ABSTRACT

Title of Document:
SPIN: LEXICAL SEMANTICS, TRANSITIVITY, AND THE IDENTIFICATION OF IMPLICIT SENTIMENT

Stephan Charles Greene
Doctor of Philosophy, 2007

Directed By:
Professor Philip Resnik, Department of Linguistics and Institute for Advanced Computer Studies

Current interest in automatic *sentiment analysis* is motivated by a variety of information requirements. The vast majority of work in sentiment analysis has been specifically targeted at detecting subjective statements and mining opinions. This dissertation focuses on a different but related problem that to date has received relatively little attention in NLP research: detecting *implicit sentiment*, or spin, in text. This text classification task is distinguished from other sentiment analysis work in that there is no assumption that the documents to be classified with respect to sentiment are necessarily overt expressions of opinion. They rather are documents that might reveal a *perspective*. This dissertation describes a novel approach to the identification of implicit sentiment, motivated by ideas drawn from the literature on lexical semantics and argument structure, supported and refined through psycholinguistic experimentation. A relationship predictive of sentiment is established for components of meaning that are thought to be drivers of verbal argument selection and linking and to be arbiters of what is foregrounded or backgrounded in discourse. In computational experiments employing targeted lexical selection for verbs and nouns, a set of features reflective of these components of meaning is extracted for the terms. As observable proxies for the underlying semantic components, these features are exploited using machine learning methods for text classification with respect to perspective. After initial experimentation with manually selected lexical resources, the method is generalized to require no manual selection or hand tuning of any kind. The robustness of this linguistically motivated method is demonstrated by successfully applying it to three distinct text domains under a number of different experimental conditions, obtaining the best classification accuracies yet reported for several sentiment classification tasks. A novel graph-based classifier combination method is introduced which further improves classification accuracy by integrating statistical classifiers with models of inter-document relationships.