Spatio-Temporal OLAP Queries
Similarity Measure and Algorithm

Olfa Layouni, Université de Tunis, ISG-Tunis, BESTMOD, Tunis, Tunisia
Jalel Akaichi, College of Computer Science, University of Bisha, Bisha, Saudi Arabia

ABSTRACT

Spatio-temporal data warehouses store large volumes of consolidated and historized multidimensional data, to be explored and analyzed by various users in order to make the best decision. A spatio-temporal OLAP user interactively navigates a spatio-temporal data cube (Geo-cube) by launching a sequence of spatio-temporal OLAP queries (GeoMDX queries) in order to analyze the data. One important class of spatio-temporal analysis is computing spatio-temporal queries similarity. In this article, the authors focus on assessing the similarity between spatio-temporal OLAP queries in term of their GeoMDX queries. The problem of measuring spatio-temporal OLAP queries similarities has not been studied so far. Therefore, this article aims at filling this gap by proposing a new similarity measure and its corresponding algorithm. The proposed measure and algorithm can be used either in developing query recommendation, personalization systems or speeding-up query evolution. It takes into account the temporal similarity and the basic components of spatial similarity assessment relationships.

KEYWORDS

Data Warehouse, Geo-Cube, GeoMDX Queries, Spatial Data, Temporal Data

INTRODUCTION

Data warehouse and OLAP tools provide access to historical data in order to support and improve the productivity of the decision-making process. In a data warehouse, data are organized and manipulated in accordance with concepts and operators provided by a multidimensional data model which allows to express user’s needs by exploiting a data cube (Inmon, 2002; Marketos, 2009). A data cube allows data to be modeled and viewed in multiple dimensions where each dimension is implemented by adopting a star or a snowflake schema model. Adding to that, Bill Inmon (Inmon, 2002) mentioned that in data warehouses, data analysis is performed interactively by applying a set of OLAP (On Line Analytical Processing) operators and tools that represent an effective solution for Business Intelligence. OLAP tools allows users to explore and analyze data in different perspective and level of aggregation by using special operators.

Nowadays, the popularity of spatial data, such as names of cities, postal codes, the position of individual objects in space, the maps created from satellite images, can be stored in a spatial data warehouse. Stefanovic et al. (Stefanovic et al., 2000) defined a spatial data warehouse as a collection of spatial and thematic, integrated, nonvolatile and historical data to support the spatial decision making process. A spatial data warehouse contains both spatial and alphanumeric data types. It is made from a multidimensional spatial model which defines the concepts of spatial measures and dimensions in order to take into account the spatial component. According to Caron (Caron, 1998)

DOI: 10.4018/IJDWM.2019040102

Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
OLAP tools don’t give the possibility to analyze and explore spatial data in order to help the user in his decision making. The SOLAP (Spatial OLAP) has been identified as an effective way to interrogate and explore data from a spatial data warehouse. Also, it enables to enhance analysis capabilities of a traditional OLAP system, by combining both the multi-dimensional analysis system with the cartographic visualization of Geographic Information Systems (GIS) (Badard, 2011). According to Bardar (Badard, 2011) a SOLAP server was defined as a visual platform realized spatially to support the exploration and to analyze spatio-temporal data in a rapid and flexible way. It is realized by using a multidimensional approach and taking account of different aggregations via a cartographic, graphic and tabular display types.

A ST-OLAP system has the same functionality and uses the same tools as a SOLAP system. Besides, a ST OLAP explores a spatio-temporal data warehouse, which we find a moving data type for the movement of an object in the time. So, moving objects are geometries that change their position and shape continuously over the time; the time could be an instant or a set of time intervals. In order to support spatio-temporal data, a data model and associated query language is needed for supporting moving objects. A ST-OLAP system is the enhancement for considering moving objects, so it takes into account the spatial objects which evolve over time. For the exploration of a data cube, the user should have operations and a manipulation language. Bardar (Badard, 2011) defined that the language used for a spatial data warehouse and a spatio-temporal data warehouse is the MDX (Multi-dimensional eXpressions) with spatial functions language, called also GeoMDX. In this case, a query can contain spatial, no spatial and temporal data types. The GeoMDX language represents the evolution of MDX language to support spatial and temporal data types.

ST-OLAP users interactively navigate a spatio-temporal data cube by launching sequence of ST-OLAP queries, which is often tedious since the user may have no idea of what the forthcoming query should be. Adding to that, spatio-temporal data cubes store a big amount of data that have become increasingly complex to be explored and analyzed. The notion of similarity has been considered as an important component for the development of recommendation systems. In our context, similarity measures are used to identify the degree of similarity between two ST-OLAP queries. To the best of our knowledge, there is no proposed similarity measure between ST-OLAP queries (GeoMDX queries (Tranchant, 2011)). So, in this paper we aim at filling this gap.

The remainder of this paper is structured as follows. Section 2 briefly reviews related work; this section presents the different similarity assessment models proposed in the literature for comparing between queries. Section 3 presents the basic definitions in the context of spatio-temporal data warehouses and ST OLAP systems. Section 4 presents our proposal of the new spatio-temporal similarity measure and the new algorithm for the similarity measures between ST-OLAP queries, section 5 presents the performance evaluation. Finally, Section 6 concludes this paper.

BACKGROUND: COMPARING QUERIES

Comparing queries has attracted a lot of attention in different areas like information retrievals (Marcel et al., 2012; Sapia, 2000; Jerbi, 2012; Sarawagi, 2000; Aligon et al., 2014), bio-informatics (Moreau et al., 2008), etc. We note that the most proposed approach focused on assessing the similarity between queries. This section reviews the literature for similarity functions that could possibly be used to compare ST-OLAP queries. We note that a ST-OLAP query can contain one or more spatial relations. So, in this section we begin by presenting some methods for comparing queries. Then, we present some methods proposed for comparing queries based on spatial relations.

Comparing OLAP Queries

In the literature, we find two different motivations that could be used for comparing queries. The first one is query optimization (Sapia, 2000). This motivation is based on comparing a query \( q \) to another \( q' \) in order to find a better way to evaluate the query \( q \). The second motivation is the most interesting
Related Content

Towards Comparative Mining of Web Document Objects with NFA: WebOMiner System
[www.igi-global.com/article/towards-comparative-mining-web-document/74753?camid=4v1a](www.igi-global.com/article/towards-comparative-mining-web-document/74753?camid=4v1a)

The Pedagogy of English Teaching-Learning at Primary Level in Rural Government Schools: A Data Mining View
P. Sunil Kumar, Sateesh Kumar Pradhan and Sachidananda Panda (2016). *Research Advances in the Integration of Big Data and Smart Computing* (pp. 105-119).
[www.igi-global.com/chapter/the-pedagogy-of-english-teaching-learning-at-primary-level-in-rural-government-schools/139398?camid=4v1a](www.igi-global.com/chapter/the-pedagogy-of-english-teaching-learning-at-primary-level-in-rural-government-schools/139398?camid=4v1a)
Enhancing the Process of Knowledge Discovery in Geographic Databases Using Geo-Ontologies
www.igi-global.com/chapter/enhancing-process-knowledge-discovery-geographic/7577?camid=4v1a

Management of Medical Website Quality Labels via Web Mining
Vangelis Karkaletsis, Konstantinos Stamatakis, Karampiperis, Karampiperis, Pythagoras Karampiperis and Pythagoras Karampiperis (2009). Data Mining and Medical Knowledge Management: Cases and Applications (pp. 206-226).
www.igi-global.com/chapter/management-medical-website-quality-labels/7534?camid=4v1a