca *et al.*, 2003b].