

Une base pour les règles d'association valides au sens de la mesure de qualité M_{GK}

Daniel R. Feno ^{*,**}, Jean Diatta^{*}, André Totohasina^{**}

^{*}Université de la Réunion
15 avenue René Cassin - BP 7151
97715 Saint-Denis messag cedex 9-France
fenodaniel2@yahoo.fr, jean.diatta@univ-reunion.fr
^{**}Université d'Antsiranana- BP O
201 Antsiranana-Madagascar
totohasina@yahoo.fr

Résumé. Ce papier concerne les règles d'association valides au sens de la mesure de qualité M_{GK} ¹. D'une part, M_{GK} est normalisée en ce sens que ses valeurs sont comprises entre -1 et $+1$ et reflètent les situations de référence telles que l'incompatibilité, la dépendance négative, l'indépendance, la dépendance positive et l'implication logique entre la prémisse et le conséquent d'une règle. D'autre part, ses propriétés permettent non seulement de considérer les règles positives et négatives (à droite, à gauche, à gauche et à droite), mais aussi de se restreindre uniquement aux règles positives et à celles négatives à droite. Ainsi, nous proposons une base pour les règles positives exactes, une base pour les règles négatives exactes, une base pour les règles positives approximatives et une base pour les règles négatives approximatives. La réunion de ces quatre bases constitue une base pour toutes les règles d'association (positives et négatives) valides au sens de M_{GK} .

1 Introduction

La fouille de données est un domaine de recherche actif dont l'importance n'a cessé de croître ces dernières années, du fait de son rôle comme outil approprié pour faire face à la croissance explosive de la taille de données stockées. Plusieurs techniques de fouille de données ont été proposées dans la littérature. La fouille des règles d'association en est l'une des plus populaires. Originellement appliquée dans le domaine de Marketing (Agrawal et al., 1993), de nos jours, la fouille des règles d'association s'applique dans différents domaines telles des bases de données médicales (Azé et al., 2003), des bases de données spatiales (Salleb, 2003), des taxonomies (Gras et Totohasina, 1995). Les règles d'association sont utiles pour la découverte de relations au sein de très grandes bases de données. Plusieurs algorithmes de fouille de règles d'association, fondés sur les mesures de qualité support et confiance, ont été proposés dans la littérature : APRIORI (Agrawal et al., 1993), CLOSE (Pasquier et al., 1999), CLOSET

¹ M_{GK} : Mesure de Guillaume-Kenchaff.