

Une approche de construction d'espaces de représentation multidimensionnels dédiés à la visualisation

Riadh Ben Messaoud, Kamel Aouiche, Cécile Favre

Laboratoire ERIC, Université Lumière Lyon 2
5 avenue Pierre Mendès-France
69676 Bron Cedex
{rbenmessaoud | kaouiche | cfavre}@eric.univ-lyon2.fr

Résumé. Dans un système décisionnel, la composante visuelle est importante pour l'analyse en ligne OLAP. Dans cet article, nous proposons une nouvelle approche qui permet d'apporter une solution au problème de visualisation des données engendré par l'éparsité. En se basant sur les résultats d'une analyse des correspondances multiples (ACM), nous tentons d'atténuer l'effet négatif de l'éparsité en organisant différemment les cellules d'un cube de données. Notre méthode ne cherche pas à réduire l'éparsité mais plutôt à construire un espace de représentation se prêtant mieux à l'analyse et dans lequel les faits du cube sont regroupés. Pour évaluer l'apport de cette nouvelle représentation des données, nous proposons un indice d'homogénéité basé sur le voisinage géométrique des cellules d'un cube. Les différents tests menés nous ont montré l'efficacité de notre méthode.

1 Introduction

Dans un contexte concurrentiel développé, les entreprises telles que les banques¹ doivent aujourd'hui être capables de prendre des décisions pertinentes, de façon réactive. La mise en place d'un processus décisionnel est alors nécessaire pour gérer une masse de données de plus en plus conséquente. Le stockage et la centralisation de ces données dans un entrepôt constitue un support efficace pour l'analyse de ces dernières. En effet, à partir d'un entrepôt de données, on dispose d'outils permettant de construire des contextes d'analyse multidimensionnels ciblés, appelés communément cubes de données. Ces cubes de données répondent à des besoins d'analyse prédéfinis en amont.

L'analyse en ligne OLAP (On Line Analytical Processing) est un outil basé sur la visualisation permettant la navigation, l'exploration dans ces cubes de données. L'objectif est d'observer des faits, à travers une ou plusieurs mesures, en fonction de différentes dimensions. Il s'agit par exemple d'observer les niveaux de ventes en fonction des produits, des périmètres commerciaux (localisations géographiques) et de la période d'achat.

De cette visualisation dépend la qualité d'exploitation des données. Or, différents facteurs peuvent dégrader cette visualisation. D'une part, la représentation multidimensionnelle

¹Nous remercions Michel Rougié, représentant du Crédit Lyonnais, pour les données fournies afin de valider ce travail.