

Cubes de Données Convexes Non-Dérivables Fermés

Hanen Brahmi, Sadok Ben Yahia

Faculté des Sciences de Tunis.
Département des Sciences de l'Informatique.
Campus Universitaire 1060.
hanenbrahmi@gmail.com, sadok.benyahia@fst.rnu.tn

Résumé. De nombreuses approches sont proposées pour pré-calculer des cubes de données afin de répondre efficacement aux requêtes OLAP. La notion de cube de données a été déclinée sous différentes appellations: cubes icebergs, cubes différentiels ou encore cubes émergents. Les cubes convexes permettent de focaliser l'attention de l'utilisateur sur un ensemble particulier de tuples intéressants. Dans cet article, nous étudions les représentations concises des cubes convexes. À cet effet, nous introduisons une nouvelle structure d'un cube de données: le Cube Convexe Non-Dérivable Fermé ($CCND$ -Cube). Ce dernier permet de capturer tous les tuples d'un cube de données satisfaisant une combinaison de contraintes monotones et/ou antimonotones. Les expériences montrent que notre proposition fournit la représentation la plus compacte d'un cube de données de manière à optimiser à la fois le temps de calcul ainsi que l'espace de stockage nécessaire.

1 Introduction

La gestion des grandes masses de données est devenue une tâche difficile et assez coûteuse à maintenir. Ce problème a conduit aux développements *des entrepôts de données*. Ces derniers ont apporté une solution adéquate et efficace au problème de la croissance continue des données. Le contenu d'un entrepôt est analysé par les applications *On Line Analytical Processing* (OLAP), qui fournissent aux utilisateurs des moyens pour naviguer dans les données multidimensionnelles afin d'y découvrir des connaissances interprétables, exploitables et utiles à la prise de décision (Chaudhuri et Dayal, 1997).

En pré-calculant tous les agrégats possibles à différents niveaux de granularité, les cubes de données permettent une réponse efficace aux requêtes OLAP et sont donc un concept clef pour la gestion des entrepôts (Chaudhuri et Dayal, 1997). Étant donné un contexte d'extraction R contenant n attributs, le nombre de tuples dans un cube à k -attributs ($0 < k \leq n$), est le nombre de tuples dans R qui ont des valeurs d'attributs distinctes pour les k attributs. La taille d'un cuboïde est presque égale à la taille de R . Puisque le cube complet, construit à partir de R , consiste en 2^n cuboïdes, alors la taille de l'union des 2^n cuboïdes est beaucoup plus élevée que la taille de R .

Il est bien connu que le calcul des cubes de données est un problème combinatoire. En effet, la taille d'un cube augmente exponentiellement en fonction du nombre des dimensions.