**THE CONCORDIA INSTITUTE FOR INFORMATION SYSTEMS ENGINEERING
IS PLEASED TO PRESENT THE FOLLOWING GUEST LECTURE IN
OUR CIISE DISTINGUISHED SEMINAR SERIES**

Dr. Xiangliang Zhang

King Abdullah University of Science and Technology (KAUST)

Mining Large-scale Streaming and Complex Data Sets

In this "big-data" era, vast amount of continuously arriving data can be found in various fields, such as sensor networks, network management, web and financial applications. Keeping terabytes to petabytes of data in memory is unacceptable. Therefore, the development of algorithms for processing large-scale streaming data instantaneously becomes highly important. In this talk, two novel algorithms for managing streaming data will be introduced.

The first algorithm, named StrAP, is an online clustering algorithm. It is able to summarize streaming data and extract the main patterns with quasi-linear complexity. An application to streaming jobs running on a grid computing system show that StrAP can be used as an online monitoring system to report abnormal usage and device errors.

The second algorithm was proposed to estimate the dynamic density over data streams, named KDE-Track as it is based on a conventional and widely used Kernel Density Estimation (KDE) method. KDE-Track solved the quadratic complexity issue of KDE and can timely catch the dynamic density of synthetic and real-world data. With the estimated density, KDE-Track was applied to outlier detection in sensor networks. Sensor errors can be identified and noisy data can be cleaned.

The large-scale complex data is more challenging to study. An algorithm for predicting the relevance of query-URL pairs in the largest Chinese search engine (Baidu.com) will be presented in the end of this talk. Through using a combination of click-through and mouse-trajectory data, the algorithm can model the interactions between users and search results, label query-URL pairs more accurately and provide more satisfactory search results are to users.

The presented work has been published at SIGKDD2009, CIKM2012 and AAAI2012, respectively.

Biography: **Xiangliang Zhang** is an assistant professor and directs the Machine Intelligence and knowledge Engineering (MINE) Laboratory in the Division of MCSE, King Abdullah University of Science and Technology (KAUST), Saudi Arabia. Prof. Zhang earned her Ph.D. degree in computer science from INRIA-University Paris-Sud 11, France, in July 2010. She received M.S. and B.S. degrees from Xi'an Jiaotong University, China, in 2006 and 2003, respectively. Dr. Zhang's research mainly focuses on developing automated methods for machine learning and data mining to discover and manage knowledge from complex and large-scale data sets for diverse applications. The general research goal is to enable computer machines to learn, such as building an autonomic cloud computing system to make the system manage itself, automatically detecting outliers/errors in a sensor network, enabling search engines to predict if users are satisfied by the retrieved results given their queries, efficiently searching in a graph database with millions of vertices and billions of edges. All these novel techniques, owing to the ability of learning, will facilitate our work and life.

Monday, August 6, 2012   **16:00 – 17:00**   **EV002.260**

Concordia
UNIVERSITY

1455 De Maisonneuve Blvd. West, EV007.640, Montreal, Quebec, Canada H3G 1M8
Tel 514-848-2424 ext. 5847 Fax 514-848-3171 www.ciise.concordia.ca