

Cubes Fermés / Quotients Émergents

Sébastien Nedjar, Alain Casali, Rosine Cicchetti & Lotfi Lakhal

Laboratoire d'Informatique Fondamentale de Marseille (LIF), CNRS UMR 6166
Aix-Marseille Université,
IUT d'Aix-en-Provence, Avenue Gaston Berger, 13625 Aix-en-Provence Cedex
prenom.nom@lif.univ-mrs.fr

Résumé. Le concept de Cube Émergent a été introduit afin de comparer deux data cubes. Dans cet article, nous introduisons deux nouvelles représentations réduites du Cube Émergent sans perte des mesures : le Cube Fermé Émergent et le Cube Quotient Émergent. La première représentation est basée sur le concept de fermeture cubique. C'est la plus petite représentation possible du cube de données émergent. À partir du Cube Fermé Émergent et donc en stockant le minimum d'informations, il est possible de répondre efficacement aux requêtes qui peuvent être exécutées sur le Cube Émergent lui-même. La seconde représentation s'appuie sur la structure du Cube Quotient qui a été proposé pour résumer un cube de données. Le Cube Quotient est revisité afin de le doter d'une sémantique basée sur la fermeture cubique et donc adapté au contexte du Cube Émergent. Le Cube Quotient Émergent résultant est moins réduit que le Cube Fermé Émergent mais il préserve la propriété de " spécialisation/généralisation " du data cube qui permet la navigation au sein du Cube Émergent. Nous établissons également le lien entre les deux représentations introduites et celle basée sur les bordures classiques en fouille de données. Des expérimentations effectuées sur divers jeux de données visent à comparer la taille des différentes représentations.

1 Introduction

Afin de capturer les renversements de tendances dans les bases de données OLAP, le concept de Cube Émergent (Nedjar et al., 2009) a été proposé. Ce concept résulte du couplage de deux structures intéressantes, le data cube (Gray et al., 1997) et les motifs émergents (Dong et Li, 2005), dans le même esprit que le Skycube (Pei et al., 2006) qui combine le concept de data cube et la dominance de Pareto (Skyline (Börzsönyi et al., 2001)). À partir des cubes de deux relations d'une base de données pourvues de dimensions (ou d'attributs catégories), le Cube Émergent rassemble tous les tuples satisfaisant une *double contrainte d'émergence* : la valeur de leur mesure est faible dans une relation (contrainte C_1) et significative dans l'autre (contrainte C_2).

Offrir aux décideurs des Cubes Émergents est loin d'être trivial car deux data cubes, probablement volumineux donc coûteux à obtenir, à stocker et à gérer, doivent être calculés pour ensuite être comparés. En abordant cette problématique, notre idée est d'étudier des représen-