



NORTHWESTERN INSTITUTE
ON COMPLEX SYSTEMS (NICO)

CHAMBERS HALL
600 FOSTER STREET
EVANSTON, IL 60208-4057
T . 847 491 2527
F . 847 556 1280



NORTHWESTERN
UNIVERSITY

**Northwestern Institute on Complex Systems
invites you to its**

Weekly Seminar

Speaker: Daniel Diermeier, NICO Co-Director, Bei Yu, Post Doc, and
Stefan Kaufmann, Asst. Professor
Title: Mass Opinion - Extraction, Classification, and Measurement
When: Wednesday, February 7, 2007
Where: Chambers Hall, 600 Foster Street, Lower Classroom Level
Time: 12:00 – 1:00

*A light lunch will be available and a NICO Coffee Hour will follow for questions,
networking, and collaboration*

Abstract

In this talk we introduce opinion classification, a subarea of text classification, and our exploratory study of opinion classification in political and business domains. Different from topic classification, opinion classification aims to assign tones (positive, negative, or neutral) to text pieces. We define three dimensions for a taxonomy of opinion texts: 1) single or multiple subjects of comment; 2) subjective or objective writing style; and 3) consistent or conflicting ideologies in the author community. The pioneering opinion classification research used the online customer reviews as the testbed, which were characterized by single subject of comment, subjective writing and consistent ideology among authors. In consequence the previous research focused on the extraction of subjective adjectives and the construction of document opinions based on a simple additive model of individual opinion indicators. Recently opinion classification has been extended to other domains, such as the public comments in eRulemaking and business news. Our preliminary work in opinion classification of congressional speeches and business news articles demonstrate that the characteristics of opinion expression in these domains (e.g. objective writing and conflicting ideologies) strongly affect the classification strategies. After comparing the classifiers learned from movie reviews and speeches in the Senate and House debates, we found that "issue words" mostly represented by nouns are better indicators for classifying opinions of conflicting ideologies and objective writing styles. This result also poses the difficulties of learning general-purpose opinion classifier based on bag-of-words representation. We need to explore more complex linguistic features in order to catch generic opinion expressions.

Posted on Plan-It Purple