Predict 452 Section 55: Web and Network Analytics

Assignment 4: Option 1: Text and Sentiment Analysis, Jack Doig

Management Problem

Detecting emotions in text is a hugely valuable problem to solve. It has business applications as wide ranging as informing negotiation strategy, political tactics and a range of marketing decisions. This project seeks to analyse and categorize text from an emotional point of view, using political speeches and validating results against opinions of emotional angle of the speaker.

Data

Barack Obama was widely praised for inspiring and mobilising a non-political class into the voting booth in 2004. It is widely agreed that this was achieved on an emotional message of hope, inspiration and opportunity. In contrast, Donald Trump, it is sometimes assumed, seems to run a negative–emotion-based campaign, tapping into the fear of the constituency: fear of outsiders, fears of job loss, fears of being exploited or taken advantage of (by China, Wall Street, immigration or the political class). This projects seeks to determine if these emotional undertones can be detected and seeks to classify the underlying emotions. Speeches have been preferred, as they are a structured way in which the candidates try to encourage support and send their message to voters, and are therefore the most likely place to find a consistent emotional signal.

Research design and methods

There is no recognized existing package to model emotion, however there has been some work in the past we can learn from. Staiano and Gerini (2014) have developed a lexicon for emotion analysis from crowd-annotated news. They have developed a 37,000 word lexicon which labels each word with a fraction allocation to each of (AFRAID,AMUSED,ANGRY,ANNOYED,DONT_CARE,HAPPY,INSPIRED,SAD). The initial test I did while developing a method was on the Donald Trump Phrase “Our enemies are

getting stronger and stronger by the day and we as a country are getting weaker". In a simple bag-of-words, additive model, this phrase accrued 11.2% of its points for AFRAID, 11.8% for ANNOYED, 14.3% for SAD, and 22.6% for HAPPY. Similarly, the Obama phrase "in this country, justice can be won against the greatest of odds; hope can find its way back to the darkest of corners" scores 16.9% for DON'T_CARE and 14.3% for INSPIRED. This Lexicon and approach did not pass this initial validation, so I discarded this approach in favour of a richer predictor feature set.

By using a self-developed method based on learning from Obama and Trump speeches, I sought to categorize 2 emotions: Fear and Hope (with a view to expanding the model to other emotions if successful). The features in the model as a result of this process were:

a) **Vocabulary**: Existence of words unambiguously associated with that emotion (for example: Hope: trust, hope, opportunity. Fear: lie, horrible, nonsense, stupid, kill, destructive)
b) **Use of Questions**: Trump uses excessive questions to convey fear and uncertainty, a technique rarely used by Obama – questions can be measured in the text
c) **Vision versus dystopia**: Obama appeals to an uplifting vision of the future, but not focussing on the negatives of other candidates or the external "threats" of immigration and China, so the topics discussed are important predictors also

Results/Findings, next steps

Using a combination of context-specific topics, vocabulary scoring and language structure predictors allowed me to distinguish between the emotions of hope and fear better, and measure them in the speeches of Clinton, Palin and Cruz. I still think this model is still in a very primitive form, however this study discovers some of the predictors that need to be considered in order to understand the emotion in a political speech.