CCSU
DEPARTMENT OF MATHEMATICAL SCIENCES

COLLOQUIUM

Friday, April 24
10:15 – 10:45 AM
Davidson Hall, Room 203

OPINION MINING OF UNSTRUCTURED TEXT WITH APPLICATION TO EXTRACTED USER ARTICLE COMMENT TEXT FROM THE NEW YORK TIMES WEBSITE

RICK ROUNTREE

(Data Mining MS Thesis Presentation)

CENTRAL CONNECTICUT STATE UNIVERSITY

Abstract: Opinion mining, also known as sentiment analysis, is the quantification of subjective intent from written material. This thesis presents the building, implementation and application of a sentiment analysis model to quantify the positive and negative opinion of comments posted to New York Times articles. New York Times article comments are accessible via Application Programming Interface (API) along with metadata associated with each comment. A representative sample of comments for 2012 and 2013 from the New York Times was classified as having positive or negative sentiment and used to build a logistic regression to predict sentiment for new comments. The final sentiment classification model shows ~80% expected classification accuracy with four input variables. Inputs to the model are two sentiment classification methods available in R and two significant comment metadata variables. When applied to two cohorts of extracted New York Times user article comments pertaining to the "The Patient Protection and Affordable Care Act" (ACA) and "Global Warming" the sentiment classification model shows promising results. The opinion expressed in New York Times comments over time for the ACA shows evidence of expected variation at significant milestones in the implementation of the ACA. For example, a clear increase in negative opinion occurs in and after October 2013 during the difficult implementation of the ACA. The global warming cohort shows seasonal changes in global warming opinion but the trend is not as pronounced as the changes seen in the ACA cohort.

For further information:
gotchevi@ccsu.edu 860-832-2839
http://www.math.ccsu.edu/gotchev/colloquium/