

# Une approche logique pour la fouille de règles d'association

Abdelhamid Boudane\*, Said Jabbour\*  
Lakhdar Sais\*, Yakoub Salhi\*

\*CRIL-CNRS, Université d' Artois F-62307 Lens Cedex France  
{boudane,jabbour,sais,salhi}@cril.fr

**Résumé.** La découverte de règles d'association à partir de données transactionnelles est une tâche largement étudiée en fouille de données. Les algorithmes proposés dans ce cadre partagent la même méthodologie en deux étapes à savoir l'énumération des itemsets fréquents suivie par l'étape de génération de règles. Dans cet article, nous proposons une nouvelle approche basée sur la satisfiabilité propositionnelle pour extraire les règles d'association en une seule étape. Pour montrer la flexibilité et la déclarativité de notre approche, nous considérons également deux autres variantes, à savoir la fouille de règles d'association fermées et la fouille de règles indirectes. Les expérimentations sur plusieurs jeux de données montrent que notre approche offre de meilleures performances comparée à des approches spécialisées.

## 1 Introduction

L'extraction des règles d'association est l'une des tâches fondamentales de la fouille de données. Elle vise à découvrir des relations intéressantes cachées dans de grandes bases de données. Ces relations entre les items (ensembles d'attributs) sont présentées sous forme d'implications, appelées règles d'association. Depuis la première application, largement connue sous la dénomination de panier de la ménagère (Agrawal et Srikant, 1994), plusieurs nouveaux domaines d'application ont été identifiés comme la bioinformatique, le diagnostic médical, la détection d'intrusion, la fouille du web et l'analyse des données scientifiques.

L'extraction de règles d'association a connu de nombreux développements théoriques et algorithmiques. Parmi ces algorithmes, Apriori (Agrawal et Srikant, 1994) et FP-Growth (Han et al., 2004) sont les plus connus. Tous ces algorithmes partagent une même méthodologie à deux étapes. La première est consacrée à la recherche de tous les itemsets fréquents, et la seconde consiste à générer les règles avec une grande confiance en combinant ces itemsets fréquents. On peut citer aussi l'approche logique GUHA proposée par Hájek et al. (Hájek et al., 2010) il y a plus d'une trentaine d'années.

Comme souligné dans (Raedt et al., 2011), les contraintes font souvent partie de la spécification de plusieurs problèmes de fouille de données. Cette observation a conduit à un nouveau domaine de recherche actif et multidisciplinaire initié dans (Raedt et al., 2008), et permettant une fertilisation croisée entre la fouille de données et l'intelligence artificielle (IA). Deux modèles de représentation et de résolution d'IA ont été utilisés pour modéliser et résoudre plusieurs problèmes de fouille de données, à savoir la programmation par contrainte (PPC) et la