Expert Systems With Applications 59 (2016) 119–144

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Hair-oriented data model for spatio-temporal data representation

Abbas Madraky[a,*], Zulaiha Ali Othman[b], Abdul Razak Hamdan[b]

[a] Computer Engineering Department, Kavosh University, Mahmoudabad, Mazandaran, Iran
[b] Data Mining and Optimization Research Group (DMO), School of Computer Science, Faculty of Information Science and Technology, National University of Malaysia (UKM), Malaysia

A R T I C L E I N F O

Article history:
Received 10 January 2015
Revised 21 April 2016
Accepted 22 April 2016
Available online 25 April 2016

Keywords:
Spatio-temporal data models
File size reduction
Query execution time
Hair-oriented data model
Nested tables

A B S T R A C T

Having an effective data structure regards to fast data changing is one of the most important demands in spatio-temporal data. Spatio-temporal data have special relationships in regard to spatial and temporal values. Both types of data are complex in terms of their numerous attributes and the changes exhibited over time. A data model that is able to increase the performance of data storage and inquiry responses from a spatio-temporal system is demanded. The structure of the relationships between spatio-temporal data mimics the biological structure of the hair, which has a 'Root' (spatial values) and a 'Shaft' (temporal values) and undergoes growth. This paper aims to show the mathematical formulation of a Hair-Oriented Data Model (HODM) for spatio-temporal data and to demonstrate the model's performance by measuring storage size and query response time. The experiment was conducted by using more than 178,000 records of climate change spatio-temporal data that were implemented in implemented in an object-relational database using nested tables. The data structure and operations are implemented by SQL statements that are related to the concepts of Object-Relational databases. The performances of file storage and execution query are compared using a tabular and normalized entity relationship model that engages various types of queries. The results show that HODM has a lower storage size and a faster query response time for all studied types of spatio-temporal queries. The significances of the work are elaborated by doing comparison with the generic data models. The experimental results showed that the proposed data model is easier to develop and more efficient.

1. Introduction

Spatio-temporal data are complicated and exist in large sets, because this data type is usually allocated to large areas, and the volume of data generated tends to increase over time (Rakêt & Markussen, 2014). Furthermore, real-world geographic objects and their relationships increase the complexity of such data (Parent, Spaccapietra, & Zimányi, 2006a). One important concern regarding spatio-temporal models is the approach used to add the time and space dimensions to the model. This concept refers to the necessary independence among the modeling dimensions of the data structures, namely, space and time (Parent, Spaccapietra, & Zimányi, 1999). However, problem solving is required when using spatio-temporal data in decision making (Triglav, Petrovič, & Stopar, 2011). The demand of having a decision support system for spatio-temporal data leads to an increase in the system's ability about data mining. The use of un-normalized tables for data mining causes data redundancy (Han, Kamber, & Pei, 2006).

The data definition/manipulation concerned on how spatio-temporal data are defined, store and retrieved, the execution time for storing, retrieving and querying the data (Wikle, 2015). Design and development of robust spatio-temporal data structures are the fundamental issues for spatio-temporal data handling. As a first issue, the volume of spatio-temporal data has fast growing and the relationships between spatio-temporal values are also increase data volume due to data redundancy. Data redundancy is the repetition of data leads to data size increasing. If the data size is increased, data manipulation has more difficulties because of the data replication and complexity (Perumal, Velumani, Sadhasivam, & Ramaswamy, 2015). Reducing data redundancy is a general solution for data manipulation problems that leads to decrease data volume and index managing costs. The second issue is about performance of the data model while querying especially response time. The response time problem has occurred due to inappropriate indexing regards to large number of records for data processing. The time based values are gradually added to the spatio-temporal database and the index management will be more

* Corresponding author.
E-mail addresses: madraky@yahoo.com (A. Madraky), zao@ukm.edu.my (Z.A. Othman), arh@ukm.edu.my (A.R. Hamdan).

http://dx.doi.org/10.1016/j.eswa.2016.04.028
0957-4174/

difficult when the number of rows in a table is increased. Therefore, the main reason of query response time problem is about indexing (Guo, Papaioannou, & Aberer, 2014).

Therefore, the reduction of redundancy and inconsistency is an additional issue that should be considered by using suitable storage and retrieval systems. Reliable and on-time query responses are important aspects of very large databases (de Caluwe & Bordogna, 2004). This issue increases in importance with increasing data volume, because consistent query answering is an important data accuracy factor with respect to data redundancy (Del Mondo, Rodríguez, Claramunt, Bravo, & Thibaud, 2013). Additionally, the data access time increases with data volume, resulting in longer-than-expected query response times (Bertossi, 2011). A suitable data structure design is a common solution for data growth problems, because it leads to a reduction in data redundancy and inconsistency, resulting in increased data integrity and reliability (Ramakrishnan & Gehrke, 2003).

In this paper, we use a bio-inspired data model to reduce the above-mentioned problems. Many structures and functions in the various branches of science are designed based on biological principles (Fei & Ma, 2007). Experience has shown that nature adopted ideas that require fewer tests due to the problems being gradually removed due to evolution and optimization over time (Steer, Wirth, & Halgamuge, 2009). Additionally, bio-inspired ideas are interesting and attractive to proposers or users, but these plans also have other benefits in application. One of the advantages of bio-inspired models is understandability, because it is easier to explain a biological model to experts or users due to their previous knowledge of the natural subject (Floreano & Mattiussi, 2008).

Furthermore, system predictability is increased when using a bio-inspired model, because it is possible to forecast behavior or results by comparing the model to the natural system. This ability is very useful for system analysis or risk avoidance (de Castro, 2007). Therefore, many biological models are utilized in computing algorithms or when designing plans for improving or deploying the algorithms (Chiong, 2009). Some famous examples of such models are molecular, DNA-based and nano-scale structures that can be found in the design field and in neural networks, as well as algorithms based on ant colonies, bees and bats that are found in computing fields (Hasançebi, Teke, & Pekcan, 2013; Zomaya, 2006).

Spatio-temporal data possess two main properties. The first property refers to data being in a particular place, because the data are allocated to specific positions due to spatial properties, such as coordinates, GIS specifications and geographic values (Sagar, 2013). The second property refers to the continual insertion of new data over time (Parent et al., 1999). For example, daily temperature values of a region over a month could be saved and processed using a time-series. We will review several data models related to these issues in our concept-related literature study, which includes entity-relational, object-oriented and object-relational conceptual models.

The simplest general data structure for representing data models is the entity-relational category, which is designed based on tables and relationships. MADS (Modeling of Application Data with Spatio-temporal features) is an example of an entity-relational data model that exists in the spatio-temporal environment. The structural dimension of MADS includes attributes, functions, integrity constraints, n-ary relationships, is-a links and aggregation links (Parent et al., 1999). This model has the ability to store and retrieve hierarchical data and semantics of measurements. Using the MADS conceptual model allows for a multi-representation data model, which has been presented in previous research (Parent, Spaccapietra, & Zimányi, 2006b) and is named MurMur. The main goal of the MurMur project has been to propose a new approach to the manipulation of geographical databases. The MurMur query interface is supported by a query editor tool, which provides facilities that can be used to formulate a spatio-temporal query. MurMur also has MADS limitations. In regard to another entity-relational data model, a graph-oriented spatio-temporal data representation has been proposed in previous research (Del Mondo et al., 2013). The model is implemented by extending the relational database specification. Entity-relational data models focus on data manipulation and query processing, but they possess a collection of normalized tables that must be joined, aggregated and transformed in the process of data mining, representing costly and complex tasks (Ordonez, Maabout, Matusevich, & Cabrera, 2013). Previous reports also have not covered object-oriented concepts due to the complexity of spatio-temporal data types.

The second category of data models is object-oriented. This category is based on real-world object properties, which are suitable for spatio-temporal areas. A formal object-oriented data model for temporal information was proposed in 2003 (Grandi & Mandreoli, 2003). Types, subtypes and schema versions of this model are represented by semantic class. A fuzzy spatio-temporal model was presented as an object-oriented data model in previous research (Ribarić & Hrkać, 2011). This model focuses on knowledge representation in high-level Petri nets, and it is suitable for designing a knowledge base in real-time and multi-agent-based intelligent systems. The model may include expert or user human-like knowledge. The main feature of the model is its knowledge representation scheme, called FuSpaT, which supports representation and reasoning for domains that include imprecise and fuzzy spatial, temporal and spatio-temporal relationships. Other fuzzy spatio-temporal models have been presented in previous research (de Tré, de Caluwe, Hallez, & Verstraete, 2003; Sözer, Yazıcı, Oğuztüzün, & Taş, 2008; Verstraete, Tré, & Hallez, 2006). A newer object-oriented data model in the spatio-temporal field was proposed in 2011 for mining monitor-based data (Yang-Ming & Qin-Lin, 2011). This presentation is suitable for an object-oriented programming environment, and the data models in this category are also able to cover complex tasks with a corresponding increase in the costs of time and space (Park, Whang, Lee, & Song, 2001).

Object-relational modeling is the third category, and this approach is designed based on both features of the two previous categories. Most of the recent research in spatio-temporal modeling is concerned with object-relational data models (Chau & Chittayasothorn, 2008; Cuevas, Marín, Pons, & Vila, 2008; Harrington, 2010; Mok, 2007; Philippi, 2005; Yu, Davis, Wilson, & Cole, 2008). A conceptual model for temporal data warehouses was proposed in previous research (Malinowski & Zimányi, 2008). The conceptual model used multidimensional definitions for time values, in terms of other attributes. In 2009, Zhou et al. proposed another object-relational prototype for representation, organization and access to disaster information (Zhou, Liu, Fu, & Zhang, 2009). In 2013, Le et al. presented a geosciences data model that referred to space and time (Le, Gabriel, Gietzel, & Schaeben, 2013). This model is based on topology-generalized maps that use geo-objects, and it also uses a relational structure for events and space-time relationships. Madraky et al. proposed a spatio-temporal data model for data analysis that was named the Hair-Oriented Data Model (Madraky, Othman, & Hamdan, 2012). All of the object-relational data models are compatible with relational DBMS, and they also accept the object-oriented features in relationships and operations.

In this paper, we extend a certain approach to this model and manipulate spatial and temporal objects to reach better performances on the use of storage and data querying. We combine the spatio-temporal specifications with the natural properties and functions of hair to create a better knowledge representation that is based on object-relational features. The specific contributions of our work include a reduction of data redundancy and database size to obtain more integrity and a decrease in query execution time in relevant categories. Additionally, we present a formal definition of