CCSU
DEPARTMENT OF MATHEMATICAL SCIENCES

DATA MINING
THESIS PRESENTATION

Friday, May 6
2:00 – 3:00 pm in MS 101

Comparing Unsupervised Multivariate Normal Cluster Results *between* Datasets and Consolidating Similar Clusters

Eric Taylor
(MS candidate in Data Mining)

Abstract

Data availability and/or computing constraints may necessitate unsupervised cluster analysis be performed using stratified subsets of the data. When the analysis is done, each stratified dataset has its own cluster results. How can the analyst compare and combine the information between datasets – while retaining the information gleaned from each? We will examine a method that compares (via Hotelling's T^2) multivariate normal cluster results *between* datasets and consolidates (via Bayes' Theorem) those that are similar. The result allows the analyst to perform incremental analyses on large/sporadic datasets – with a little upfront planning.

Refreshments will be served.

For further information:
gotchevi@ccsu.edu (860) 832-2839