MTP Progress

Nikhilkumar Jadhav

Topic Modeling

Process

- Given a set of documents
  - set # of topics = N
  - Remove stopwords
  - Gibbs sampling
  - topic-document distribution
  - word-topic distribution

Topic Modeling for Sentiment Analysis

Process

- Given a set of documents
  - set # of topics = 2 (pos and neg)
  - Remove stopwords
  - Remove objective words
  - Gibbs Sampling
  - topic-document distribution
  - word-topic distribution
- Infer topics for new documents

Experiments

- Implementation of LDA in Mallet
- Amazon reviews Dataset
  - SNAP
    - Standard Network Analysis Platform
  - 8 million reviews
  - subsets of this Dataset have been used
- SentiWordNet was used to extract subjective words

Word lists

Positive

great good love bad plot fun
funny worth enjoy pretty hard
loved recommend classic favorite
enjoyed fans nice entertaining
amazing wrong happy humor
awesome fantastic disappointed
fast cool dead die top scary loves
poor enjoyable worst boring
hilarious zombie glad decent
excellent interesting super hate
waste terrible stupid horrible

Negative

work wonderful excellent perfect
beautiful interesting dark lost works
death fine brilliant evil won top easy
strong important greatest problem
modern superb lead murder hero
killer crime kill clear romantic doubt
incredible miss powerful mystery
problems supporting killed
believable realistic lack outstanding
difficult masterpiece perfectly
strange talent fall worked

Results

Training Size (N/2 positive, N/2 negative)	Testing	Accuracy
20000	4000	83.65
60000	12000	65.07
200000	40000	43.71

Topic Modeling for Sentiment Analysis with Sentiment Prior

Process

- Given a set of documents
  - set # of topics = 2 (pos and neg)
  - Remove stopwords
  - Remove objective words
  - Gibbs Sampling
    - Using SentiWordNet, assign topics to words initially
  - topic-document distribution
  - word-topic distribution
- Infer topics for new documents

Word lists

Positive

great good love work fun excellent worth classic enjoy pretty loved recommend interesting wonderful favorite enjoyed perfect beautiful top amazing nice happy humor entertaining works fine fantastic fast won brilliant loves strong enjoyable important greatest modern hero superb lead hilarious glad clear romantic incredible super awesome powerful supporting recommended

Negative

bad plot funny hard fans wrong dead dark lost death problem disappointed cool evil die easy scary poor worst boring murder zombie decent killer crime kill slow doubt miss hate mystery sad problems waste killed free hell terrible lack stupid difficult horrible strange fall worse fiction missed twist falls

Results

Training Size (N/2 positive, N/2 negative)	Testing	Without Prior	With Prior
20000	4000	83.65	88.97
60000	12000	65.07	79.73
200000	40000	43.71	81.69

Conclusion

- Purely Unpsupervised topic modeling is not suitable for Sentiment Analysis
- Adding sentiment prior increases accuracy of the inference

Future Work

- Implement JST
  - *Lin, Chenghua, and Yulan He. "Joint sentiment/topic model for sentiment analysis." Proceedings of the 18th ACM conference on Information and knowledge management. ACM, 2009.*