# Mesure formelle de la robustesse des règles d'association

Yannick Le Bras[*,***] Patrick Meyer[*,***] Philippe Lenca[*,***] Stéphane Lallich[**]

[*]Institut Télécom, Télécom Bretagne
UMR CNRS 3192 Lab-STICC
Technopôle Brest Iroise CS 83818, 29238 Brest Cedex 3
{yannick.lebras || patrick.meyer || philippe.lenca}@telecom-bretagne.eu
[**]Université de Lyon, Laboratoire ERIC, Lyon 2, France
stephane.lallich@univ-lyon2.fr
[***]Université européenne de Bretagne, France

**Résumé.** Nous proposons dans cet article une définition formelle de la robustesse pour les règles d'association, s'appuyant sur une modélisation que nous avons précédemment définie. Ce concept est à notre avis central dans l'évaluation des règles et n'a à ce jour été que très peu étudié de façon satisfaisante. Il est crucial car malgré une très bonne évaluation par une mesure de qualité, une règle peut être très fragile par rapport à des variations légères des données. La mesure de robustesse que nous proposons dépend de la mesure de qualité utilisée pour évaluer les règles et du seuil d'acceptation minimal. Il est alors possible à partir de ces deux seuls éléments et de la valeur prise par la règle sur la mesure d'évaluer sa robustesse. Nous présentons plusieurs propriétés de cette robustesse, montrons sa mise en œuvre et illustrons celle-ci par les résultats d'expériences sur plusieurs bases de données pour quelques mesures. Nous donnons ainsi un nouveau regard sur la qualification des règles.

## 1 Introduction

Depuis sa définition originale par Agrawal et al. (1993) et l'algorithme APRIORI (Agrawal et Srikant, 1994) les motifs fréquents et les règles d'association ont suscité de très nombreux travaux algorithmiques (voir par exemple les synthèses proposées par Hipp et al. (2000), Goethals (2005) et Han et al. (2007)). Malheureusement, les nombreux algorithmes, déterministes et performants, du type d'APRIORI produisent de trop grandes quantités de règles. Par ailleurs l'intérêt d'une large proportion de ces règles est souvent discutable : Brin et al. (1997) mettent par exemple en évidence ce problème en montrant l'importance de l'étude des corrélations en complément du couple support confiance, c'est-à-dire la comparaison de la confiance à la probabilité du conséquent plutôt qu'à un seuil fixe. Face à ce problème à la fois quantitatif et qualitatif de nombreux efforts ont porté sur l'évaluation de l'intérêt des règles d'association afin de sélectionner idéalement toutes les règles intéressantes et uniquement celles-ci.

Une méthode très populaire pour évaluer l'intérêt des règles d'association consiste à quantifier cet intérêt à l'aide de mesures objectives (Piatetsky-Shapiro, 1991; Hilderman et Hamilton, 2000). Définies à partir de la contingence des règles, les mesures objectives permettent de