

NATIONAL UNIVERSITY OF SINGAPORE

School of Computing

C S S E M I N A R

Title: **Stream Data Mining and Applications: A Big Data Perspective**

Speaker: Professor Latifur Khan
 Department of Computer Science
 University of Texas at Dallas, USA

Date/Time: 23 July 2015, Thursday, 01:00 PM to 02:30 PM

Venue: Executive Classroom, COM2-04-02

Chaired by: Dr Zimmermann, Roger, Associate Professor, School of Computing
 (rogerz@comp.nus.edu.sg)

Abstract:

Data streams are continuous flows of data. Examples of data streams include network traffic, sensor data, call center records and so on. Data streams demonstrate several unique properties that together conform to the characteristics of big data (i.e., volume, velocity, variety and veracity) and add challenges to data stream mining. In this talk we will present an organized picture on how to handle various data mining techniques in data streams. Most existing data stream classification techniques ignore one important aspect of stream data: arrival of a novel class. We address this issue and propose a data stream classification technique that integrates a novel class detection mechanism into traditional classifiers, enabling automatic detection of novel classes before the true labels of the novel class instances arrive. Novel class detection problem becomes more challenging in the presence of concept-drift, when the underlying data distributions evolve in streams. In this talk we will show how to make fast and correct classification decisions under this constraint.

Furthermore, we will present a semi supervised framework which exploits change detection on classifier confidence values to update the classifier intelligently with limited labeled training data. We will present a number of stream classification applications such as website fingerprinting, real time monitoring, evolving insider threat detection and textual stream classification. This research was funded in part by USA National Science Foundation (NSF), NASA, Air Force Office of Scientific Research (AFOSR), Sandia National Lab (via DOE) and Raytheon.

Biodata:

Dr. Latifur Khan is currently a full Professor (tenured) in the Computer Science department at the University of Texas at Dallas, USA where he has been teaching and conducting

research since September 2000. He received his Ph.D. and M.S. degrees in Computer Science from the University of Southern California (USC) in August of 2000, and December of 1996 respectively. Dr. Khan is an ACM Distinguished Scientist. He has received prestigious awards including the IEEE Technical Achievement Award for Intelligence and Security Informatics. Dr. Khan has published over 200 papers in prestigious journals, and in peer reviewed conference proceedings. Currently, his research area focuses on big data management and analytics, data mining, complex data management including geo-spatial data and multimedia data. More details can be found at: www.utdallas.edu/~lkhan/