

UW CENTER FOR PATTERN ANALYSIS AND MACHINE INTELLIGENCE

GRADUATE SEMINAR SERIES

Strategies for Mining Big Data

Speaker: Ahmed Elgohary

Date: February 6, 2013

Time: 4:00 pm – 4:30 pm

Place: E5 (5128) Refreshments will be served

Abstract :

Recent advancements in Cloud-based storage and computing Infrastructure have facilitated collecting, maintaining and analyzing massive amounts of data. Businesses in various domains can dramatically benefit from information and insights gained by mining such amounts of data. Due to changes in the computing model, adapting data analytics algorithms to work efficiently and effectively on cloud computing environments turns out to be a non-trivial task. The special characteristics of data mining algorithms (e.g. memory intensiveness, iterative nature, etc) impose additional limitations and constraints on implementing data mining algorithms over that scale of data.

MapReduce and its open source implementation Hadoop are considered the most successful and widely deployed cloud analytics framework. In this talk, I will discuss the challenges of developing efficient and effective data mining algorithms using MapReduce. Then, I will present some approximation strategies that have been employed in previous works to optimize the execution of various data mining algorithms over MapReduce.