IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITB10782

Chapter V

Indexing Multi-Dimensional Trajectories for Similarity Queries

Michail Vlachos, IBM T.J. Watson Research Center, USA

Marios Hadjieleftheriou, University of California - Riverside, USA

Eamonn Keogh, University of California - Riverside, USA

Dimitrios Gunopulos, University of California - Riverside, USA

Abstract

With the abundance of low-cost storage devices, a plethora of applications that store and manage very large multi-dimensional trajectories (or time-series) datasets have emerged recently. Examples include traffic supervision systems, video surveillance applications, meteorology and more. Thus, it is becoming essential to provide a robust trajectory indexing framework designed especially for performing similarity queries in such applications. In this regard, this chapter presents an indexing scheme that can support a wide variety of (user-customizable) distance measures while, at the same

time, it guarantees retrieval of similar trajectories with accuracy and efficiency.

Introduction

Multi-dimensional trajectory data are prevalent in diverse fields of interest, such as environmental information systems, meteorology, traffic supervision, wireless telephony, video tracking (Betke, 2002), and so forth. In such applications it is very common to record multiple measurements concurrently (for example, spatio-temporal datasets that record the *x* and *y* positions of objects over time). Therefore, any dataset that involves storage of data with multiple attributes can be considered and treated as a set of multi-dimensional trajectories. An example of 2-dimensional trajectories appears in Figure 1. Recent advances in wireless communications, sensor devices and GPS technology have made it possible to collect large amounts of spatio-temporal data, increasing the interest in performing data mining tasks (Barbara, 1999; Roddick, 2000). Examples include tracking animals, recording weather conditions, gathering human motion data generated by tracking various body joints (Shimada, 2000; Arikan, 2002), tracing the evolution of migrating particles in biological sciences and more.

Some very important data mining operations for multi-dimensional trajectories involve the discovery of objects that *move similarly* or others that *follow closely a given query pattern*. An important consideration for these operations is the *similarity/distance measure* that will be used for discovering the most appropriate trajectories (for example, Euclidean distance). A major difficulty that affects the choice of a good similarity measure is the presence of noise (introduced due to electromagnetic anomalies, transceiver problems and so forth). Another obstacle is that objects may follow similar motion patterns (spatial domain) but at different rates (temporal domain). Hence, the similarity models should be robust to noise, and should support elastic and imprecise matches. For example, in Figure 1 all three trajectories follow similar paths; quantifying the similarity of these paths depends on the similarity measure that will be used. In addition, it is clear that existing outliers might distort the "expected" value of the similarity between the trajectories.

Some of the most widely used similarity measures are functions with large computational cost that make similarity query evaluation a challenging task. Consequently, in order to produce query results promptly, it is crucial to speed up the execution of these functions. For that purpose, low-cost *upper/lower-bounding* approximations of the similarity functions can be used initially to help prune most dissimilar trajectories faster.

Related Content

Multidimensionality in Statistical, OLAP and Scientific Databases
Arie Shoshani (2003). *Multidimensional Databases: Problems and Solutions (pp. 46-68).*
www.irma-international.org/chapter/multidimensionality-statistical-olap-scientific-databases/26964

Ontology for Data Mining and its Application to Mining Incomplete Data
Hai Wangand Shouhong Wang (2008). *Journal of Database Management (pp. 81-90).*
www.irma-international.org/article/ontology-data-mining-its-application/3396

A Space-Efficient Protocol for Consistency of External View Maintenance on Data Warehouse Systems: A Proxy Approach
Shi-Ming Huang, David C. Yenand Hsiang-Yuan Hsueh (2007). *Journal of Database Management (pp. 21-47).*
www.irma-international.org/article/space-efficient-protocol-consistency-external/3373

Conceptual Modeling for XML: A Myth or a Reality
Sriram Mohanand Arijit Sengupta (2009). *Database Technologies: Concepts, Methodologies, Tools, and Applications (pp. 527-549).*
www.irma-international.org/chapter/conceptual-modeling-xml/7930

An Exploration of a Set Entropy-Based Hybrid Splitting Methods for Decision Tree Induction
Kweku-Muata Osei-Brysonand Kendall Giles (2004). *Journal of Database Management (pp. 1-17).*
www.irma-international.org/article/exploration-set-entropy-based-hybrid/3313