Cooperation of Geographic and Multidimensional Databases

Elaheh Pourabbas
National Research Council, Rome, Italy

INTRODUCTION

In recent years, the enormous increase of independent databases widely accessible through computer networks has strongly motivated the interoperability among database systems. Interoperability allows the sharing and exchange of information and processes in heterogeneous, independent, and distributed database systems. This task is particularly important in the field of decision support systems. These systems through the analysis of data in very large databases identify the unusual trends in particular applications for creating opportunities for new business or for forecasting production needs.

Currently, in the research community, geographic information systems (GISs) and multidimensional databases (MDDBs) are seen as the most promising and efficient information technologies for supporting decision making. Geographic information systems, which are geographic-database- (GDB) dependent, through graphic display functionalities and complex spatial data structures, facilitate the storage and manipulation of geographic data.

Multidimensional databases refer either to statistical databases (Chan & Shoshani, 1981; Rafanelli & Shoshani, 1990), which mostly represent applications in the socio-economic area, or OLAP (online analytical processing) databases (Agrawal, Gupta & Sarawagi, 1997; OLAP Council, 1997), which emphasize business applications. Similar to statistical databases, OLAP databases have a data model that represents one or more "summary measures" over a multidimensional space of "dimensions," where each dimension can be defined over a hierarchy of "levels" (Shoshani, 1997). In this area, mostly the aspects of handling the multidimensional data and summarizations over the dimensions have been largely investigated. However, unlike OLAP databases, statistical databases may have only the summarized data available for reasons of privacy. These databases are often referred to by the term "summary databases."

The common key elements between geographic and multidimensional data that allow effective support in data cooperating are basically *time* and *space*. In literature, space has been considered as a bridge element for cooperating GDB and MDDB, on which our attention will be focused.

A feature notably lacking in most GDBs is the capability of accessing and manipulating business data, which are stored in MDDBs. We tackle this task by a novel approach that shares a number of characteristics and goals with the approaches proposed in literature. They aimed at defining a set of operators applicable to either spatial or summary data without dealing with the "logical organization" of databases at all. Similar to these models, our approach is addressed for cooperative query answering but it provides a data model for the summary data manipulation in the context of GDB. In addition, the above-mentioned models are based on multidimensional data formed by solely one location dimension, whereas in our approach we also consider data defined by more than one location dimension and we analyze their effect on data modeling and query answering.

BACKGROUND

In the database community, the cooperation between GDB and MDDB is indicated by taking into account the notion of map generalization. Map generalization is intended to consider the impact of scale and resolution on spatial data querying (see Muller, Lagrange, & Weibel, 1995). In this context, some attempts have been made to look for a standard set of multidimensional (or statistical) operators based on aggregation and disaggregation.

Gargano, Nardelli, & Talamo (1991), for instance, have extended the relational algebra essentially by defining two algebraic operators that are able to manipulate either the spatial extension of geographic data or summary data. They are named *G-Compose* and *G-Decompose*. The first operator is denoted by $G-Compose_X(F_y;Y)$, where X and Y are two nonintersecting subsets of attributes of a relation R. In the case of spatial data, it "merges" all tuples of R which are already projected on Y in a single one whose Y value is generated by the application of the spatial fusion function. This function, which is represented by F_y, takes a subset of geometric attributes of a given type and returns a single geometric attribute of the same type. In the case of summary data, F_y aggregates the numeric value of Y attributes. The effect of G-Decompose is that all tuples of R projected on Y are "decomposed." In this proposal, the fundamental issues of hierarchies and data aggregation for either spatial or summary data have been omitted.

These issues are discussed later in an approach proposed by Rigaux and Scholl (1995). In this work, the authors make the bridge between the geographic and statistical disciplines by defining an aggregation technique over a hierarchy of space partitions. Their model is based on the concept of "partition" that is used for partitioning either geometric space or other sets (e.g., a set of people). They introduced a relation called *cover*, which is represented by the schema $O = \{A_1, \ldots, A_n, A_g\}$, such that $\pi_{A_g}(O)$ is a partition and there is a bi-univocal functional dependency between the attributes $A_1, \ldots, A_n$ and A_g. They defined the *geometric projection* operator on a subset of attributes $S = \{A_1, \ldots, A_q\}$ as follows:

$$apply_{\Sigma_{Geo}}\ (nest_S(\pi_{S,A_g}(O)),$$

where $\pi_{A_g}(O)$ is the N1NF grouping operation (see Abiteboul & Bidoit, 1986) on S, $nest_S(\pi_{S,A_g}O)$ gives the result with the schema $\{A_1, \ldots, A_q, B\}$ where $B = set(A_g)$, and Σ_{Geo} performs the spatial aggregation function UNION on attribute $B = set(A_g)$. They used the relation cover for representing summary data, in which each descriptive attribute A can be defined on a hierarchical domain. The same operator is redefined as below:

$$apply_{\Sigma_{Geo}}\ (nest_S\,(gen_{A:A'}(O))),$$

where before applying the nest operator, the abstraction level of hierarchy to which the attribute A belongs is changed. It is indicated by $gen_{A:A'}(O)$. Note, in this case Σ_{Geo} performs the numeric aggregation function SUM.

The model proposed by Rigaux and Scholl (1995) is addressed to generate maps in multiple representations of data using the hierarchy of space partitions and the hierarchy induced by a partial order relationship in the domain of an attribute. In this proposal only one location dimension hierarchy for summary data is considered.

While the above models give a formal definition for the cooperation between spatial and multidimensional environments, some other works consider the architectural aspects of an integration system. For instance, Kouba, Matousek, and Miksovsky (2000) tried to identify some requirements for the correct and consistent functionality of system interconnection. They proposed an integration module that has two different roles: One is the transformation of data from external data sources, and the other refers to the integration of GIS and data warehouse through their common components, that is, location. The GIS under consideration is based on an object-oriented model that identifies the basic GIS elements that are classes and objects. In the GIS system, the structure of the geographical class hierarchy is stored in a metadata object for accessing directly from the integration module. Furthermore, the implementation aspects of the integration of the Microsoft SQL Server 7 Data Warehouse and ArcView GIS System are discussed.
Sindoni G., De Francisci S., Paolucci M., and Tininini L. (2001) have considered the integration of several spatiotemporal data collections of the Italian National Statistics Institute. The integration system is defined mainly by a historical database containing the temporal variation of territorial administrative partitions, a statistical data warehouse providing statistical data from a number of different surveys, and a GIS providing the cartography of the Italian territory up to census tract level. The implemented cooperative system manages the maps of the temporal evolution of a certain number of administrative regions and links to these maps the content of the above-mentioned statistical database of a given year.

A LOGICAL APPROACH

For describing the logical data model, we consider an object-oriented GDB and a cube-based MDDB, the main components of which are indicated in Table 1.

Cooperation by Binding Elements

In order to clarify the need for such cooperation, let us consider the case of a GDB user wishing to display a query result on a map with different combinations of geographic and nongeographic data. Let the query be as follows: "Find all the Italian regions which are adjacent to the Tuscany region, in which the number of cars sold in 1990, in the case of <Corolla>, was greater than 10,000."

For answering this query, it is necessary not only to retrieve the adjacent regions, but also to perform some OLAP operations on the time and product dimensions of the Car_Sales cube shown in Figure 1a. The former analyzes the topological spatial relationship (i.e., adjacency) and can be performed only in GDB. The solution of such queries depends essentially on a data model that enables the component databases to cooperate but remain independent. The main idea of our approach is to explore the additional input that can come to GDB from MDDB.

As we have shown in Table 1 and in Figure 1b, an object-oriented geographic database is characterized by

Related Content