

# Fouille de Données Multidimensionnelles : Différentes Stratégies pour Prendre en Compte la Mesure

Marc Plantevit, Anne Laurent, Maguelonne Teisseire

LIRMM, Université Montpellier 2, CNRS,  
161 Rue Ada 34392 Montpellier, France  
{plantevi, laurent, teisseire}@lirmm.fr

**Résumé.** Les entrepôts de données contiennent de gros volumes de données historisées stockées à des fins d'analyse. Des techniques d'extraction de motifs séquentiels multidimensionnels ont été développées afin de mettre en exergue des corrélations entre des positions sur des dimensions au cours du temps. Même si ces méthodes offrent une meilleure appréhension des données sources en prenant en compte certaines spécificités des cubes de données (e.g. multidimensionnalité, hiérarchies, relation d'ordre), aucune méthode ne permet de prendre directement en compte la valeurs des agrégats (mesure) dans l'extraction des motifs. Dans cet article, nous définissons deux méthodes de comptage du support d'une séquence multidimensionnelle en s'appuyant sur les valeurs des agrégats des cellules qui supportent cette séquence. Des expérimentations sont décrites et montrent l'intérêt de notre proposition.

## 1 Introduction

Il est largement reconnu que la technologie OLAP fournit de puissants outils d'analyse dans le but d'extraire des connaissances dans d'importants volumes de données hétérogènes et distribués. Plusieurs avantages viennent confirmer le potentiel d'un tel modèle d'analyse. En effet, OLAP a l'avantage de pouvoir représenter de manière assez naturelle les ensembles de données de la vie réelle, qui sont multidimensionnels, définis sur plusieurs niveaux de hiérarchies et fortement corrélés. Ainsi, les données peuvent être analysées suivant plusieurs niveaux d'agrégation à l'aide des nombreux opérateurs OLAP (roll-up, drill-down, etc.) ou par des requêtes complexes (top- $k$ , iceberg, gradient). De plus, l'une des principales forces d'OLAP est son intégration avec des outils plus complexes d'analyse issus des statistiques, de l'analyse des séries temporelles et de la fouille de données. Beaucoup de travaux visent à coupler OLAP à des outils et des algorithmes de fouilles de données à partir de la proposition de Han (1998). Cette proposition consiste à combiner la puissance d'OLAP à l'efficacité des algorithmes de fouilles de données permettant de découvrir des connaissances intéressantes dans de vastes volumes de données (e.g., l'ensemble des cellules d'un cube de données).

Dans ce contexte, l'extraction de motifs séquentiels permet de mettre en exergue des corrélations entre événements suivant leur chronologie d'apparition Agrawal et Srikant (1995). Extraire de tels motifs est particulièrement adapté quand on souhaite découvrir des tendances