Extending a Conceptual Multidimensional Model for Representing Spatial Data

Elzbieta Malinowski
Universidad de Costa Rica, Costa Rica

Esteban Zimányi
Université Libre de Bruxelles, Belgium

INTRODUCTION

Data warehouses keep large amounts of historical data in order to help users at different management levels to make more effective decisions. Conventional data warehouses are designed based on a multidimensional view of data. They are usually represented as *star* or *snowflake schemas* that contain relational tables called fact and dimension tables. A *fact table* expresses the focus of analysis (e.g., analysis of sales) and contains numeric data called *measures* (e.g., quantity). Measures can be analyzed according to different analysis criteria or *dimensions* (e.g., by product). Dimensions include attributes that can form *hierarchies* (e.g., product-category). Data in a data warehouse can be dynamically manipulated using on-line analysis processing (OLAP) systems. In particular, these systems allow automatic measure aggregations while traversing hierarchies. For example, the roll-up operation transforms detailed measures into aggregated data (e.g., daily into monthly sales) while the drill-down operation does the contrary.

Data warehouses typically include a location dimension, e.g., store or client address. This dimension is usually represented in an alphanumeric format. However, the advantages of using spatial data in the analysis process are well known since visualizing data in space allows users to reveal patterns that are difficult to discover otherwise. Spatial databases have been used for several decades for storing and managing spatial data. This kind of data typically represents geographical objects, i.e., objects located on the Earth's surface (such as mountains, cities) or geographic phenomena (such as temperature, altitude). Due to technological advances, the amount of available spatial data is growing considerably, e.g., satellite images, and location data from remote sensing systems, such as Global Positioning Systems (GPS). Spatial databases are typically used for daily business manipulations, e.g., to find a specific place from the current position given by a GPS. However, spatial databases are not well suited for supporting the decision-making process (Bédard, Rivest, & Proulx, 2007), e.g., to find the best location for a new store. Therefore, the field of spatial data warehouses emerged as a response to the necessity of analyzing high volumes of spatial data.

Since applications including spatial data are usually complex, they should be modeled at a conceptual level taking into account users' requirements and leaving out complex implementation details. The advantages of using conceptual models for database design are well known. In conventional data warehouses, a multidimensional model is commonly used for expressing users' requirements and for facilitating the subsequent implementation; however, in spatial data warehouses this model is seldom used. Further, existing conceptual models for spatial databases are not adequate for multidimensional modeling since they do not include the concepts of dimensions, hierarchies, and measures.

BACKGROUND

Only a few conceptual models for spatial data warehouse applications have been proposed in the literature (Jensen, Klygis, Pedersen, & Timko, 2004; Timko & Pedersen, 2004; Pestana, Mira da Silva, & Bédard, 2005; Ahmed & Miquel, 2005; Bimonte, Tchounikine, & Miquel, 2005). Some of these models include the concepts presented in Malinowski and Zimányi (2004) and Malinowski and Zimányi (2005), to which we will refer in the next section; other models extend non-spatial multidimensional models with different aspects, such as imprecision (Jensen *et al.*, 2004), location-based data (Timko & Pedersen, 2004), or continuous phenomena such as temperature or elevation (Ahmed & Miquel, 2005).

Other authors consider spatial dimensions and spatial measures (Stefanovic, Han, & Koperski, 2000; Rivest, Bédard, & Marchand, 2001; Fidalgo, Times, Silva, & Souza, 2004); however, their models are mainly based on the star and snowflake representations and have some restrictions, as we will see in the next section.

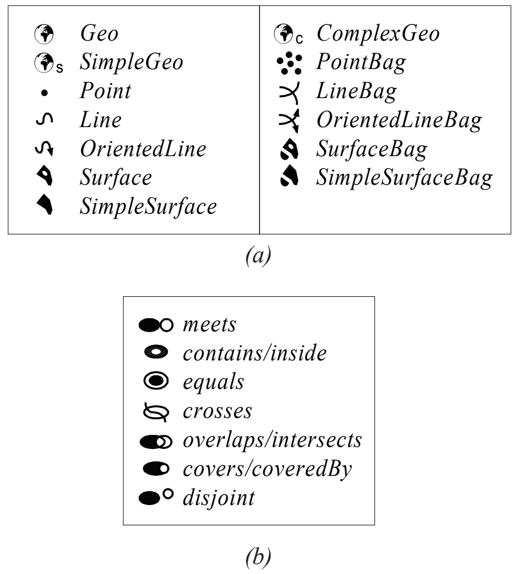
We advocate that it is necessary to have a conceptual multidimensional model that provides organized spatial data warehouse representation (Bédard, Merrett, & Han, 2001) facilitating spatial on-line analytical processing (Shekhar & Chalwa, 2003; Bédard *et al.*, 2007), spatial data mining (Miller & Han, 2001), and spatial statistical analysis. This model should be able to represent multidimensional elements, i.e., dimensions, hierarchies, facts, and measures, but also provide spatial support.

Spatial objects correspond to real-world entities for which the application needs to keep their spatial characteristics. Spatial objects consist of a *thematic* (or descriptive) component and a *spatial* component. The thematic component is represented using traditional DBMS data types, such as integer, string, and date. The spatial component includes its geometry, which can be of type point, line, surface, or a collection of these types. Spatial objects relate to each other with topological relationships. Different topological relationships have been defined (Egenhofer, 1993). They allow, e.g., determining whether two counties touches (i.e., share a common border), whether a highway crosses a county, or whether a city is inside a county.

Pictograms are typically used for representing spatial objects and topological relationships in conceptual models. For example, the conceptual spatio-temporal model MADS (Parent *et al.* 2006) uses the pictograms shown in Figure 1.

The inclusion of spatial support in a conceptual multidimensional model should consider different aspects not present in conventional multidimensional models, such as the topological relationships existing between the different elements of the multidimensional model or aggregations of spatial measures, among others. While some of these aspects are briefly mentioned in the literature, e.g., spatial aggregations (Pedersen & Tryfona, 2001), others are neglected, e.g., the influence on aggregation procedures of the topological relationships between spatial objects forming hierarchies.

Figure 1. Pictograms for a) spatial data types and b) topological relationships



MAIN FOCUS

The MultiDim model (Malinowski & Zimányi, 2008a, 2008b) is a conceptual multidimensional model that allows designers to represent fact relationships, measures, dimensions, and hierarchies. It was extended by the inclusion of spatial support in the different elements of the model (Malinowski & Zimányi, 2004; Malinowski & Zimányi, 2005). We briefly present next our model.

A Conceptual Multidimensional Model for Spatial Data Warehouses

To describe the MultiDim model, we use an example concerning the analysis of highway maintenance costs. Highways are divided into highways sections, which at their turn are divided into highway segments. For each segment, the information about the number of cars and repairing cost during different periods of time is available. Since the maintenance of highway segments is the responsibility of counties through which the highway passes, the analysis should consider the administrative division of the territory, i.e., county and state. The analysis should also help to reveal how the different

Related Content