Supervised Topic Models

Advanced Machine Learning for NLP

Jordan Boyd-Graber

MULTILINGUAL APPROACH

Representations

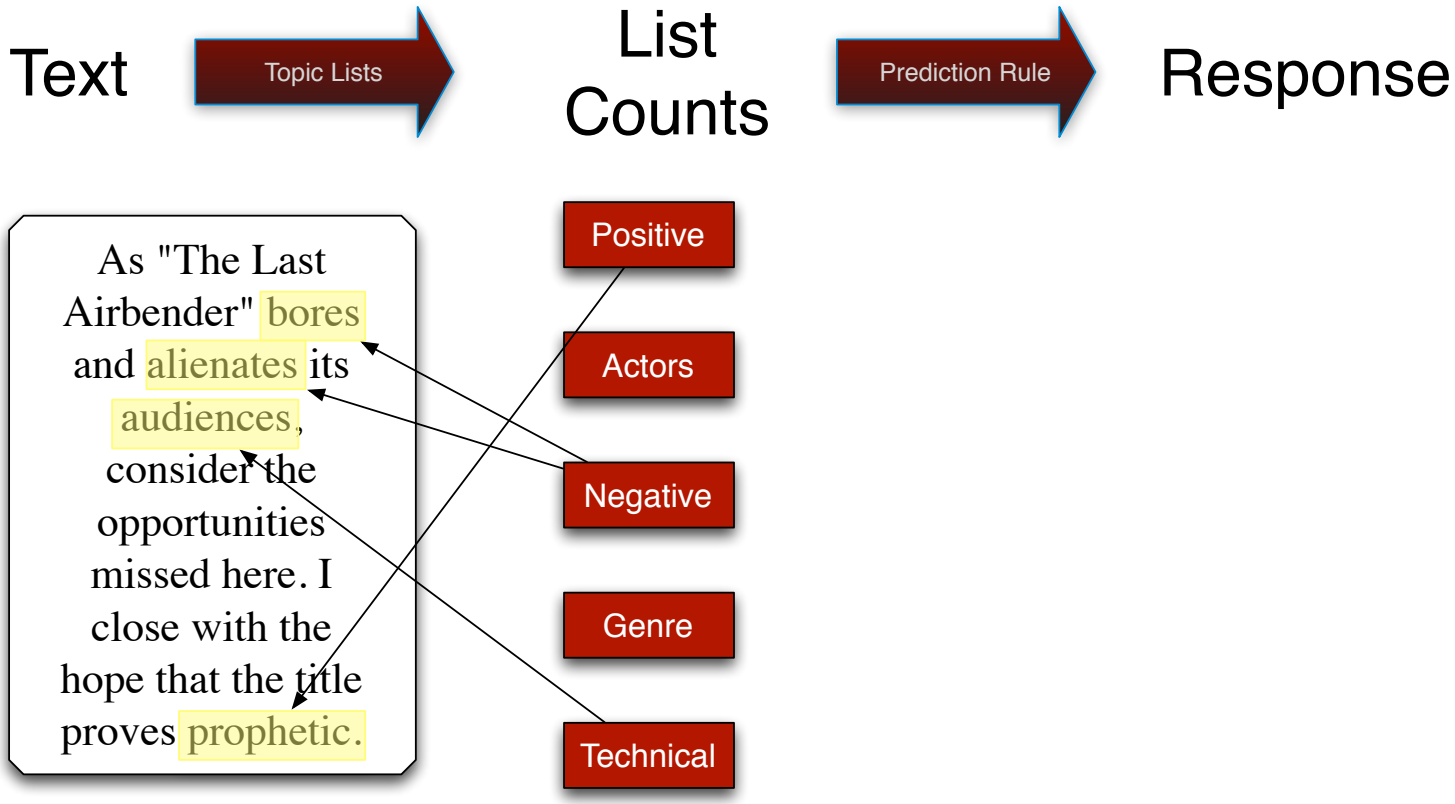
- Last couple of weeks: probabilistic representations
- Today: combining with supervised **response**
  - Rating of a product
  - Movie review
  - Vote on bill
  - Percentage of people who retweet a tweet
  - Percentage of people consider a comment “extreme”
- More advanced represenatations:
  - Multiple langauges
  - Hierarchy

Conceptual Approach

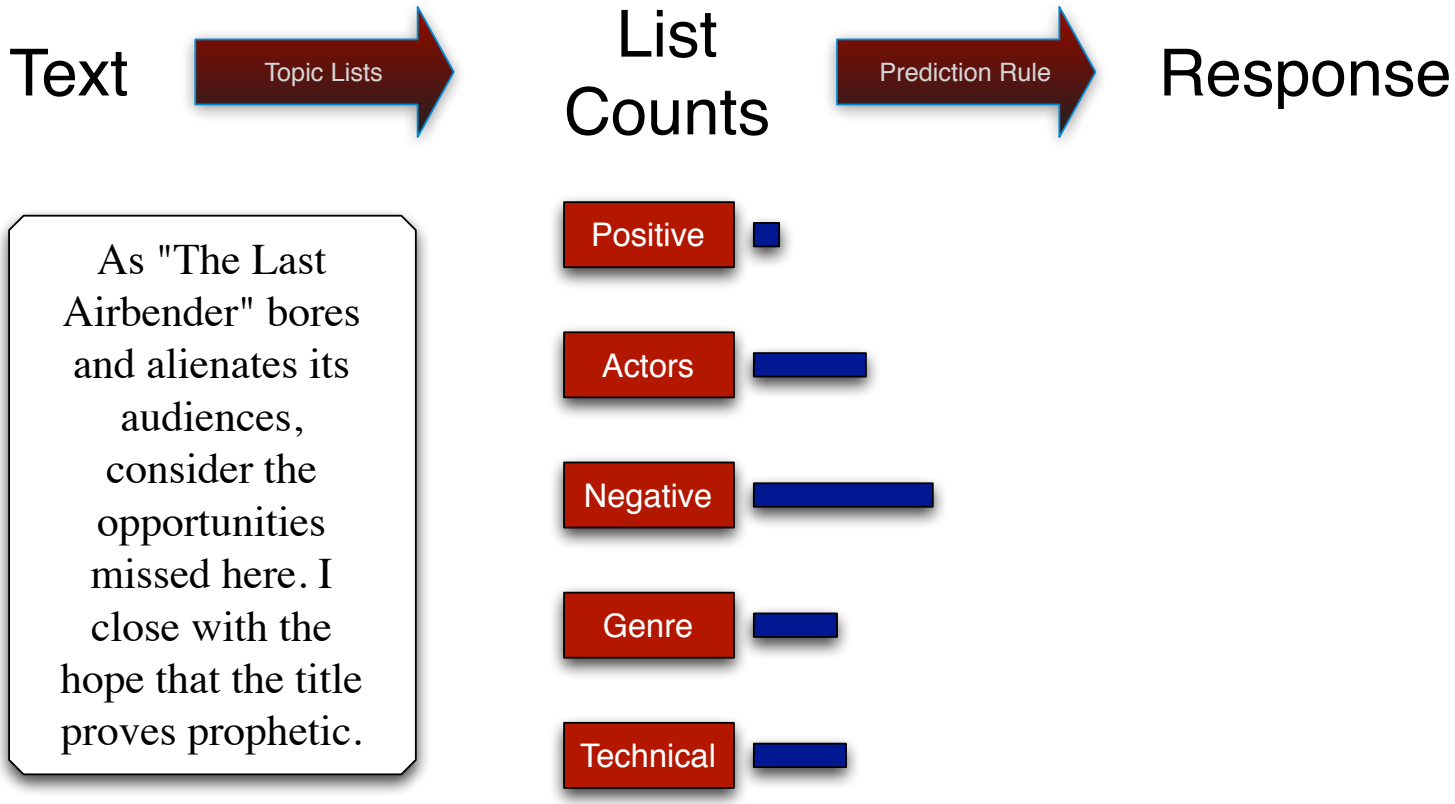
Conceptual Approach

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

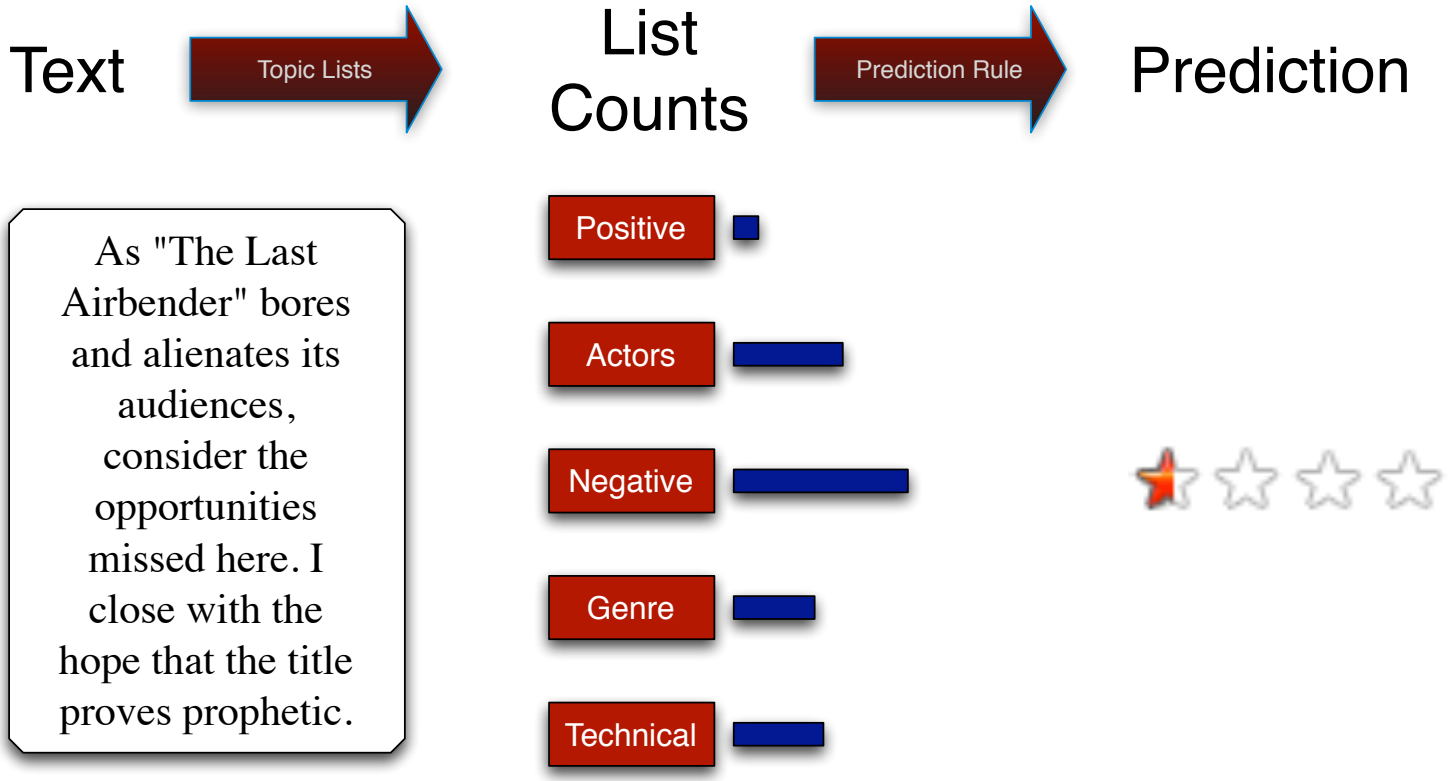
Conceptual Approach



Conceptual Approach

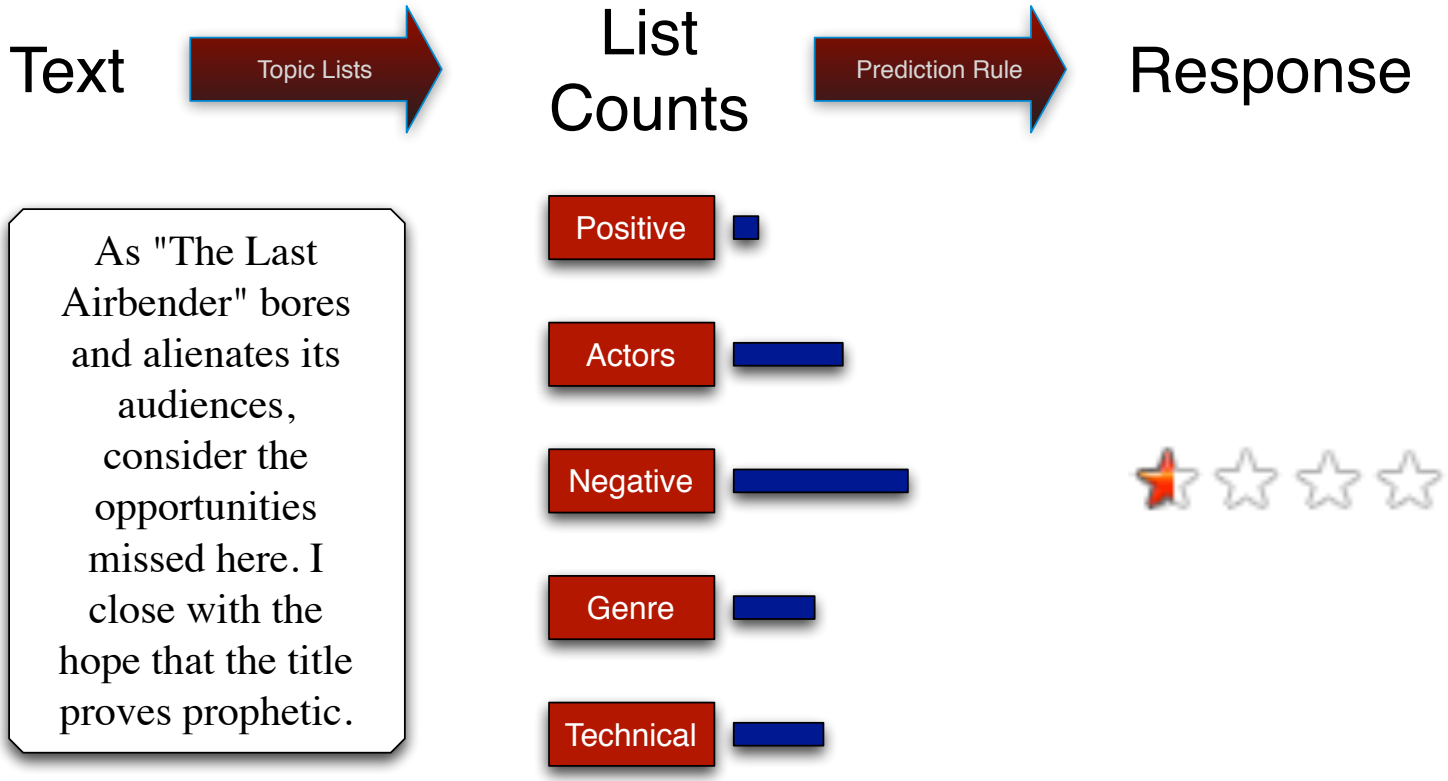


Conceptual Approach



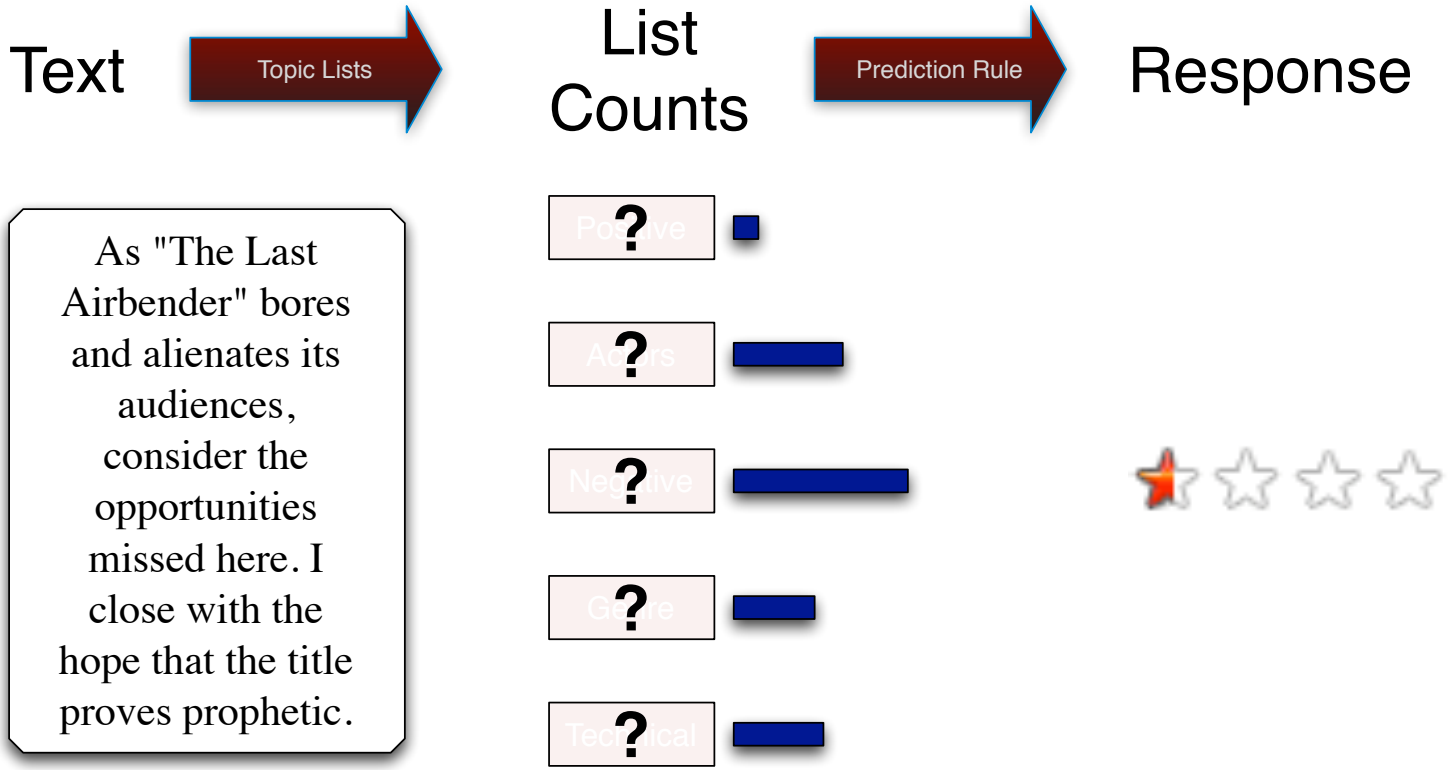
Similar to social science methodology LIWC

Conceptual Approach



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

Conceptual Approach



What if we don't know the representation?

Putting Pieces Together

- How do we learn the representation?
- 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 representation?
- 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 representation?
  - 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 representation?
  - 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 representation?
  - 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