**COLLOQUIUM**

Jiayang Sun, Ph.D

Case Western Reserve University

**“Efficient Adaptive Mixture Estimation”**

Thursday, October 30, 2008 at 3:30 p.m. in CH 025

Refreshments will be served in Math 108 at 3:00 p.m.

**Abstract**

Data mining is important in scientific research, knowledge discovery and decision making. In this article, we develop and study new adaptive estimation procedures, the partial EM (PEM) and its Bayesian variants (BMAP & BPEM) for analyzing large data from heterogeneous populations. The adaptive procedures are computationally fast, and provide good alternatives to a full EM when a full EM procedure can and cannot be run, for large or streaming data. Under mild conditions, the PEM is further shown to be consistent and efficient, and has a “super-efficiency” in the case that a mixture distribution for the second batch of the data has an extra component (a.k.a.”contaminations'' or ``intrusions') to that of the first batch of the data.  The applications to a small data set and a network intrusion data set will also be shown.  (This is a joint work with Peng Liu and Jiahua Chen.)