



Sentiment Analysis: Memorability

RAIFY HIDALGO

Summary

Using movie scripts, can a line be identified as memorable?

Are there frequent parts of speech tags that are used in memorable lines?

Can the memorability of a line be predicted?

- ▶ Memorable: Worth remembering or easily remembered

Process

Data Mining

- ← Use 40 movie scripts from online database
- ← Break up scripts into a corpus of sentences
- ← Scrape yahoo search results to get hits per sentence
- ← Two main corpus for training:
 - ← Top Yahoo hits w/o manual annotation
 - ← Top Yahoo hits w/manual annotation

Feature Analysis

- ← Part of speech frequencies for each corpus
 - ← What are the most common breakdowns (tuples or more) for each type of sentence?
 - ← What tags are most frequently used for each type of sentence?

Prediction

- ← Approach:
 - ← split up sentences into data set of words per memorability type
 - ← Eliminate low information
 - ← Use Information gain/Kullback-Leibler divergence
- ← Return predictions based off of corpus with and without annotations

Evaluation/Difficulties

► Difficulties:

- Defining memorability
- Consistency across annotators
- Getting proper evaluation
- Search engine hits
 - Which one is best?
 - Inherent problem of popularity bias
 - Getting enough data – throttled from too much scraping

► Evaluation

- Use of recall and f-score for annotator agreement
- Use of same annotators and yahoo hits for new scripts
- Compare predictions against manual annotations & engine hits of new scripts