

Week 4 Readings

C. Hidey

February 11, 2015

[SA2] Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. 2010:

O'Connor et al. focused on analyzing the relationship between sentiment in Twitter and public opinion polls. They extract Tweets containing specific words that embody concepts and then use a simple score to measure polarity. They use an average of the last k days to smooth the polarity. Using linear regression they measure the correlation between daily sentiment and public opinion polls. They also attempt to predict future polls using current poll information and sentiment analysis. One thing I liked about the paper was that the techniques should extend to other applications. The work could be expanded to look into other concepts (sports, pop culture) without much difference. One thing I didn't like was that they used a simple moving average to measure sentiment. Since the average is still a linear combination of features, they could use each day of sentiment as a separate feature and they would thus get assigned their own weight. They could also have experimented with feature transformation (i.e. using a higher order polynomial for more recent days).

[SA3] Theresa Wilson , Janyce Wiebe, and Paul Hoffmann (2005):

Wilson et al. looked at the classification of subjective phrases as having contextual polarity, which may be different from the polarity of other phrases in the same sentence. They used a generic polarity lexicon to determine the prior polarity out of context, both as a baseline and a feature. They adopted a two-step classification process, first classifying subjectivity phrases as either having polarity or not, then classifying for positive, negative, both, or neutral (still). They used an AdaBoost classifier with many lexical, structural, and syntactic features. One thing I liked about the paper was that it recognizes that sentiment is often dependent on context, as compared to sentiment analysis methods that consider only individual words as independently expressing sentiment. One thing I didn't like about the paper is the difficulty in reproducing the experiment (data set is not available and it is not clear how to extract features).