Thesis: Sentiment, Opinion And Emotion Analysis.

Abstract: The sentiment, opinion, and emotion and their related concepts are central t0 human psychology and significantly impact our behaviors. With the rapid growth of social media, sentiment analysis, also called opinion mining, has become one of the most active research areas in natural language processing (NLP). Its applications are also widespread, from business services to political campaigns. This survey details different approaches for sentiment, opinion, and emotion analysis through reviewing the literature on their applications and discusses their advantages and disadvantages.

Specifically, we review approaches to construct sentiment lexicons that have influenced for lexicon and feature-based sentiment approaches. Sentiment values from the dictionary are assigned to words and to combine them in order to make the final prediction regarding the overall sentiment for the document. Other approaches include machine learning (ML) and Deep Learning (DL) methods applied in NLP. The sentiment analysis is formulated as a classification problem and methods such as Naïve Bayes, Logistic Regression, Maximum Entropy, and the Support Vector Machines classifiers. The Deep Neural Network approaches such as convolution neural networks (CNN) and recurrent neural networks (RNN) are shown for the sentiment analyses. The issues and various experimental results are compared.

The aspect-based sentiment analysis approach is to recognize several polarity values in a document. Its tasks include identification of multiple aspects (topics) in a document first, then apply the sentiment analysis. We review the unsupervised approaches such as the Latent Dirichlet Allocation (LDA) to identify aspects, and a few supervised M.L methods for aspect detection. The emotion detection task faces further challenges, e.g. lack of data with particular emotion labels. We present emotion theories and how they are used in major emotion detection approaches, Keyword-based, Emotion lexicons, Learning-based and hybrid methods.

Committee:

- Professor Soon Ae Chun, Mentor, The College Of Staten Island
- Professor Robert M. Haralick, The Graduate Center
- Professor Ronak Etemadpour, The City College Of New York