Melissa Iori
NLP Final Project Pitch

“Wikipedia Promotional Content Classifier”

Motivation:
The Wikipedia community largely aims to provide unbiased, quality content in its various articles. However, despite the best attempts of contributors, many articles face problems with tone and style, as well as veracity of content. In particular, some articles have been found to be heavily biased in favor of certain people, organizations, or products. Wikipedia calls these “articles with promotional tone”. These are articles that have to be manually identified and flagged by a Wikipedia community member for further revision or deletion.

Back-end:
My project will be to implement an n-gram document classifier which can classify a Wikipedia article as having “promotional tone” or not.
This has already been researched by the following team in the following paper, and they met with success in terms of finding an accurate classifier by using a combination of content, structural, and Wikipedia network (links to other articles) features:
http://www.cs.utexas.edu/~ml/papers/bhosale.emnlp13.pdf
Rather than exactly replicate their work, my goals are:

- To see if accurate classification is still possible with less or different features
- Compare some of NLTK’s classifiers for accuracy
- Compare unigram, bigram, trigram and quadgram for accuracy
- Train on a new set of the latest articles (since this study was done in 2013). It’s possible that editors who create promotional content could have created new tactics since then to get around these classifiers, since there are ways to write promotional content while still adhering to a generally encyclopedic tone.

The back-end will be written in Python. The webpage can be served using Django, so that the classifier can run directly from Python and display the results back to the user. The model will basically only be trained once, and re-loaded every time a user makes a request. If it had to be trained on every request, that would take too much time and the request would time out. Also, the model being used on the webpage will be the most accurate one out of my tests.

Front-end:
The webpage will have a brief description of the problem that the classifier addresses, and a text box in which the user can input a Wikipedia article URL and click a button. The page then shows popular n-grams identified, along with the decision (label) whether the article contains promotional content or not. If time, will also include a version of the article’s text with key n-grams highlighted for greater comprehension.
The front-end will be in HTML, CSS and JavaScript.

Data Sources (subsets of):

These are designated articles of high quality, which contain minimal promotional content, if any. They are deemed by the community to be written in the proper style for Wikipedia, and wouldn't be flagged.
Negatives:
https://en.wikipedia.org/wiki/Wikipedia:Featured_articles
https://en.wikipedia.org/wiki/Wikipedia:Good_articles
These are the articles that have been hand-picked and identified as promotional in tone, and should be flagged by the classifier.
Positives:
https://en.wikipedia.org/wiki/Category:Articles_with_a_promotional_tone
https://en.wikipedia.org/wiki/Category:Articles_with_a_promotional_tone_from_October_2017