



Supervised Topic Models

Advanced Machine Learning for NLP

Jordan Boyd-Graber

MULTILINGUAL APPROACH

Problem

- We have documents in multiple languages
- They are annotated with the same **continuous response**
 - Rating of a product
 - Movie review
 - Percentage of people who retweet a tweet
 - Percentage of people consider a comment “extreme”
- Can learning a model in multiple languages at once help?

Problem

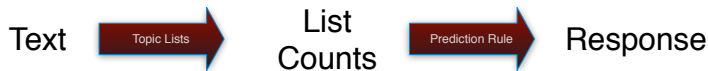
- We have documents in multiple languages
- They are annotated with the same **continuous response**
 - Rating of a product
 - Movie review
 - Percentage of people who retweet a tweet
 - Percentage of people consider a comment “extreme”
- Can learning a model in multiple languages at once help?

For languages where you don't have many resources, **yes!**

Conceptual Approach

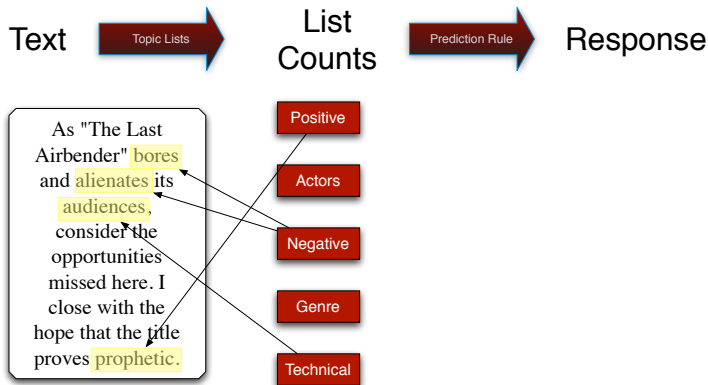


Conceptual Approach

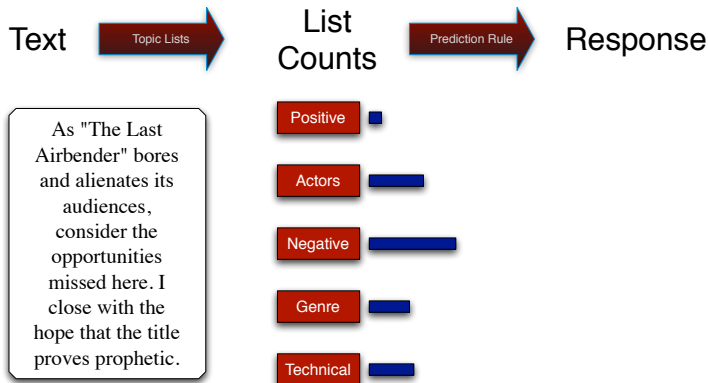


As "The Last
Airbender" bores
and alienates its
audiences,
consider the
opportunities
missed here. I
close with the
hope that the title
proves prophetic.

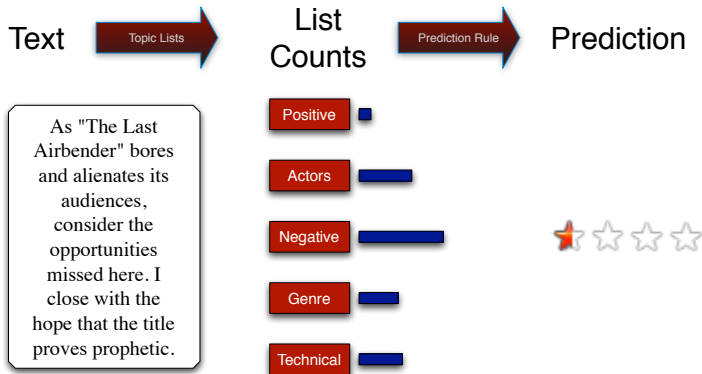
Conceptual Approach



Conceptual Approach

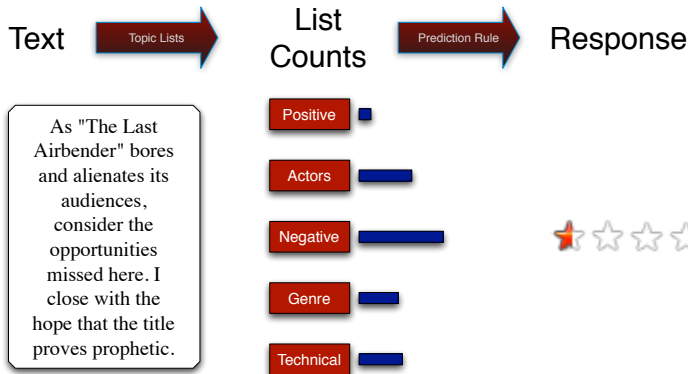


Conceptual Approach



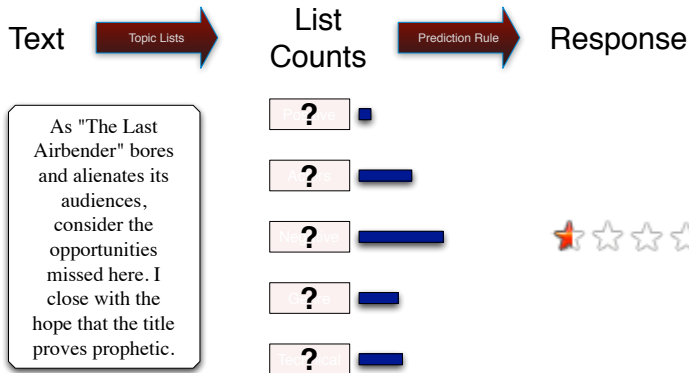
Similar to social science methodology LIWC ?

Conceptual Approach



- **Assumption:** We can create list counts from documents in any language
- **Observation:** Once we have list counts, underlying language doesn't matter

Conceptual Approach



What if we don't know the lists?

Putting Pieces Together

- How do we learn the word lists?
- How do ensure that the word lists reflect sentiment?
- How do make the word lists make sense across languages?

Putting Pieces Together

- How do we learn the word lists?
- How do ensure that the word lists reflect sentiment?
- How do make the word lists make sense across languages?

Putting Pieces Together

- How do we learn the word lists?
 - Topic Models
- How do ensure that the word lists reflect sentiment?
- How do make the word lists make sense across languages?

Putting Pieces Together

- How do we learn the word lists?
 - Topic Models
- How do ensure that the word lists reflect sentiment?
 - Supervised Topic Models
- How do make the word lists make sense across languages?

Putting Pieces Together

- How do we learn the word lists?
 - Topic Models
- How do ensure that the word lists reflect sentiment?
 - Supervised Topic Models
- How do make the word lists make sense across languages?
 - Semantic Resources

Why do this?

- Topic models embed documents in low dimensional space
- These spaces are often useful for prediction
- But not designed for it!
- Can we use different objective functions to optimize embedding
- Understanding interplay between

Overview of today

- Supervised topic models
- Using multiple languages
- Hierarchical non-parametric models