Text Sentiment Analysis System Report

Solomon Foster & Lazar Zamurovic

**Research Question:** Using a set of words with sentiment scores, can we create a resonably accurate system to analyze the positivity or negativity of peoples’ texts? Once we have attained a primitive model, what are some techniques we can use to improve the sophistication and accuracy of our system?

**Background:** We wanted a way to analyze our text messages and look for trends in texting patterns: people who are more or less positive, when we are more or less positive, etc. Essentially we created a tool for people interested in analyzing their texts; it could be used by teenagers or people doing exploratory data analysis, or really by anyone who has a corpus. We found a scored set of 10,000 words from a study done by an NLP group at Cornell. The scored set averaged peoples’ reponses to words on a scale between 1 and 9. Words like love score around an 8.2, while blood and murder score in the 1s and low 2s.

**Methodology:** Our system goes sentence-by-sentence and word-by-word using a dictionary with our scored words to look up and assign scores to our candidate words. When a word is not in our tagset, we generate synonyms for the word using wordnet and check whether any of those words is in our score set. If it is, we assign the word that score. We also have a very slow method (O(10,000n)) to compute edit distances of a candidate word to words in our score set. This method is effective on misspellings like “suuuuper” but ineffective on many other types of misspellings. Failing all that, we assign a word a <unk> tag. We also have a function to map inputted scores to qualitative judgements of positivity. This function uses Solomon’s texts as training data and assesses how many standard deviations from the mean positivity of his texts a given score is. It then outputs (Somewhat|Very|Extremely) (Positive|Negative) depending on the result.

**Results:** Our system is somewhat difficult to evaluate analytically. One thing we can say is that the word-by-word recall of the system is about .94, leaving about 6% of words unscored. The main mathod of evaluation we have used is the eyeball test; especially with qualititative output, it’s fairly easy to compare one’s expectation with the system’s judgement. This makes sense because humans generally should be the gold standard of judging positivity.

**Conclusion:** Our system is a solid step in the direction of developing a sophisticated sentiment analysis tool. It performs substantially better than our first attempt. Like many natural language tools, however, it struggles with subtlety and other things like sarcasm and irony. It further is better at correctly identifying positive sentences than correctly identifying negative sentences. We think that this is caused in part by the fact that positive utterances are usually composed of at least a few very positive words (like happy, awesome, etc.), whereas negative utterances are often an amalgamation of neutral words and/or negation words (e.g. “I don’t like you very much”). Our system could be broadly used by people looking to identify positivity in any corpus, but probably in developing it with text messages we biased the system towards texts (especially our qualitative analysis, for instance, which uses a corpus of text messages as its baseline).

Our system could be improved by using a chunker to bring together negative phrases and evaluating several words as one negative block. We also could involve ngram models and/or machine learning to identify words we have tagged as unknown. This would result in some, but probably not massive, improvement.