

***BoolLoader* : un chargeur efficace dédié aux bases de transactions denses**

Zahir Maazouzi¹ Ansaf Salleb Christel Vrain

LIFO Rue Léonard de Vinci, BP 6759, 45067 Orléans Cedex 02
{salleb, cv}@lifo.univ-orleans.fr

Résumé. Nous nous intéressons à la représentation et au chargement de bases de transactions en mémoire. Pour cela, nous proposons d'utiliser un format condensé fondé sur les diagrammes de décision binaires et nous présentons un algorithme que nous avons implanté en un système baptisé *BoolLoader*, pour charger des bases de transactions. Nous donnons également des résultats expérimentaux de notre système sur des bases éparses et denses.

Mots clés : Base de transactions, fonction booléenne, diagramme de décision binaires, densité d'une base de transactions, itemset fréquent.

1 Introduction

L'idée sous-jacente de la fouille de données est d'extraire des connaissances enfouies dans des volumes de données, en croissance continue. De nombreux algorithmes de recherche de connaissances ont été proposés, reposant sur divers modes de représentation de données et de résultats. Ces algorithmes rivalisent en performances car la tâche est coûteuse en temps et en espace mémoire, étant donnée la taille importante des données à exploiter.

Nous nous intéressons à la représentation et à la gestion des bases dites de transactions. À cette fin, nous avons prospecté une nouvelle approche de représentation de bases de transactions en mémoire, basée sur un format condensé des transactions. L'idée est de représenter une table de transactions par une fonction à variables binaires et à valeur entière. Une telle fonction est ensuite représentée et manipulée à l'aide d'une structure de données compacte et canonique appelée diagramme de décision binaire (DDB). À notre connaissance, l'utilisation de cette structure n'a pas encore été envisagée pour représenter ni les bases de transactions, ni les connaissances en fouille de données. Nous présentons un algorithme que nous avons implanté en un système baptisé *BoolLoader*, pour charger des bases de transactions en mémoire, en utilisant les DDB. Nous donnons ici des résultats expérimentaux de notre système sur des bases éparses et denses. L'étude empirique conduit à la conclusion suivante : *BoolLoader* est efficace sur les bases de transactions *denses*, et de façon plus générale, il donne une estimation assez intéressante de la densité d'une base.

¹À la mémoire de Zahir Maazouzi qui nous a quittés depuis.