Movie Review Sentiment Visualization

by Dan Prendergast

Team: Dan Prendergast

Objectives & Motivating Problem: Interpreting subtle meaning and sentiment in conversation is one of the most complex cognitive tasks in which humans engage. Having a real time sentiment meter could help people assess how a conversation is going. Techniques in natural language processing (NLP) have been applied to sentiment analysis of movie reviews, but top performing systems are deep neural nets. These would require significant processing power and may not execute in real time.

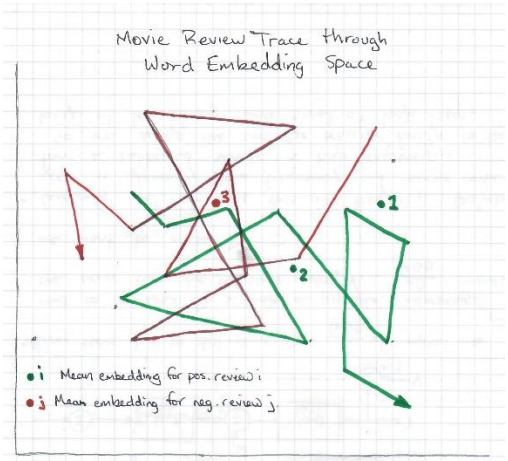
In this project, I would like to visualize multiple, simple-to-measure metrics of the movie reviews to see if any features of the reviews will allow simple and fast discrimination of review sentiment. Of course, any results found on a movie review dataset do not necessarily translate to general conversation; however, it would represent an initial proof of concept on readily available movie review datasets. Further work would involve applying the same processing and visualization to a dataset of general conversation with sentiment annotations.

Rough Plan of Work and Milestones: The work can be broken into 3 phases – data collection and processing, coding the visualizations, and project writeup. Movie reviews with annotated sentiments are available from Kaggle.com, and pretrained word embeddings are available from multiple sources depending on the algorithm used (e.g., GloVe, word2vec). The milestones are as follows:

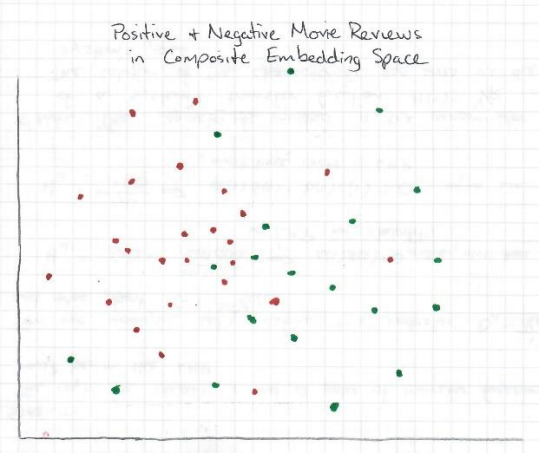
- Data Collection/Processing: 21 Apr
- Coding the Visualizations: 28 Apr
- Project Writeup: 29 Apr

Deliverables: The project deliverable will be a Jupyter Notebook that will run the processing and display the visualizations. The notebook will allow the user to select a movie, then display visualizations of the following movie review features:

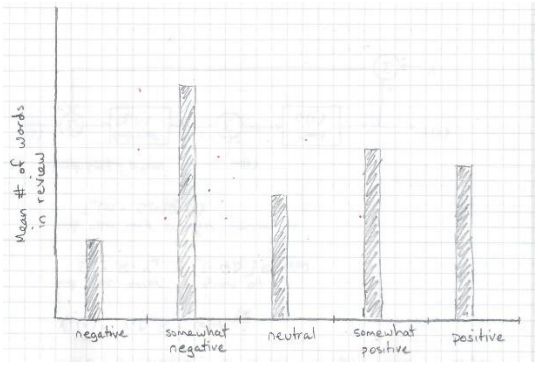
1. Traces of review paths through word embedding space (with overlay of mean word embedding values for each review)



2. Composite embeddings for each review. Composite embeddings calculated by element-wise multiplication of each word embedding in the review.



3. Mean number of words for positive and negative sentiment reviews



4. Distribution of words by frequency of use