

Approche binaire pour la génération des fortes règles d'association

Thabet Slimani, Boutheina Ben Yaghlane, Khaled Mellouli

LARODEC, IHEC Carthage, Carthage Présidence 2016, Tunisia

thabet.slimani@issatm.rnu.tn

boutheina.yaghlane@ihed.rnu.tn

khaled.mellouli@ihed.rnu.tn

Résumé. Dans ce papier, nous proposons une nouvelle méthode d'extraction des règles d'association dans des bases de données relationnelles basée sur la technologie des *arbres de Peano* (Ptree). La structure de données utilisée pour représenter la base de données est un ensemble de Ptrees de base représentant chacun un vecteur binaire et tous ces Ptrees sont stockés dans des fichiers binaires. Nous montrons que la structure Ptree combinée avec la technique de réduction appelée *élagage par support minimum* produisent des règles d'association fortes et réduisent considérablement le temps de construction de l'association. En effet, notre approche présente l'avantage de ne pas effectuer des parcours coûteux de la base de données. Cette approche a été testée à travers un prototype que nous avons implémenté. Les résultats expérimentaux montrent que les règles d'association fortes sont générées dans un temps minimum comparativement à d'autres travaux.

Keywords : Fouille de données, Règle d'association, Itemsets fréquents, Arbres de Peano (Ptree), Règles ANDing.

1 Introduction

L'extraction de connaissances à partir des données ECD (KDD en anglais pour "Knowledge Discovery in Data") constitue un thème important du domaine de la fouille de données. Elle consiste en la recherche des connaissances implicites, cachées dans un ensemble volumineux d'informations stockées dans des bases de données. L'extraction des règles d'association est l'un des principaux problèmes traités de l'ECD. Ce thème a été initialement abordé par (Agrawal et al 1993) pour l'analyse des bases de données de transactions de ventes où chaque transaction est constituée d'une liste d'articles achetés (*items*), le but étant de découvrir les groupes d'articles (*itemsets*) achetés le plus fréquemment ensembles. Une *règle d'association* est une relation d'implication $X \rightarrow Y$ entre deux itemsets disjoints X et Y . Dans l'exemple du panier de la ménagère, cette règle indique que les transactions qui contiennent les items de l'ensemble X ont tendance à contenir les items de l'ensemble Y . Il existe deux mesures de qualité associées à chaque règle, *confiance* et *support*. Une règle apparaît dans l'ensemble de transactions D avec une confiance c si $c\%$ des transactions dans D qui contiennent X contiennent aussi Y . Une règle d'association $X \rightarrow Y$ a comme support s si $s\%$ des transactions dans D contiennent $X \cup Y$. Une *règle d'association forte* est une règle dont le support