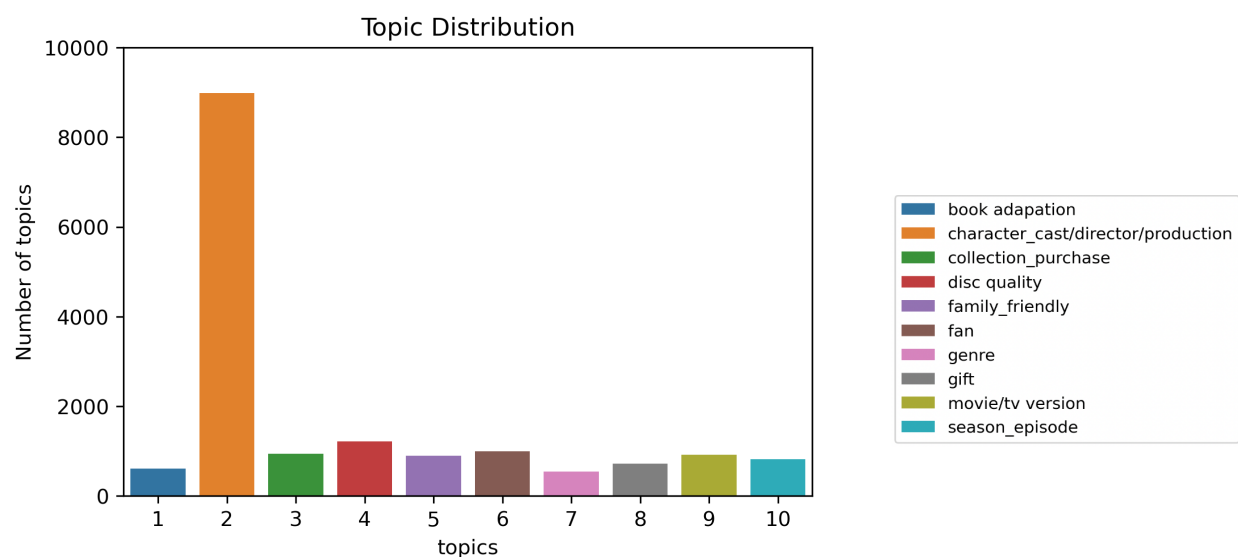


NLP Project Write Up

Amazon Movie&TV Reviews Sentiment Analysis and Topic Modeling

Abstract

I performed natural language processing on the Amazon movie&TV reviews data. I was able to identify the overall customer sentiment on the movies&TVs and find out the most outstanding topics that customers talk about as shown below.



Data and Design

I accessed the Amazon metadata through <https://nijianmo.github.io/amazon/index.html> (Jianmo Ni, Jiacheng Li, Julian McAuley *Empirical Methods in Natural Language Processing (EMNLP)*, 2019) under the 'Movies and TV' session and later focused on the smaller cleaned dataset. The final cleaned data I used for analysis consists of 16905 rows of customer IDs, reviews, and ratings data.

Work Flow and Algorithms

EDA and Data Visualization

- Matplotlib, WordCloud

Text Preprocessing

- Used including word tokenization, lemmatization, pos tagging, stop words removal
- Filtered words down to nouns, adjectives, verbs and adverbs

Sentiment Analysis:

- VADER, TextBlob used and settled with VADER

Topic Modeling:

- With only nouns
- LDA, LSA, NMF, CorEx used and settled with NMF

Communication:

<https://github.com/zouma-spec/NLP/blob/main/Presentation.pdf>