Research Colloquium
Department of Computer Science

Matt Otey
Department of Computer Science and Engineering
Ohio State University

will speak on:

Approaches to Abnormality Detection with Constraints

11:00 a.m. on Friday, February 17 in Houser 108

A common problem in data analysis is that of discriminating between modes of normal behavior and modes of abnormal behavior. Of particular interest are techniques that can automatically detect abnormal activity in data. This is important since, for example, abnormal data may be indicative of measurement error in scientific data, or malicious activity in security audit data. Efficient detection of such abnormalities reduces the risk of making poor decisions based on erroneous data, and aids in identifying, preventing, and repairing the effects of malicious or faulty behavior. Furthermore, many data mining algorithms and techniques for statistical analysis may not work well in the presence of these abnormalities, as they may introduce skew or complexity into the model, making it difficult, if not impossible, to fit an accurate model to the data in a computationally feasible manner.

In this research we examine abnormality detection in the presence of constraints. Such constraints include (a) limits on the amount of time an algorithm can spend finding abnormalities, (b) dynamic data sets, (c) distributed data sets, and (d) data sets containing a mixture of categorical and continuous attributes. In this talk, I will discuss several different approaches to abnormality detection that we have developed that address these constraints in different ways. In particular, I will discuss novel dissimilarity measures for outlier detection, detecting network intrusions using programmable network interface cards, and LOADED, which is our approach for detecting anomalies in dynamic and distributed mixed-attribute data sets.