

# Comparaison des mesures d'intérêt de règles d'association : une approche basée sur des graphes de corrélation

Xuan-Hiep Huynh\*, Fabrice Guillet\*, Henri Briand\*

\*LINA CNRS FRE 2729 - Ecole polytechnique de l'université de Nantes  
La Chantrerie, BP 50609, 44306 Nantes cedex 3, France  
{xuan-hiep.huynh,fabrice.guillet,henri.briand}@univ-nantes.fr

**Résumé.** Le choix des mesures d'intérêt (MI) afin d'évaluer les règles d'association est devenu une question importante pour le post-traitement des connaissances en ECD. Dans la littérature, de nombreux auteurs ont discuté et comparé les propriétés des MI afin d'améliorer le choix des meilleures mesures. Cependant, il s'avère que la qualité d'une règle est contextuelle : elle dépend à la fois de la structure de données et des buts du décideur. Ainsi, certaines mesures peuvent être appropriées dans un certain contexte, mais pas dans d'autres. Dans cet article, nous présentons une nouvelle approche contextuelle mise en application par un nouvel outil, ARQAT, permettant à un décideur d'évaluer et de comparer le comportement des MI sur ses jeux de données spécifiques. Cette approche est basée sur l'analyse visuelle d'un graphe de corrélation entre des MI objectives. Nous employons ensuite cette approche afin de comparer et de discuter le comportement de trente-six mesures d'intérêt sur deux ensembles de données a priori très opposés : un premier dont les données sont fortement corrélées et un second aux données faiblement corrélées. Alors que nous attendions des différences importantes entre les graphes de corrélation de ces deux jeux d'essai, nous avons pu observer des stabilités de corrélation entre certaines MI qui sont révélatrices de propriétés indépendantes de la nature des données observées. Ces stabilités sont récapitulées et analysées.

## 1 Introduction

Dans la dernière décennie, la conception de mesures d'intérêt adaptées à l'évaluation de la qualité des règles d'association est devenue un défi important dans le contexte d'ECD. Bien que le modèle des règles d'association (Agrawal et al., 1993) permette une extraction non supervisée de tendances implicatives dans les données, il produit malheureusement de grandes quantités de règles, ce qui les rend inexploitable sans la mise en oeuvre d'une étape lourde de post-traitement. Le post-traitement doit aider l'utilisateur (un décideur ou un analyste) à choisir les meilleures règles en fonction de ses préférences. Une manière de faciliter la tâche de choix de l'utilisateur consiste à lui offrir des indicateurs numériques sur la qualité des règles d'association : des mesures d'intérêt adaptées à ses buts et aux données étudiées.

Dans les travaux précurseurs sur les règles d'association (Agrawal et al., 1993; Agrawal et Srikant, 1994), deux premières mesures statistiques sont introduites : le support et la confiance.