Spatio-Temporal Indexing - Current Scenario, Challenges and Approaches

Aditya Telang, Deepak Padmanabhan, Prasad Deshpande
IBM Research – India
Bangalore, INDIA
{adtelang, deepak.s.p, prasdesh}@in.ibm.com

1. MOTIVATION

With rapid advancements in computing hardware, tracking devices such as GPS receivers and sensors have become pervasive, generating a large amount of spatio-temporal data, such as measurements of temperature, pressure, air quality, traffic, etc. using sensors, GPS data from mobile phones and data from radars that capture location information about people and other moving objects such as cars and aeroplanes. This has enabled a wide variety of spatio-temporal applications, resulting in a renewed interest in techniques for handling spatio-temporal data. Over the past two decades or so, a large number of indexes for supporting spatial, temporal and spatio-temporal data have been independently proposed in the database and data mining communities. However, there exists no clear-cut guidelines or a prescriptive formula for pointing out which index should be chosen when specific needs of the underlying application are known. In addition, since spatio-temporal indexes have been proposed under various domains, it is hard for researchers and practitioners to determine whether some specified indexes are indeed available to address the problem at hand. For instance, an index like PO-Tree [7] is suitable for monitoring static spatio-temporal objects (such as sensors, cell-phone towers, etc.) but it is completely undesirable for handling moving object data (e.g., location tracking of cell-phone users, GPS tracking of vehicles and so on). Likewise, if the semantics of the application require indexing trajectories of moving objects, only a specific set of indexes (such as PA-Tree [6]) are useful whereas others such as (APR-Tree [3]) are undesirable.

We design this tutorial to expose the audience to the vast reservoir of spatio-temporal indexing techniques that are available in literature. In addition, apart from introducing the various indexes, our aim is to analyze the pros and cons of different indexing mechanisms when applied to various diverse scenarios. Given the large recent interest in spatio-temporal data analytics among corporates and academia, we hope that this tutorial is well-positioned in time to enhance and enrich the understanding of spatio-temporal data processing. Further, we think that the subject matter of this tutorial is a perfect fit for the COMAD conference that has a focused track for data management and its disciplines.

The 18th International Conference on Management of Data (COMAD), 14th-16th Dec, 2012 at Pune, India.

2. TUTORIAL ORGRANIZATION

We propose to organize this tutorial for a duration of 3 hours. A brief outline of the organization of the tutorial is as follows:

1. **Motivation**: (10 minutes)
2. **Spatial Indexing**: (30 minutes)
   - Spatial Data Types
   - Spatial Query Categories
   - Classification of Spatial Indexes
     - Grid-based technique
     - Tree-based technique
   - Analysis of Different Spatial Indexes (such as Geodesic Grid, R Tree, KD-Tree and so on)
     - Semantics of each index
     - Typical Usage
     - Applicability for real-time applications
       - Pros & Cons
3. **Temporal Indexing**: (20 minutes)
   - Need for temporal indexing
   - Types of Temporal Indexes
     - Semantics of each index
     - Typical Usage
     - Applicability for real-time applications
       - Pros & Cons
4. **Spatio-temporal Indexing**: (75 minutes)
   - Motivation & Basic Techniques
     - Native Space Indexing
     - Parametric Space Indexing
   - Types of Spatio-Temporal Indexes
     - Semantics of each index
     - Typical Usage
     - Applicability for real-time applications
       - Pros & Cons
5. **Comparative Analysis**: (35 minutes)
   - Different Spatio-Temporal Application Scenarios
     - Application of Different Indexes
     - Implications
     - Pros & Cons

6. **Conclusion & Discussion** (10 minutes)
    - Summary of the Tutorial
    - Pointers to Exciting New Problems

A set of transparencies(in PDF format) and a recommendation of papers will be made available to the participants.

3. TUTORIAL CONTENT

Here we present a detailed description of the material presented in this tutorial.

3.1 Motivation

In this segment, we introduce the problem of managing and handling spatio-temporal data. We present the different types of query scenarios that are typically posed on such data and illustrate the need for indexing mechanisms for organizing this data for effective retrieval of results. We provide a brief overview of the different contexts in which spatio-temporal data management has been addressed i.e., organizing historical data for analysis, warehousing data for mining, maintaining real-time data for frequent updates and queries, isolating and organizing trajectory data as well as individual data points for moving objects, and so on.

3.2 Spatial Indexing

In this part of the tutorial, we dig deeper into different types of spatial indexes (such as Geodesic Grid, R-tree [4], R+Tree [12], R*Tree [11], KD-Tree [8] along with its derivatives such as the Quad-Tree [9] and Oct-Tree [5]) that have been purely proposed for organizing different kinds of spatial data such as – surface of the earth (e.g., volcanic zones, earthquake regions, etc.), natural entities (e.g., forests, rivers, etc.), man-made entities (e.g., universities, castles, etc.) and moving spatial entities (e.g., cars on roads, ships in oceans, etc.). We discuss the semantics associated with each index . Specifically, we demonstrate how each of this index behaves when subjected to standard paradigms of spatial querying i.e., range queries and k-nearest neighbour queries. Further, we also provide insights as to which index to select (i.e., either a grid-based or a tree-based) depending on the needs of the problem setting.

3.3 Temporal Indexing

Similar to spatial indexing, temporal indexing has received a lot of attention for organizing database tuples based on their timestamps. We briefly touch base with Allen's Algebra [2] in order to understand the type of temporal queries typically issued on databases. Accordingly, we survey the different temporal indexing techniques and their performance aspects when handling such queries.

3.4 Spatio-Temporal Indexing & Comparative Analysis

This section forms the core component of this tutorial. We elaborate of the different types of indexing techniques for different kinds of spatio-temporal needs i.e., indexing data for statistical analysis, organizing trajectory-related data, managing data with respect to constantly moving and frequently updating objects, and so on. We discuss the semantics of each of these techniques, and provide a comparative analysis of different spatio-temporal indexing mechanisms (such as the TPR-Tree [10], the TPR*-Tree [13], the COLR-Tree [1], the Q+R-Tree [14] and others such as RT-Tree, 3DR-Tree, MV3R-Tree, HR-Tree, etc [3]) with respect to their performance, their ability to support range and kNN queries, and their overall applicability to different kinds of real-time monitoring of moving objects in the context of a spatio-temporal setting.

3.5 Analysis and Conclusions

Here, we summarize the contents of the tutorial and present a various pointers for future work.

4. TARGETED AUDIENCE & EXPECTATIONS

This tutorial is mainly targeted at several kinds of audience such as researchers, graduate students and industry professionals working in and/or interested in the area of handling, maintaining and working with spatio-temporal data in the context of real-time applications. The tutorial is organized in a self-contained way and does not assume any particular expertise from the audience. At the end of the tutorial, we hope that the attendees will be equipped with insights into different aspects involved in indexing spatio-temporal data, and would have a clear picture in terms of what indexing techniques to select for specific needs of applications using such data. We attempt our best to maintain a striking balance between theoretical concepts and practical importance of the problems in the tutorial. Thus, we hope that practitioners also get benefited from this tutorial.

5. BRIEF BIOGRAPHY

Aditya Telang: Aditya is a researcher at IBM Research India since 2011. Prior to joining IBM, he finished his PhD from University of Texas at Arlington. His current research interests include Spatio-Temporal Data Analytics, Information Management, and Business Analytics.

Deepak Padmanabhan: Deepak works with the Information Management Group at IBM Research India at Bangalore. He obtained his masters degree from IIT Madras prior to joining IBM.

Prasad Deshpande: Prasad M Deshpande is a Senior Researcher at IBM Research - India and Manager of the Information Analytics group. His areas of expertise lie in data management, specifically data integration and warehousing, OLAP, data mining and text analytics. He received a B. Tech in Computer Science and Engineering from IIT, Bombay and a M.S. and Ph.D. in Database systems from the University of Wisconsin, Madison. He has worked at several companies, including startups, IBM Almaden Research Center and currently at IBM Research - India. He has more than 35 publications in reputed conferences and journals and has several patents to his name. He has served on the Program Committee of many conferences, most recently being the PC Chair for COMAD 2011 and ACM Compute 2010.

6. REFERENCES

[1] Y. Ahmad and S. Nath. Colr-tree: Communication-efficient spatio-temporal indexing for a sensor data web portal. In *ICDE*, pages 784–793, 2008.

[2] J. F. Allen. Maintaining knowledge about temporal intervals. *Commun. ACM*, 26(11):832–843, 1983.

[3] H.-J. Cho, J.-K. Min, and C.-W. Chung. An adaptive indexing technique using spatio-temporal query workloads. *Information & Software Technology*, 46(4):229–241, 2004.

[4] A. Guttman. R-trees: A dynamic index structure for spatial searching. In *SIGMOD Conference*, pages 47–57, 1984.

[5] C. L. Jackins and S. L. Tanimoto. Quad-trees, oct-trees, and k-trees: A generalized approach to recursive decomposition

of euclidean space. *IEEE Trans. Pattern Anal. Mach. Intell.*, 5(5):533–539, 1983.

[6] J. Ni and C. V. Ravishankar. Pa-tree: A parametric indexing scheme for spatio-temporal trajectories. In *SSTD*, pages 254–272, 2005.

[7] G. Nol, S. Servigne, and R. Laurini. The po-tree: a real-time spatiotemporal data indexing structure. In *11th International Symposium on Spatial Data Handling*, pages 259–270, 2005.

[8] B. C. Ooi. Spatial kd-tree: A data structure for geographic database. In *BTW*, pages 247–258, 1987.

[9] M. H. Overmars and J. van Leeuwen. Dynamic multi-dimensional data structures based on quad- and *k - d* trees. *Acta Inf.*, 17:267–285, 1982.

[10] S. Saltenis, C. S. Jensen, S. T. Leutenegger, and M. A. Lopez. Indexing the positions of continuously moving objects. In *SIGMOD Conference*, pages 331–342, 2000.

[11] T. K. Sellis. Review - the r*-tree: An efficient and robust access method for points and rectangles. *ACM SIGMOD Digital Review*, 2, 2000.

[12] T. K. Sellis, N. Roussopoulos, and C. Faloutsos. The r+-tree: A dynamic index for multi-dimensional objects. In *VLDB*, pages 507–518, 1987.

[13] Y. Tao, D. Papadias, and J. Sun. The tpr*-tree: an optimized spatio-temporal access method for predictive queries. In *Proceedings of the 29th international conference on Very large data bases - Volume 29*, VLDB '03, pages 790–801. VLDB Endowment, 2003.

[14] Y. Xia and S. Prabhakar. Q+rtree: Efficient indexing for moving object database. In *DASFAA*, pages 175–182, 2003.