Welcome

Seminar Series of the Department of Computer Science Electrical Engineering

Presents

SketchTree: Approximate Tree Pattern Counts over Streaming Labeled Trees

October 26, 2007. Time 2:00-3:00PM. FH 557.

Praveen Rao
Assistant Professor
Computer Science and Electrical Engineering
School of Computing and Engineering
University of Missouri-Kansas City

Abstract

In recent years, there has been a rising interest in developing online approximation algorithms for data streams. Some of the key challenges are posed by the fact that streaming data can be read only once in a fixed order of arrival and only a limited amount of memory is available for storage. In this talk, I shall present the problem of approximately counting tree patterns over a stream of labeled trees (e.g., XML documents). I present a new approximation algorithm called SketchTree (ICDE '06) that computes a synopsis of the stream in a single pass by processing each tree only once. Using a limited amount of memory, SketchTree provides approximate answers for both ordered and unordered tree pattern counts. Theoretical analyses is provided to show that SketchTree has provably strong guarantees on the error bounds. Experiments on real datasets demonstrate that SketchTree can indeed estimate tree pattern counts within 10-15% relative error with high confidence under various situations.

Bio: Praveen Rao is an Assistant Professor in Computer Science Electrical Engineering Department at the University of Missouri-Kansas City. He received his BE degree in Computer Engineering from the University of Pune in 1999. He received his MS and PhD degrees in Computer Science from the University of Arizona in 2001 and 2007 respectively. For a year, he worked as a software developer at Amazon.com. His research interests include indexing, query processing, filtering, aggregation on large-scale semi-structured data (eg., XML). He is also interested in data management for scientific disciplines.