Sentiment Analysis of Movies Review (Midterm Presentation)

CSC 664/864: Multimedia Systems
By
Phyo Htut
Sanjay Mirani
Jose Pascual

So Far...

- Started looking for classifiers along with movie reviews to train and test our model.
- Started working with library called NLTK in python, which provides all the tools and methods for natural language processing.
- Also, it contains 2000 movie reviews - positive and negative, which we have used for now to train our model
- Also have collected dataset of movies which we still need to look into.

Method Design

- Started off playing with the NLTK library which contains an in-built movie review dataset.
- Divided the dataset into positive and negative data list.
- Applied the data preprocessing techniques like tokenizing, removing stop-words and converting tokens into lowercase.
- Next, dividing this dataset into training and test set.
- Passing on the training dataset to the Naive Bayes Classifier which comes in the NLTK library.

contd....

- Next, we test our classifier on the test dataset.
- And, as a result, our accuracy hit around 70% to 75%.
- Here is the link to the github repo -
  https://github.com/sanjay417/Sentiment-Analysis

Next

- Focusing on improving our ML accuracy
  - Have used bag-of-words feature for now but still have to take a look into N-grams feature.
  - Still have to compare naive bayes with Support Vector Machine (SVM).
- Building up backend
  - Flask
- Benchmark against Python sentiment library
  - Textblob
  - Vader
- Building up frontend
  - Vue Framework

Thanks!