Semantics-Aware Advanced OLAP Visualization of Multidimensional Data Cubes

Alfredo Cuzzocrea, University of Calabria, Italy
Domenico Saccà, University of Calabria, Italy
Paolo Serafino, University of Calabria, Italy

ABSTRACT

Efficiently supporting advanced OLAP visualization of multidimensional data cubes is a novel and challenging research topic, which results to be of interest for a large family of data warehouse applications relying on the management of spatio-temporal (e.g., mobile) data, scientific and statistical data, sensor network data, biological data, etc. On the other hand, the issue of visualizing multidimensional data domains has been quite neglected from the research community, since it does not belong to the well-founded conceptual-logical-physical design hierarchy inherited from relational database methodologies. Inspired from these considerations, in this article we propose an innovative advanced OLAP visualization technique that meaningfully combines (i) the so-called OLAP dimension flattening process, which allows us to extract two-dimensional OLAP views from multidimensional data cubes, and (ii) very efficient data compression techniques for such views, which allow us to generate “semantics-aware” compressed representations where data are grouped along OLAP hierarchies.

Keywords: approximate query answering; data cube compression; OLAP; OLAP visualization

INTRODUCTION

OLAP systems (Chaudhuri & Dayal, 1997; Codd, Codd, & Salley, 1993; Inmon, 1996; Kimball, 1996) have rapidly gained momentum in both the academic and research communities, mainly due to their capability of exploring and querying huge amounts of data sets according to a multidimensional and multi-resolution vision. Research-wise, three relevant challenges of OLAP have captured the attention of researchers during the last years: (i) the data querying problem, which concerns with how data are accessed and queried to support summarized knowledge extraction from massive data cubes; (ii) the data modeling problem, which concerns with how data are represented and, thus, processed inside OLAP servers (e.g., during query evaluation); and (iii) the data visualization problem, which concerns with how data are presented to OLAP users and decision makers in data warehouse environments. Indeed, research communities
have mainly studied and investigated the first two problems, whereas the last one, even if important-with-practical-applications, has been very often neglected.

**Approximate query answering (AQA)** techniques address the first challenge, and can be reasonably considered as one of the most important topics in OLAP research. The main proposal of AQA techniques consists in providing approximate answers to resource-consuming OLAP queries (e.g., range- (Ho, Agrawal, Megiddo, & Srikant, 1997), top-k (Fang, Shivakumar, Garcia-Molina, Motwani, & Ullman, 1998), and iceberg (Xin, Han, Cheng, & Li, 2006) queries) instead of computing exact answers, as decimal precision is usually negligible in OLAP query and report activities (e.g., see Cuzzocrea, 2005). Due to a relevant interest from the data warehouse research community, AQA techniques have been intensively investigated during the last years with the achievement of important results. Among the others, **histograms** (e.g., Acharya, Poosala, & Ramaswamy, 1999; Bruno, Chaudhuri, & Gravano, 2001; Gunopulos, Kolios, Tsotras, & Domeniconi, 2000; Muralikrishna & DeWitt, 1998; Poosala & Ioannidis, 1997), **wavelets** (Vitter, Wang, & Iyer, 1998), and **sampling** (e.g., Babcock, Chaudhuri, & Das, 2003; Chaudhuri, Das, Datar, Motwani, & Rastogi, 2001; Cuzzocrea & Wang, 2007; Gibbons & Matias 1998) are the most successful techniques, and they have also inducted several applications in contexts even different from OLAP (e.g., Cabibbo & Torlone, 1998), or theoretical failures of popular data cube operations, like aggregation functions (e.g., Lehner, Albrecht, & Wedekind, 1998; Lenz & Shoshani, 1997; Lenz & Thalheim, 2001).

Contrarily to data querying and modeling issues, since **data presentation models** do not properly belong to the well-founded conceptual-logical-physical design hierarchy for relational databases (which has also been inherited from multidimensional models (Vassiliadis et al., 1999)), the problem of OLAP data visualization has been studied and investigated so far only (Gebhardt, Jarke, & Jacobs, 1997; Inselberg, 2001; Keim, 1997; Maniatis, Vassiliadis, Skiadopoulos, & Vassiliou, 2003a, 2003b). On the other hand, being OLAP a technology focused at supporting decision making, thus based on (sensitive) information exploration and browsing, it is easy to understand that, in future years, tools for advanced visualization of multidimensional data cubes will quickly conquer the OLAP research scene.

Starting from fundamentals of data cube compression techniques and OLAP data visualization research issues, in this article we argue to meaningfully exploit the main results coming from the former and the goals of the latter in a combined manner, and propose a novel technique for supporting advanced OLAP visualization of multidimensional data cubes. The basic motivation of such an approach is realizing that (i) compressing data is an efficient way of visualizing data, and (ii) this intuition is well-founded at large (i.e., for any data-intensive system relying on massive data repositories), and, more specifically, it is particularly targeted to the OLAP context where
A Multidimensional Data Warehouse Development Methodology
Jose Maria Cavero, Carmen Costilla, Esperanza Marcos, Mario G. Piattini and Adolfo Sanchez (2003). Managing Data Mining Technologies in Organizations: Techniques and Applications (pp. 188-201).
www.igi-global.com/chapter/multidimensional-data-warehouse-development-methodology/25766?camid=4v1a