Week 11 Readings
C. Hidey

April 2, 2015

[SP1] Bramsen, Philip, Martha Escobar-Molana, Ami Patel, and Rafael Alonso. 2011.

Bramsen et al. attempt to identify variations in language that allow for the identification of power relationships. They term these variations UpSpeak, DownSpeak, and PeerSpeak. The authors examined mixed n-grams for use in identifying these relationships, where the n-grams consist of a combination of POS and word. They use a weight for the features that corresponds to how often that bigram appears relative to the total number of bigrams. For example, one mixed bi-gram may be *please,VB* and for a sentence that has 9 total bigrams of this type, it would have a weight of $1/9$. They determined what n-grams to use based on relative frequency and absolute frequency both being greater than a threshold. Finally, they binned n-grams into sets according to their relative frequency and used information gain to divide these n-grams into classes. They created a classifier trained on the Enron E-mail corpus and experimented with different subsets of features: n-grams, mixed n-grams, binned n-grams, and manually selected n-grams. I liked that the paper was straightforward and well-written. I didn't like that they used a simple method to bin n-grams instead of considering other clustering techniques.

[SP2] Danescu-Niculescu-Mizil, Cristian, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013.

Danescu-Niculescu-Mizil et al. examined politeness and social power using an annotated corpus of Wikipedia edits and Stack Exchange posts. They used English-speaking annotators to rate the politeness of text using a continuous value between very impolite and very polite and normalized the scores so that the mean was 0 and standard deviation was 0.7 They analyzed linguistically-motivated politeness strategies and extracted them using regular expressions on a dependency parse. They created an SVM classifier and found that including the linguistically-motivated features improved over BOW and is close to human performance. Finally, they used their results to determine that politeness decreases as a user gains power. I liked that they showed a clear relationship between politeness and social power and were able to determine this relationship using linguistically motivated features, I didn't like that the idea of politeness as a continuous feature is subjective.