What's the Word?

Topic model discussion on movie reviews

Introduction

Motivation:

- Disney wants to take a deeper look at what audiences think of the new movie Shang-Chi.
- I've been tasked to see what audiences are talking about.

Objective/Goals:

- Generate topic model from looking at audience reviews
- Sentiment analysis of reviews to see what topics are being discussed in positive and negative reviews

Methodology

01 **Data**
- 2,000+ audience reviews from Google reviews

02 **Metrics**
- Document polarity
- Topic Correlation

04 **Tools & Models**
- BeautifulSoup & Selenium
- Preprocessing tools(NLTK, Spacy,TextBlob)
- NMF & Corex

Results

NMF

- Strong topics not formulated, but some words seemed worth exploring

Corex

- Use of anchor words helped solidify topics

Topics

- Action sequences, Plot, Cast, Asian Culture

Polarity and Topic Correlation

Polarity
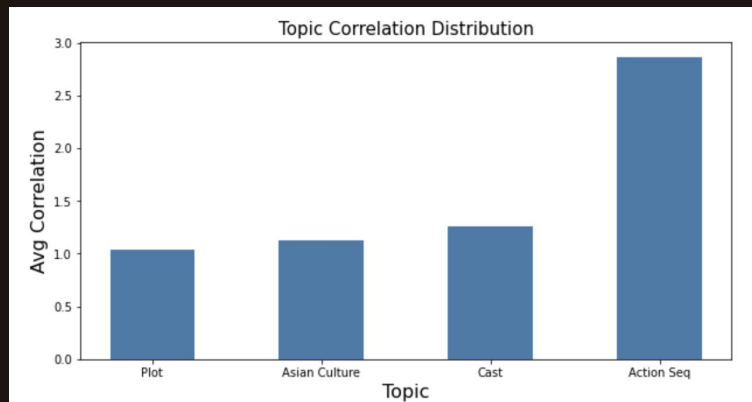- Obtained polarity for each document to determine if review was positive or negative

Polarity = 1:
   "A perfect Marvel movie as always."
Polarity = -1:
   "This is the most boring movie I've ever watched."
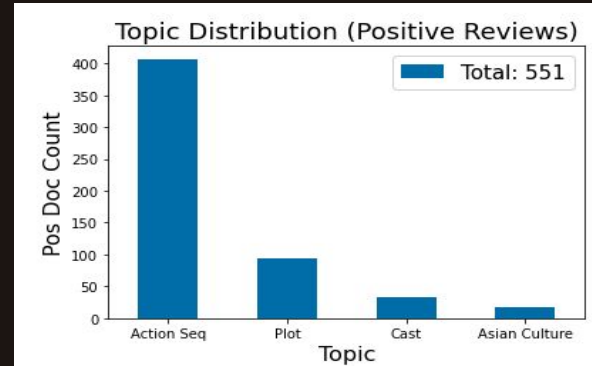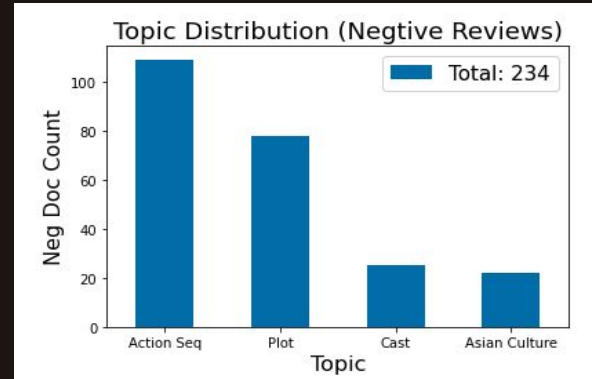
Topic Correlation
- TC Dist → Topic Correlation

Results

- Overall, both positive and negative reviews centered around action sequences and plot

Topic Distribution (Negtive Reviews) — Total: 234

Topic Distribution (Positive Reviews) — Total: 551

Examples

STRENGTHS

- Asian Culture (Positive)
  - *"It was visually awesome and it enlightens a lot about Taiwanese and Chinese culture."*

WEAKNESSES

- Plot (Negative)
  - "Outrageously bad it is difficult to give it one star."

- Action Sequence (Positive)
  - "Brilliant movie IMax is a must."

Conclusion

- Audiences generally only talk about a few topics.
- Some reviews don't really have "topics" at all

Future Work

- Take a look at critic reviews instead of audience reviews. They may give more informative topics

- Take a different approach when assigning topics (i.e allowing a document to be more than one topic, or discard if not related to any topics)

- Take a closer look at neutral reviews as they might be a combination of negative and positive

Thank you!