Chapter 104

A State-of-the-Art in Spatio-Temporal Data Warehousing, OLAP and Mining

Leticia Gómez
Instituto Tecnológico de Buenos Aires, Argentina

Bart Kuijpers
Hasselt University and Transnational University of Limburg, Belgium

Bart Moelans
Hasselt University and Transnational University of Limburg, Belgium

Alejandro Vaisman
Universidad de la Republica, Uruguay

ABSTRACT

Geographic Information Systems (GIS) have been extensively used in various application domains, ranging from economical, ecological and demographic analysis, to city and route planning. Nowadays, organizations need sophisticated GIS-based Decision Support System (DSS) to analyze their data with respect to geographic information, represented not only as attribute data, but also in maps. Thus, vendors are increasingly integrating their products, leading to the concept of SOLAP (Spatial OLAP). Also, in the last years, and motivated by the explosive growth in the use of PDA devices, the field of moving object data has been receiving attention from the GIS community, although not much work has been done to provide moving object databases with OLAP capabilities. In the first part of this paper we survey the SOLAP literature. We then address the problem of trajectory analysis, and review recent efforts regarding trajectory data warehousing and mining. We also provide an in-depth comparative study between two proposals: the GeoPKDD project (that makes use of the Hermes system), and Piet, a proposal for SOLAP and moving objects, developed at the University of Buenos Aires, Argentina. Finally, we discuss future directions in the field, including SOLAP analysis over raster data.

INTRODUCTION

Geographic Information Systems (GIS) have been extensively used in various application domains, ranging from economical, ecological and demographic analysis, to city and route planning (Rigaux, Scholl, & Voisard, 2001; Worboys, 1995). Spatial information in a GIS is typically stored in different so-called *thematic layers* (also called *themes*). Information in themes can be stored in data structures according to different data models, the most usual ones being the *raster model* and the *vector model*. In a thematic layer, spatial data is annotated with classical relational attribute information, of (in general) numeric or string type. While spatial data is stored in data structures suitable for these kinds of data, associated attributes are usually stored in conventional relational databases. Spatial data in the different thematic layers of a GIS system can be mapped univocally to each other using a common frame of reference, like a coordinate system. These layers can be overlapped or overlayed to obtain an integrated spatial view.

On the other hand, OLAP (On Line Analytical Processing) (Kimball, 1996; Kimball & Ross, 2002) comprises a set of tools and algorithms that allow efficiently querying multidimensional databases containing large amounts of data, usually called data warehouses. In OLAP, data is organized as a set of *dimensions* and *fact tables*. In the multidimensional model, data can be perceived as a *data cube*, where each cell contains a measure or set of (probably aggregated) measures of interest. As we discuss later, OLAP dimensions are further organized in hierarchies that favor data aggregation (Cabibbo & Torlone, 1997). Several techniques and algorithms have been developed for multidimensional query processing, most of them involving some kind of aggregate precomputation (Harinarayan, Rajaraman, & Ullman, 1996).

The need for OLAP in GIS

Different data models have been proposed for representing objects in a GIS. ESRI (http://www.esri.com) first introduced the *Coverage* data model to bind geometric objects to non-spatial attributes that describe them. Later, they extended this model with object-oriented support, in a way that behavior can be defined for geographic features (Zeiler, 1999). The idea of the Coverage data model is also supported by the Reference Model proposed by the Open Geospatial Consortium (http://www.opengeospatial.org). Thus, in spite of the model of choice, there is always the underlying idea of binding geometric objects to objects or attributes stored in (mostly) object-relational databases (Stonebraker & Moore, 1996). In addition, query tools in commercial GIS allow users to overlap several thematic layers in order to locate objects of interest within an area, like schools or fire stations. For this, they use indexing structures based on R-trees (Gutman, 1984). GIS query support sometimes includes aggregation of geographic measures, for example, distances or areas (e.g., representing different geological zones). However, these aggregations are not the only ones that are required, as we discuss below.

Nowadays, organizations need sophisticated GIS-based Decision Support System (DSS) to analyze their data with respect to geographic information, represented not only as attribute data, but also in maps, probably in different thematic layers. In this sense, OLAP and GIS vendors are increasingly integrating their products (see, for instance, Microstrategy and MapInfo integration in http://www.microstrategy.com/, and http://www.mapinfo.com/). Thus, aggregate queries are central to DSSs. Classical aggregate OLAP queries (like "Total sales of cars in California"), and aggregation combined with complex queries involving geometric components ("Total sales in all villages crossed by the Mississippi river and within a radius of 100 km around New Orleans") must be efficiently supported. Moreover, naviga-

Related Content

Sentiment Analysis of Tweets Using Naïve Bayes, KNN, and Decision Tree
Kadda Zerrouki, Reda Mohamed Hamouand Abdellatif Rahmoun (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines (pp. 538-554).*
www.irma-international.org/chapter/sentiment-analysis-of-tweets-using-nave-bayes-knn-and-decision-tree/308507

Opinion Mining in Tourism: A Study on "Cappadocia Home Cooking" Restaurant
Ibrahim Akn Özenand Ibrahim Ilhan (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines (pp. 1506-1527).*
www.irma-international.org/chapter/opinion-mining-in-tourism/308558

Data Warehouse and Interactive Map for Promoting Cultural Heritage in Saudi Arabia Using GIS
Nasser Allheeib, Marine Alraqdiand Mohammed Almukaynizi (2022). *International Journal of Data Warehousing and Mining (pp. 1-18).*
www.irma-international.org/article/data-warehouse-and-interactive-map-for-promoting-cultural-heritage-in-saudi-arabia-using-gis/314236

Estimating Semi-Parametric Missing Values with Iterative Imputation
Shichao Zhang (2012). *Exploring Advances in Interdisciplinary Data Mining and Analytics: New Trends (pp. 147-156).*
www.irma-international.org/chapter/estimating-semi-parametric-missing-values/61173

Finding Non-Coincidental Sporadic Rules Using Apriori-Inverse
Yun Sing Koh, Nathan Rountreeand Richard O'Keefe (2006). *International Journal of Data Warehousing and Mining (pp. 38-54).*
www.irma-international.org/article/finding-non-coincidental-sporadic-rules/1765