

Sentiment Analysis on Film Reviews

Proposal and Learning Goals

Pratool Gadtaula and Jay Woo

February 5, 2016

Due to our express interest in natural language processing and sentiment analysis, we decided to pursue the Kaggle competition for classifying the positivity of movie reviews on Rotten Tomatoes.

1 Project Goals and Alignment

First, we will apply what we already learned from class. The data may require some sort of cleaning (using a bag of words) before being processed, and we hope to try a logistic regression model for our first iteration. We expect that this model will not be very accurate, so we would like to try to come up with our own NLP model or research and implement someone else's. There is an NLP course on Coursera that we could learn bits and pieces of along with Python's NLTK library which has already implemented common models.

2 Pratool's Learning Goals

I was interested in being able to extract meaningful information from unstructured data, which is the essence of natural language processing. I also want to improve my skills in explaining very technical work, which is part of the project requirements. Through this written medium, I hope to be able to use figures appropriately to explain significant findings and learning. I'm also very aware that it is easy to be sucked into a project where only research is performed and there is a lack of implementation. I hope that by taking proactive steps to learn concepts on our own time, we can get together to fill gaps in each other's knowledge and implement those concepts in code.

3 Jay's Learning Goals

My goals for this project are to learn some of the basics of natural language processing and to at least implement a simple model that is 50% accurate or better (the highest Kaggle scores are around 75%). I want to avoid doing a ton of research in topics that are completely out-of-scope and not implement anything practical, as a result.

Sources

<http://www.vikparuchuri.com/blog/natural-language-processing-tutorial/>
<https://www.coursera.org/course/nlp>
<https://www.coursera.org/course/nlangp>
<http://nlp.stanford.edu/software/tokenizer.shtml>
<http://www.nltk.org/>
<https://www.kaggle.com/c/sentiment-analysis-on-movie-reviews/>