Algebraic and Graphic Languages for OLAP Manipulations

Franck Ravat, Université Toulouse I, France
Olivier Teste, Université Toulouse III, France
Ronan Tournier, Université Toulouse III, France
Gilles Zurfluh, Université Toulouse I, France

ABSTRACT

This article deals with multidimensional analyses. Analyzed data are designed according to a conceptual model as a constellation of facts and dimensions, which are composed of multi-hierarchies. This model supports a query algebra defining a minimal core of operators, which produce multidimensional tables for displaying analyzed data. This user-oriented algebra supports complex analyses through advanced operators and binary operators. A graphical language, based on this algebra, is also provided to ease the specification of multidimensional queries. These graphical manipulations are expressed from a constellation schema and they produce multidimensional tables.

Keywords: constellation; data warehouse; decision support systems; graphical query language; multidimensional manipulations; OLAP; OLAP algebra

INTRODUCTION

As competitiveness increases in the business world, and as faster reactivity is required more than ever, the decision-making process has become a major focus of research and is increasingly assisted with information technologies. OLAP (Online Analytical Processing) systems, aim to ease the decision-making process with a multidimensional data presentation. The use of Multidimensional DataBases (MDB) provides a global view of company data, and enables decision-makers to gain insight into an enterprise performance through fast and interactive access to data (Colliat, 1996). Unfortunately, in spite of a decade of research in OLAP systems, concepts and systems exist without uniform theoretical basis (Niemi et al., 2003; Rizzi et al., 2006).

Context and Related Works

Without a model based on a consensus for multidimensional data, many propositions have
been made. Multidimensional models rest upon cube or hyper-cube metaphor. Several surveys may be found in Chaudhuri and Dayal (1997), Blaschka et al. (1998), Vassiliadis and Sellis (1999), Pedersen et al., (2001), Torlone (2003) and Abelló et al. (2006).

The first works, based on a “cube model” that present data in the form of n-dimensional cubes (Li and Wang, 1996; Agrawal et al., 1997; Gyssens & Lakshmanan, 1997; Datta & Thomas, 1999), have the following drawbacks:

1. Weakness in modeling the fact (subject of analysis) and its Key Performance Indicators (KPI or measures)
2. Little or no conceptual modeling of dimensions (analysis axes) with no explicit capture of their hierarchical structure
3. No separation between structure and content

The second category called “multidimensional model” overcomes these drawbacks and it is semantically richer. It allows a precise specification of each multidimensional component (Lehner, 1998; Pedersen et al., 2001; Abelló et al., 2003; Trujillo et al., 2003; Abelló et al., 2006). Models of this category are based on the concepts of fact and dimension. Dimension attributes are organized in hierarchies. A hierarchy defines a point of view (or analysis perspective) of an analysis axis and is composed of the different aggregation levels of the measures. To our knowledge, hardly any multidimensional model provides a combined multi-fact and multi-hierarchy representation.

From a manipulation point of view, the first works on OLAP manipulation algebras extended relational algebra operators for the cube model (Gray et al., 1996; Li & Wang, 1996; Agrawal et al., 1997; Gyssens & Lakshmanan, 1997; Rafanelli, 2003). To counter the inadaptability of relational algebra for manipulating multidimensional structures in an OLAP context, numerous works provided operations for specifying and manipulating a cube (Cabibbo & Torlone, 1997; 1998; Pedersen et al., 2001; Abelló et al., 2003; Franzoni & Kamble, 2004). These works are not user-oriented (Abelló et al., 2003) for the following reasons: 1) they do not define an adapted structure for displaying decisional data to the user; 2) they are based on partial sets of OLAP operations; and 3) the defined operations do not easily represent OLAP manipulations of decision-makers (Ravat et al., 2006a).

Multidimensional OLAP analyses consist in exploring interactively multidimensional databases by drilling, rotating, selecting and displaying data. Although there is no consensus on a common core of a minimal set of operations for a multidimensional algebra, most papers offer a support of these operation categories:

- **Drilling**: these operations allow navigating through the hierarchical structure of the analysis axes, in order to analyze a measure with more or less precision. Drilling upwards (roll-up) consists in displaying the data with a coarser level of detail; for example, rollup allows changing corporate sales initially displayed by months into sales displayed by years. The opposite, drilling downwards (drill-down) consists in displaying the data with a finer level of detail.
- **Selections**: these operations allow the user to work on a subset of the available data. **Slice** specifies a restriction predicate on dimension data while **Dice** specifies a restriction predicate on fact data.
- **Rotations**: these operations allow changing analysis axes (rotation of dimensions), changing the subject of analysis (rotation of facts or drill-across), or changing an analysis perspective within the same dimension (rotation of hierarchies).

Some authors have also presented additional operations:

- **Fact modification**: these operations allow decision-makers to add and to remove a measure (analysis indicator or KPI) in the current analysis.
- **Dimension modification**: these operations enable the insertion of dimensional
Effectiveness of Fuzzy Classifier Rules in Capturing Correlations between Genes
www.igi-global.com/article/effectiveness-fuzzy-classifier-rules-capturing/1818?camid=4v1a