



Interdisciplinary Data Sciences



* IDSC Seminar Series *
March 3, 2017 3:00p.m.-4:00pm
Location: CUTR 102

<https://idsbigdata.com/>

Presents

Dr. Douglas W Oard

College of Information Studies and UMIACS
University of Maryland
College Park, MD 20742

Title: Extrinsic Evaluation of Text Classification for Policy Analysis based on Coding Human Values

Abstract: One way in which social scientists seek to analyze written content is to first assign labels to segments of text and then to perform statistical analysis on the results of that labeling process. Traditionally, such labels have been assigned by hand, which limits the scale at which this approach to content analysis can be performed. Deciding on which labels to assign, and how those labels should be assigned, is an inherently human process, but the actual assignment of labels is a process that is amenable to automation using text classification techniques. In this talk, I will describe experiments that we have conducted to characterize the degree to which sufficiently accurate classifiers can be built for one specific application: association of labels indicating human values to spans of text in written statements that were prepared for delivery to Congressional and Federal Communications Commission hearings on Net Neutrality. Our initial experiments sought to optimize traditional intrinsic measures of classifier effectiveness such as accuracy and the F measure, but in our more recent work we have started to examine whether further improvements in such intrinsic measures are predictive of improvements in the results of the policy analysis that motivated the classification task. We formulate that question by asking whether the same results from the policy analysis might have been achieved using less human effort, and if so how much human effort might have been saved. Our conclusion is that substantial savings are possible, in part because the actual task (a correlation analysis that associates labels with policy positions) is tolerant of some level of error in the label assignments. While this way of performing extrinsic (i.e., task-grounded) evaluation sheds light on the utility of our present classifiers, it also raises questions about how such extrinsic evaluation measures should be designed. I will therefore conclude the talk by summarizing the questions with which we are presently grappling, on which we would welcome suggestions from the audience. This is joint work with Emi Ishita, Yoichi Tomura, An-Shou Cheng, Ken Fleischmann and Yasuhiro Takayama.



Biography: Douglas Oard is a Professor at the University of Maryland, College Park, with joint appointments in the College of Information Studies and the University of Maryland Institute for Advanced Computer Studies (UMIACS). He is on sabbatical this year at the University of South Florida and the University of Florida. Dr. Oard earned his Ph.D. in Electrical Engineering from the University of Maryland. His research interests center around the use of emerging technologies to

support information seeking by end users. Additional information is available at <http://terpconnect.umd.edu/~oard/>.

IDSC Contact:
Dr. K. Ramachandran
University of South Florida
4202 E Fowler Ave, CMC317
Tampa, FL 33620-5700
E-mail: ram@usf.edu
Telephone: (813)-974-1270
Fax: (813)-974-2700