Modeling and Querying Continuous Fields with OLAP Cubes

Leticia Irene Gómez, Instituto Tecnológico de Buenos Aires, Buenos Aires, Argentina
Silvia Alicia Gómez, Instituto Tecnológico de Buenos Aires, Buenos Aires, Argentina
Alejandro Vaisman, Department of Computer & Decision Engineering (CoDE), Université Libre de Bruxelles, Brussels, Belgium

ABSTRACT

The notion of SOLAP (Spatial On-Line Analytical Processing) is aimed at exploring spatial data in the same way as OLAP operates over tables. SOLAP, however, only accounts for discrete spatial data. Current decision support systems are increasingly being needed for handling more complex types of data, like continuous fields, which describe physical phenomena that change continuously in time and/or space (e.g., temperature). Although many models have been proposed for adding spatial (continuous and discrete) information to OLAP tools, no one is general enough to allow users to just perceive data as a cube, and analyze any type of spatial data together with typical alphanumerical discrete OLAP data, using only the classic OLAP operators (e.g., Roll-up, Drill-down). In this paper the authors propose a model and an algebra supporting it, that allow operating over data cubes, independently of the underlying data types and physical data representation. That means, in this approach, the final user only sees the typical OLAP operators at the query level, whereas at lower abstraction levels the authors provide discrete and continuous spatial data support as well as different ways of partitioning the space. As far as the authors are aware of, this is the first proposal, which provides such a general framework for spatiotemporal data analysis.

Keywords: Continuous Fields, Geographic Information Systems (GIS), Map Algebra, On-Line Analytical Processing (OLAP), Spatial On-Line Analytical Processing (SOLAP)

INTRODUCTION AND BACKGROUND

Nowadays organizations need to integrate their data warehouses, typically exploited with OLAP (On-Line Analytical Processing) (Kimball & Ross, 2002) tools, with geographical information. SOLAP (standing for Spatial OLAP) (Rivest et al., 2001), aims at exploring spatial data in the same way as OLAP operates over tables and charts. Continuous Fields (from now on, fields) describe physical phenomena that change continuously in time and/or space, like temperature, pressure, and land elevation. Fields

DOI: 10.4018/jdwm.2013070102
can be of different dimensionality, e.g., elevation in a 2D spatial domain, pollution defined in a 3D spatial domain, or temperature in a 4D spatiotemporal domain. More generally, the N-dimensional domain can be any continuous one. For example, we can even model time series as a 1D field. Some efforts to extend SOLAP to continuous fields have been carried out (Vaisman & Zimányi, 2009; Gómez et al., 2010). In real-world practice, scientists and practitioners register the value of a field taking samples at (generally) fixed locations, and inferring the values at other points in space using some interpolation method. Different discrete data models have been proposed to represent continuous fields, based on sampling and interpolation. The most popular one is the raster model, where the 2D space is divided into regular squares. To operate with fields represented as raster data, Tomlin (1990) proposed Map Algebra. Basically, Map Algebra consists in three operators: Local, Focal and Zonal. Given a collection of grids, and an aggregate function, the Local operator returns a grid such that the value in each cell is the result of applying the aggregate function to values of the cell at the same location in the input grids. Focal operators receive a field and a region (represented as a contiguous group of cells), and generate a new grid where the value at each cell is calculated by summarizing with some function, the values of the neighboring cells in the input grid. Zonal operators receive two input raster data fields, such that one of them plays a reference role, and produce a table where all the values of the cells in the non-referential grid are summarized (taking into account the values in the referential field) using some function. Map Algebra operators have been implemented in different Geographic Information Systems (GIS), like ArcGIS. Extensively used due to its simplicity, raster representation has some problems. For instance, given that the space is arbitrary partitioned without considering the sampled points, the true sampled values may be lost. Therefore, alternatives have been proposed. Ledoux and Gold (2006) redefined the Local, Focal and Zonal operations to support Voronoi diagrams, also giving methods for creating a 3D discrete Voronoi diagrams. In these proposals, however, no mixed representations are allowed (that means, all arguments must be of the same type). Other discrete representations exist, like TIN, usually representing an “altitude” field (Kumler, 1994). We are not aware of any algebra supporting TIN. Mennis et al. (2005) generalized Map Algebra, adding a temporal dimension. Also Mennis (2010) proposed a Multidimensional Map Algebra (MMA), an extension to Map Algebra that allows working with raster data of different dimensionality (although no other kinds of tessellations). The algebras above have some important drawbacks: (a) there is not a formal definition for them; (b) they are not closed; (c) input grids must have identical extents and cell sizes (in the case of raster data); (d) The arguments of the operators must be fields of the same kind. Thus, our first contribution is a formal generic framework for field data, independent of the underlying representation. We present a data model and a closed algebra for fields such that at a low level of abstraction, fields can be of any kind (e.g., raster and Voronoi, Voronoi and TIN). New representations may appear in the future, and they will still be supported by this algebra. A key contribution of our work is that we show that fields can be considered a particular kind of data cube, over which the typical OLAP operators can be applied. We go further, and show that mapping mechanisms, based on the work by Abelló et al. (2002), allow us to redefine the Drill-Across operator, and analyze together field cubes and traditional OLAP cubes. To the best of our knowledge this is the first proposal that formalizes the many map algebras proposed so far, and that introduces the idea that fields are actually data cubes, therefore they can be queried together with other cubes by using typical cube operators, provided that certain mappings are defined.

The paper is organized as follows. We first describe the discrete data model we use in the paper to represent continuous fields. Then, we define the algebra to operate with these “discretized” fields, based on the classic map algebra operators (Tomlin, 1990), after which we describe the multidimensional data model
Text Mining to Define a Validated Model of Hospital Rankings
[www.igi-global.com/chapter/text-mining-definemodeled-validated-model/10186?camid=4v1a](www.igi-global.com/chapter/text-mining-definemodeled-validated-model/10186?camid=4v1a)

Automated Integration of Heterogeneous Data Warehouse Schemas
[www.igi-global.com/article/automated-integration-heterogeneous-data-warehouse/1815?camid=4v1a](www.igi-global.com/article/automated-integration-heterogeneous-data-warehouse/1815?camid=4v1a)