

SE&T Colloquium Series-Winter 2012

Speaker	Dr. Timothy Havens Department of Computer Science and Engineering Michigan State University Hosts: Dr. Morteza Marzjarani
Title	<i>Incremental Fuzzy C-Means for Large Scale Data</i>
Abstract	<p>Since the early 1990's, the ubiquity of personal computing technology has produced an abundance of staggeringly large data sets—the Library of Congress has stored over 160 terabytes of web data, it is estimated that Facebook alone logs over 25 terabytes of data per day, and large bioinformatics data sets that integrate microarrays, sequences, and ontology annotations continue to grow. To compound this fact, these data sets are populated from disparate, often unknown, sources and are in a wide-range of formats. There is a great need for systems by which one can elucidate the similarity among and between groups in these next-generation data sets and compute meaningful clustering results. In this talk, I will discuss a method for efficiently and accurately approximating the solution of the fuzzy c-means (FCM) clustering algorithm using sampling and incremental algorithms based on weighted FCM. Comparisons will be made between literal FCM and the incremental algorithms on manageably sized data sets (~million objects), including magnetic resonance imaging volumes. I will then briefly discuss an extension of the incremental methods to kernel clustering and show some preliminary results. Last, I will apply the clustering methods to bioinformatics data composed of genes described by Gene Ontology annotations to show how they can be used for comparative genomics.</p>
Date	Tuesday, February 7
Time	4:10-5:00pm
Place	Pioneer 240
	Refreshments will be served at 4:00pm.