

SENTIMENT ANALYSIS OF TWEETS AND MOVIE REVIEWS

Noelle Harp, Debzani Deb. Dept of Computer Science; nharp115@rams.wssu.edu, debd@wssu.edu

It is now well recognized that the user-generated content (e.g., product reviews, forum discussions and blogs) contains valuable consumer opinions that can be exploited to gain consumer, brand, and market insights in today's business. As an important tool of E-commerce and business intelligence, sentiment analysis (or, opinion mining) emphasizes on classification of web comments into positive and negative categories. This poster details our experiments and results of performing sentiment analysis based on two different data sets such as movie reviews and tweets. We utilized two machine learning (ML) algorithms such as Naïve Bayes (NB) and K-Nearest-Neighbor (KNN) to classify the web contents containing positive or negative sentiments. We utilized WEKA datamining tool in order to design and perform our experiments. In our experiments, the training data set is utilized to train the ML algorithms and once trained, a test data set is utilized to predict the classification. Our experiments on twitter dataset showed that there is no significant variation in classification accuracy while varying ML algorithms and other parameters. On the contrary, performing different tests on the movie review data set showed some notable variations in classification accuracy. The best accuracy was achieved when removing stop words and extracting important attributes, using 80% as training data, and using the NB algorithm to classify the data. In all cases, NB outperformed KNN algorithm. Furthermore, focusing on important words generated better accuracy than considering all words in the web content.

Funding provided in part by NSF Award #1600864, WSSU RIP Award 2016-2017