

Computer Science Seminar Series, 2009

National Capital Region

Finding Non-Trivial Patterns and Structural Similarity in Time Series Databases

Speaker: Prof. Jessica Lin
Department of Computer Science
George Mason University

Friday, September 18, 2009
1:00PM- 2:00PM, NVC 325

Abstract

Perhaps the most commonly encountered data type is time series. Apart from the obvious problem of handling the typically massive size of time series databases—gigabytes or even terabytes are not uncommon—most classic data mining algorithms do not perform or scale well on time series data. This is mainly due to the inherent structure of the data: high dimensionality and feature correlation. These intrinsic structural characteristics, combined with the measurement-induced noises that beset real-world time series data, pose challenges that render classic data mining algorithms ineffective and inefficient. The emphasis of this talk is on the discovery of important patterns in time series data. The previous body of work in this area has been mostly concentrated on the identification of previously known patterns. The major distinction of this work is that it offers the ability to discover important, unknown patterns in an effective and automated manner. We introduced SAX (Symbolic Aggregate approxImation), the first symbolic representation of time series that allows dimensionality reduction and lower-bounding distance measures. I will discuss SAX and its utilities, including recent work on finding structural similarity on time series data

Biography

Dr. Jessica Lin is an Assistant Professor of Computer Science at George Mason University. Her research interests are in temporal, spatiotemporal, and stream data mining. Her work focuses on efficient discovery of previously unknown and interesting patterns from massive temporal and spatiotemporal databases. Dr. Lin has been in the program committee for several international conferences in the area of data mining. She is the co-Chair of the 2008 and 2009 ICDM Workshops on Spatial and Spatiotemporal Data Mining, and Special Session on Data Mining for Geoinformatics at the 17th International Conference on Geoinformatics.