

Stardust: Data Stream Indexing for Sensor Networks *

Ahmet Bulut Ambuj K. Singh
Department of Computer Science
University of California Santa Barbara, CA, 93106
{bulut,ambuj}@cs.ucsb.edu

ABSTRACT

Monitoring thousands of data streams online poses a challenge in many data-centric applications such as telecommunications networks, traffic management, trend-related analysis, web-click streams, and sensor networks. Stream mining techniques employed in these applications have to be efficient in terms of space usage and per-item processing time, while providing a high quality of answers to similarity queries such as detecting correlations and finding similar patterns. We propose a new approach for summarizing a set of data streams, and for constructing a composite index structure to answer similarity queries. The features of a stream are extracted incrementally on the fly at multiple resolutions, and inserted into a family of index structures for later querying.

1. INTRODUCTION

A growing number of real world applications deal with multiple streams of data: performance measurements in network monitoring and traffic management, call detail records in telecommunications networks, transactions in retail chains, ATM operations in banks, log records generated by web servers, and sensor network data. Network management involves monitoring and configuring network hardware and software to ensure smooth operation. Sensors distributed across the network send measurements and alarms periodically to network operations centers. The operations centers use this data to monitor link bandwidth usage, to detect congestion, to detect the links that are utilized correlatively, to do load balancing and all in all to improve the utilization of network resources.

Design and implementation issues for building Data Stream Management Systems for sensor networks [2, 3] have been addressed in the database community. In our earlier work, SWAT [1], we examined how to summarize a single data stream in an online manner.

2. PROPOSED SOLUTION

In this paper, we propose a new scheme to summarize and index a set of data streams at operations centers. The features of data streams at multiple resolutions are extracted on the fly using Discrete Wavelet Transformation (*DWT*), and inserted into a high-dimensional index structure family. We reduce the maintenance cost of the index structure by computing approximation coefficients online: we compute

*Work supported partially by NSF under grants EIA-0080134 and ANI-0123985.

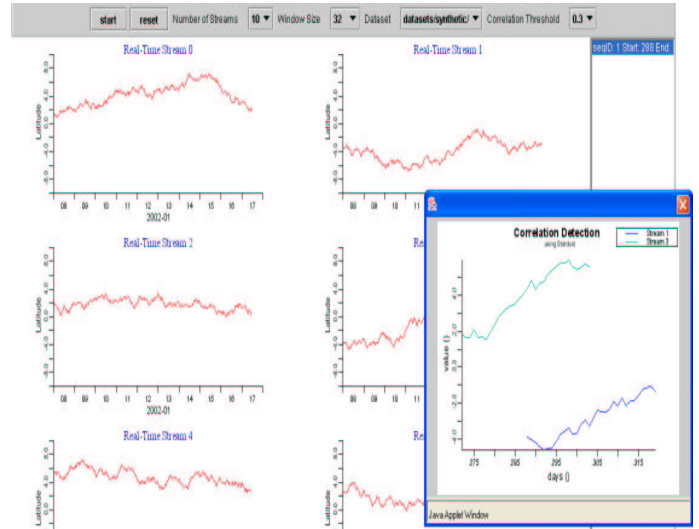


Figure 1: Stardust correlation detection framework

the coefficients on the materialized view of the continuous stream (on the index structure), thereby providing a one-pass algorithm also called a streaming algorithm for feature extraction. Our technique is cost effective in terms of its time and space requirements, and is parameterized: it can be tuned to trade accuracy for speed and space. Our focus is on similarity queries in the context of data streams, which is gaining a growing interest from the database community. New emerging data stream mining applications require that databases are enhanced with the capability to process similarity queries. These queries are useful for identifying correlations as shown pictorially in Figure 1 and discovering patterns in dynamic data streams. An online demo is available at <http://amazon.cs.ucsb.edu/~bulut/demo.html> for your reference.

3. REFERENCES

- [1] A. Bulut and A. K. Singh. SWAT: Hierarchical stream summarization in large networks. In *ICDE*, pages 303–314, 2003.
- [2] A. Deshpande, S. Nath, P. B. Gibbons, and S. Seshan. Cache-and-query for wide area sensor databases. In *SIGMOD*, pages 503–514, 2003.
- [3] S. Madden and M. J. Franklin. Fjording the stream: An architecture for queries over streaming sensor data. In *ICDE*, pages 555–566, 2002.