Math and Stat Colloquium

Thursday, February 19

3:00pm LIB 421

Refreshments will be served in the Lund Hall Foyer at 2:30pm

Speaker:

Michael W. Berry

Department of Electrical Engineering

& Computer Science

University of Tennessee, Knoxville

"Using Nonnegative Matrix and Tensor Factorizations for Topic and Scenario Detection and Tracking"

Abstract: Automated approaches for the identification and clustering of semantic features or topics are highly desired for text mining applications. Using low rank non-negative matrix factorizations (NNMFs) to retain natural data non-negativity, one can eliminate subtractive basis vector and encoding calculations present in techniques such as principal component analysis for semantic feature abstraction. Moving beyond two-way factorizations, we demonstrate how non-negative tensor factorizations (NNTFs) can be used to capture temporal and semantic proximity and thereby enable the tracking of both targeted and latent (previously unknown) discussions or communication patterns. Demonstrations of NNMF and NNTF algorithms for topic (or discussion) detection and tracking using the Enron Email Collection and the Airline Safety Reporting System (ASRS) corpus are provided. A Java-based information visualization environment called FutureLens, which incorporates NNTF outputs, is used to demonstrate the detection of terrorism-based plots/scenarios in the VAST 2007 Contest dataset.

Joint work with Andrey A. Puretskiy (University of Tennessee) and Brett W. Bader (Sandia National Laboratory)