Sentiment Analysis of Movies Review (Final Presentation)

CSC 664/864: Multimedia Systems

By
Phyo Htut
Sanjay Mirani
Jose Pascual

Previously

- Why not more advanced term analysis methods like TF-IDF or LSA?
  - Applied TF-IDF
- Which classifiers are being tested?
  - Naive Bayes Classifier (sklearn)
- Is the current prediction rate on test data or cross-validated training?
  - Test data: ~85%
  - Cross-validated data: ~90%
- How do you avoid overfitting?
  - K-fold with cross-validation
- How is the result correlated with IMDB scores?
  - More on demo...

So Far...

- Model is up and running.
- Integrated Omdb api to fetch the imdb id for a movie by passing on the movie title.
- Also, used api dojo from Rapid api, which gets us the top movie reviews from the critics and the users based on the imdb id fetched from the previous api.
- This reviews are then passed on to the model and sentiment for each review is being displayed.
- Only thing user have to do is enter a movie name.

Changes...

- Based on comments given by professor, we started our work on exploring techniques to avoid overfitting and got the best one - instead of splitting the dataset into train and set, we implemented k-fold cross validation technique. And so, the accuracy is now being calculated on cross validated training data unlike previous accuracy which being calculated on test data.
- Also, we have implemented unigram along with bigram feature which do provides a little boost to our accuracy.

Tech Stack

- Building up backend
  - Flask (Python)
- 3rd Party API endpoints
  - omdbAPI
  - apiDojo.net
- Building up frontend
  - React
- Techniques and data
  - Naive Bayes Classifier
  - Data: https://ai.stanford.edu/~amaas/data/sentiment/

Demo

Thanks!