Chapter 15
When Spatial Analysis Meets OLAP: Multidimensional Model and Operators

Sandro Bimonte
Cemagref, France

Anne Tchounikine
INSA-Lyon, France

Maryvonne Miquel
INSA-Lyon, France

François Pinet
Cemagref, France

ABSTRACT
Introducing spatial data into multidimensional models leads to the concept of Spatial OLAP (SOLAP). Existing SOLAP models do not completely integrate the semantic component of geographic information (alphanumeric attributes and relationships) or the flexibility of spatial analysis into multidimensional analysis. In this chapter, the authors propose the GeoCube model and its associated operators to overcome these limitations. GeoCube enriches the SOLAP concepts of spatial measure and spatial dimension and take into account the semantic component of geographic information. The authors define geographic measures and dimensions as geographic and/or complex objects belonging to hierarchy schemas. GeoCube’s algebra extends SOLAP operators with five new operators, i.e., Classify, Specialize, Permute, OLAP-Buffer and OLAP-Overlay. In addition to classical drill-and-slice OLAP operators, GeoCube provides two operators for navigating the hierarchy of the measures, and two spatial analysis operators that dynamically modify the structure of the geographic hypercube. Finally, to exploit the symmetrical representation of dimensions and measures, GeoCube provides an operator capable of permuting dimension and measure. In this chapter, GeoCube is presented using environmental data on the pollution of the Venetian Lagoon.

DOI: 10.4018/978-1-61350-474-1.ch015
INTRODUCTION

Data warehousing and On-Line Analytical Processing (OLAP) systems are technologies designed to support business intelligence. OLAP models (hypercubes) are based on the concepts of dimensions and facts (Inmon, 1996). A fact is a concept that is relevant for the decision-making process, and it is described by a set of numerical indicators (measures). Dimensions, composed of hierarchies, allow for the analysis of facts along different analysis axes at different levels of detail.

New information and communication technologies make it possible to collect huge amounts of geographic data. These data are generated by remote sensing systems or other computer applications (Franklin, 1992). Geographic information is described by two components (Longley et al., 2001). The spatial component is the geometry and its position on the Earth’s surface. The semantic component is a set of (1) descriptive attributes and (2) spatial, thematic and map generalization relationships. Geographic Information Systems (GIS) have been developed in order to store, organize, visualize and analyze geographic data, (Longley et al., 2001).

Spatial analysis aims to understand, estimate and predict real phenomena, showing recurrent spatial structures and shapes. Several spatial operators have been proposed (e.g., overlay, map join, etc.) but a “standard” model and allied algebra have not yet been defined (Voisard & David, 2002). Nevertheless, Longley et al. (2001) proposed a classification of spatial analysis operators, i.e., query and reasoning methods, measuring methods, transformation methods, and synthesis methods. Transformation methods modify geographic data (i.e., overlay, buffer, etc.) through logic and/or spatial rules. Query and reasoning methods exploit relationships between geographic objects to enable multigranular spatial analysis (Timpf & Frank, 1997; Camossi et al., 2008). Here, data are represented at different levels of detail (or ‘granularity’), i.e., cities and regions, etc., to allow for support spatial analysis by adding or downscaling details for particular datasets through zoom-out/zoom-in operations.

Therefore, a new kind of Decision Support Systems called Spatial OLAP (SOLAP) has been developed in order to effectively factor spatial data into multidimensional analysis. SOLAP tools integrate GIS functionalities (memorizing, analyzing and visualizing) into OLAP and data warehousing systems (Marketos et al., 2008; Rivest et al., 2005). SOLAP tools were recently successfully used to analyze agricultural, economic, seismological data (Marketos et al., 2008), etc.

These systems are based on spatio-multidimensional models composed of spatial dimensions, spatial measures, which are analyzed through spatio-multidimensional operators. Multidimensional models have been proposed for SOLAP (Ahmed & Miquel, 2005; Pourrubas, 2003; Jensen et al., 2004; Damiani & Spaccapietra, 2006; Gómez et al., 2009; Sampaio et al., 2006; Silva et al., 2008; Glorio & Trujillo, 2008) which formalize the concepts of spatial dimensions, spatial measures and spatio-multidimensional operators. In particular, they define spatial measures as geometric values and/or the result of spatial operators. Spatio-multidimensional operators are defined as extensions of OLAP operators for spatial dimensions.

However, in our opinion, the existing SOLAP models actually limit certain aspects of spatio-multidimensional capabilities, namely the semantic component of geographic information and flexibility of spatial analysis. When geographic information is used as measure, SOLAP models reduce it to geometry without taking into account its relationships. Consequently, they support multigranular analysis, but with some limitations.

Spatial and multidimensional analyses are different in terms of flexibility. SOLAP operators (spatial drill-down, spatial roll-up and spatial slice) only allow navigation in the hypercube, since multidimensional data structures (dimensions and measures) and instances (members and measures