14th International Conference, DASFAA'09, Brisbane, Australia, pp.780-783, 2009

Tutorial

Techniques for Efficiently Searching in Spatial, Temporal, Spatio-temporal, and Multimedia Databases

Hans-Peter Kriegel, Peer Kröger, Matthias Renz

Institute for Informatics
Ludwig-Maximilians-Universität München, Germany
Oettingenstr. 67, 80538 Munich, Germany
http://www.dbs.ifi.lmu.de
{kriegel,kroegerp,renz}@dbs.ifi.lmu.de

Abstract. This tutorial provides a comprehensive and comparative overview of general techniques to efficiently support similarity queries in spatial, temporal, spatio-temporal, and multimedia databases. In particular, it identifies the most generic query types and discusses general algorithmic methods to answer such queries efficiently. In addition, the tutorial sketches important applications of the introduced methods, and presents sample implementations of the general approaches within each of the aforementioned database types. The intended audience of this tutorial ranges from novice researchers to advanced experts as well as practitioners from any application domain dealing with spatial, temporal, spatio-temporal, and/or multimedia data.

1 Introduction

The management and analysis of spatial, temporal, spatio-temporal, and multimedia data is a hot topic in database research because such data types occur in many applications. Querying databases of such a content is very important for these applications. In recent years, a vast amount of research has been done to explore efficient solutions for answering similarity queries on these data types. However, the existing research mainly focuses on one data type only although many proposed approaches are conceptually rather similar. As a consequence, it is a complex task to keep track with current research results not only because of the large amount of existing approaches, but also because of very different vocabularies used to express similar concepts. This tutorial aims at providing a cooperate and comprehensive view of the state-of-the-art research in similarity search for spatial, temporal, spatio-temporal, and multimedia data by identifying the general algorithmic approaches common to all solutions. This will build the bridges between the various approaches proposed for the different data types and illustrate relationships among them.

2 Searching in Spatial, Temporal, Spatio-Temporal and Multimedia Databases

Real-world applications dealing with spatial, temporal, spatio-temporal, and multimedia data require efficient methods for similarity search. Example applications are shape

based similarity search and docking queries in Geographic and CAD databases, time-series matching, proximity queries on moving objects and image and video retrieval in multimedia databases. There are a bundle of problems emerging from the aforementioned applications that challenge the problem of designing efficient solutions for answering these basic query types.

In this tutorial, we provide a detailed introduction to basic algorithmic approaches to tackle the problem of answering similarity queries efficiently. In addition, we discuss the relationships between the approaches. Last but not least, we will present sample implementations of these basic approaches for each of the four data types mentioned above.

2.1 General concepts of query processing in complex structured data

The first part of the tutorial details the general algorithmic approaches for efficient similarity query processing including indexing and multi-step query processing. We start with the general concept of feature based similarity search which is commonly used to efficiently support similarity query processing in non-standard databases [1, 7, 9]. This section gives a general overview of state-of-the-art approaches for similarity query processing using indexing methods. Thereby, diverse approaches for different similarity query types are discussed, including distance range queries (DRQ), k-nearest neighbor queries (kNNQ), and reverse k-nearest neighbor queries (RkNNQ). Furthermore, this section addresses the concept of multi-step query processing. Here, we show the optimal interaction between access methods and multi-step query processing methods [18, 12] and give a survey of this topic showing the relationships among diverse existing approaches. Finally, the first part of the tutorial briefly sketches other types of queries that are specialized for a given data type, e.g. intersection queries for spatial objects.

2.2 Querying spatial, temporal, spatio-temporal and multi-media data

The second part of the tutorial gives an overview of the sample implementations of the previously presented basic approaches for each of the four data types mentioned above. It starts with sample solutions for query processing in spatial data implementing the general algorithmic approaches presented previously. The presented methods address queries on two- or three dimensional spatially extended objects, e.g. geographic data [16, 5], CAD data [11, 4] and protein data.

Next, the tutorial addresses query processing in time series data which is the most important data type among temporal data. Here, we sketch the general problem of indexing time series known under the term ”curse of dimensionality” and discuss diverse solutions for this problem, including dimensionality reduction and GEMINI framework [7]. In addition to matching based similarity query methods [8] we also discuss threshold based similarity search methods for time series data [3].

The sample implementations concerning spatio-temporal data mainly focuses on proximity queries in traffic networks, i.e. queries on objects moving within a spatial network. This section introduces techniques enabling efficient processing of proximity queries in large network graphs. In particular, we discuss solutions that are adequate for densely populated [17] and sparsely populated traffic networks [19, 13]. The later case

requires graph embedding techniques that allows the application of multi-step query processing concepts.

Finally, this tutorial outlines sample solutions for query processing in multimedia data implementing the general algorithmic approaches presented in the first part of this tutorial, e.g. [10, 2]. In addition to similarity filter techniques that are specialized to multimedia data we also discuss how the multi-step query processing techniques can cope with uncertainty in multimedia data. Here we give an overview of sample solutions, e.g. [6, 14, 15].

References

1. R. Agrawal, C. Faloutsos, and A. Swami. Efficient similarity search in sequence databases. In *Proc. FODO*, 1993.
2. I. Assent, M. Wichterich, T. Meisen, and T. Seidl. Efficient similarity search using the earth mover's distance for large multimedia databases. In *Proceedings of the 24th International Conference on Data Engineering (ICDE), Cancún, Mexico*, 2008.
3. J. Aßfalg, H.-P. Kriegel, P. Kröger, P. Kunath, A. Pryakhin, and M. Renz. Similarity search on time series based on threshold queries. In *Proceedings of the 10th International Conference on Extending Database Technology (EDBT), Munich, Germany*, 2006.
4. J. Aßfalg, H.-P. Kriegel, P. Kröger, and M. Pötke. Accurate and efficient similarity search on 3d objects using point sampling, redundancy and proportionality. In *Proceedings of the 9th International Symposium on Spatial and Temporal Databases (SSTD), Angra dos Reis, Brazil*, 2005.
5. T. Brinkhoff, H. Horn, H.-P. Kriegel, and R. Schneider. A storage and access architecture for efficient query processing in spatial database systems. In *Proceedings of the 3rd International Symposium on Large Spatial Databases (SSD), Singapore*, 1993.
6. R. Cheng, S. Singh, S. Prabhakar, R. Shah, J. Vitter, and Y. Xia. Efficient join processing over uncertain data. In *Proceedings of the 15th International Conference on Information and Knowledge Management (CIKM), Arlington, VA*, 2006.
7. C. Faloutsos and M Ranganathan adn Y. Manolopoulos. Fast subsequence matching in time series database. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD), Minneapolis, MN*, 1994.
8. E. Keogh. Exact indexing of dynamic time warping. In *Proceedings of the 28th International Conference on Very Large Data Bases (VLDB), Hong Kong, China*, 2002.
9. F. Korn, N. Sidiropoulos, C. Faloutsos, E. Siegel, and Z. Protopapas. Fast nearest neighbor search in medical image databases. In *Proceedings of the 22nd International Conference on Very Large Data Bases (VLDB), Bombay, India*, 1996.
10. F. Korn, N. Sidiropoulos, C. Faloutsos, E. Siegel, and Z. Protopapas. Fast and effective retrieval of medical tumor shapes. In *IEEE Transactions on Knowledge and Data Engineering*, 1998.
11. H.-P. Kriegel, S. Brecheisen, P. Kröger, M. Pfeifle, and M. Schubert. Using sets of feature vectors for similarity search on voxelized cad objects. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD), San Diego, CA*, 2003.
12. H.-P. Kriegel, P. Kröger, P. Kunath, and M. Renz. Generalizing the optimality of multi-step k-nearest neighbor query processing. In *Proceedings of the 10th International Symposium on Spatial and Temporal Databases (SSTD), Boston, MA*, 2007.
13. H.-P. Kriegel, P. Kröger, M. Renz, and T. Schmidt. Hierarchical graph embedding for efficient query processing in very large traffic networks. In *Proceedings of the 20th International Conference on Scientific and Statistical Database Management (SSDBM), Hong Kong, China*, 2008.

14. H.-P. Kriegel, P. Kunath, M. Pfeifle, and M. Renz. Probabilistic similarity join on uncertain data. (best paper). In *Proceedings of the 11th International Conference on Database Systems for Advanced Applications (DASFAA), Singapore*, 2006.
15. H.-P. Kriegel, P. Kunath, and M. Renz. Probabilistic nearest-neighbor query on uncertain objects. In *Proceedings of the 12th International Conference on Database Systems for Advanced Applications (DASFAA), Bangkok, Thailand*, 2007.
16. J.A. Orenstein and F.A. Manola. Probe spatial data modelling and query processing in an image database application. *IEEE Trans. on Software Engineering*, 14(5):611–629, 1988.
17. D. Papadias, J. Zhang, N. Mamoulis, and Y. Tao. Query processing in spatial network databases. In *Proceedings of the 29th International Conference on Very Large Data Bases (VLDB), Berlin, Germany*, 2003.
18. T. Seidl and H.-P. Kriegel. Optimal multi-step k-nearest neighbor search. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD), Seattle, WA*, 1998.
19. C. Shahabi, M. R. Kolahdouzan, and M. Sharifzadeh. A road network embedding technique for k-nearest neighbor search in moving object databases. *Geoinformatica*, 7(3):255–273, 2003.