

The Quantification of Sentiment

In the context of classic work in psychology and more recent work in computational as well as theoretical linguistics, I discuss an ongoing project to develop an automated system for sentiment analysis, i.e. a program that identifies whether a text (in general, a product review) is positively or negatively oriented. Topics include methods for building a sentiment dictionary and integrating the effects of context, including phrase-level negation, intensification, and modality, as well as higher-level concerns such as topicality, subjectivity, and discourse structure (e.g. rhetorical relations and genre stages). In addition to a theoretical discussion, I present data showing the performance of our system on commonly-used corpora, and contrast our approach to quantification with machine classifiers, focusing on both performance and the appropriateness of the features and weights used by such systems. Ultimately, my goal is to suggest that psychology and linguistics have an important role to play in these kinds of Natural Language Processing projects, and that focusing on purely computational methods is not a good strategy for long-term success.