Similarity Searches in Time Series Data

Tomasz Nykiel

April 2, 2008

Abstract

In this paper, we propose an indexing data structure within a relational system to index multidimensional data. The index is used to store and retrieve information about regions called Minimum Bounding Rectangles. The technique is used as a preprocessing step in similarity searches over time-series data. The structure is multi-resolutional, storing information with respect to different window sizes. We define a mapping of the index to relational tables, which are then indexed with B+-tree indexes.

We incorporate declarative language queries in the retrieval process to achieve sequential access patterns instead of random. In particular, we propose two SQL-based procedures for the index retrieval process: *PrefixSearch* and *MultiMatching*. We investigate the query plans issued by the DBMS optimizer. The proposed technique is compared against other existing solutions with respect to very large datasets. We also consider scalability issues and possible future extensions.

Keywords: time-series databases, relational databases, SQL, similarity searches