

OpAC : Opérateur d'analyse en ligne basé sur une technique de fouille de données

Riadh Ben Messaoud*, Sabine Rabaseda**
Omar Boussaid**, Fadila Bentayeb*

ERIC – Université Lumière Lyon 2
5, avenue Pierre Mendès-France – 69676 Bron Cedex – France
<http://eric.univ-lyon2.fr>

*{rbenmessaoud | bentayeb}@eric.univ-lyon2.fr
**{sabine.rabaseda | boussaid}@univ-lyon2.fr

Résumé. L'analyse en ligne OLAP (*On-line Analysis Processing*) et la fouille de données (*Data Mining*) sont deux champs de recherche qui ont connu, depuis quelques années, des évolutions parallèles et indépendantes. De récentes études ont montré l'importance et l'intérêt de l'association entre ces deux domaines scientifiques. A l'heure actuelle, on assiste à l'accroissement du besoin d'une analyse en ligne plus élaborée. Nous pensons que le couplage entre OLAP et la fouille de données pourra apporter des réponses à ce besoin. Dans cet article, nous proposons d'adopter ce couplage en vue de créer un nouvel opérateur, baptisé *OpAC* (*Opérateur d'Agrégation par Classification*), d'analyse en ligne des données multidimensionnelles. *OpAC* consiste particulièrement en l'agrégation sémantique des modalités d'une dimension d'un cube de données en se basant sur la technique de la classification ascendante hiérarchique.

Mots-Clés : Couplage, Analyse en ligne, Cubes de données, Fouille de données, Agrégation sémantique, Opérateur d'analyse, Classification ascendante hiérarchique.

1 Introduction

La gestion des grandes masses de données est devenue une tâche difficile et assez coûteuse à maintenir. Les entrepôts de données (*Data Warehouses*) ont apporté des solutions efficaces à ce problème [Inmon, 1996] [Kimball, 1996]. En effet, un entrepôt de données représente une structure informatique centralisée dans laquelle est emmagasiné un volume important de données historisées, organisées par sujets et consolidées à partir de diverses sources d'informations. En plus de sa vocation de stockage, un entrepôt vise aussi l'exploitation des données dans un processus d'analyse en ligne et d'aide à la décision. Des modèles particuliers, tels que le schéma en étoile ou le schéma en flocon de neige, ont été conçus afin de rendre les données d'un entrepôt prêtes à l'analyse. Ces modèles permettent de créer des vues multidimensionnelles des données appelées aussi cubes de données dont la vocation est l'analyse en ligne OLAP (*On-Line Analysis Processing*) [Chaudhuri et Dayal, 1997]. Ainsi, la grande capacité de stockage, la