An Analysis of Nothing

Chandler Ault, Yamina Katariya, Yash Manne, Aditi Shrivastava

Background

What is Seinfeld:

- American sitcom that features four friends, Jerry Seinfeld, George Costanza, Elaine Benes, and Cosmo Kramer played by real-life comedians

Problem:

- Difficult for Seinfeld fans to search and discover certain episode and series characteristics when their only resource is manual web-searching

Solution:

- Interactive web page that covers the extensive capabilities of Wikipedia, IMDb, and Netflix all in one!

Data Used

Source	Collection	Description	Limitations
Seinfeld Chronicles (Kaggle)	Downloaded	Complete scripts for all 173 episodes	Doesn't adhere to Netflix/IMDb indexing, messy fields
IMDb Datasets	Downloaded	Episode production information and IMDb scores	Scores may be biased toward either ends.
IMDb Episode Pages	Scraped	Description, summaries, keywords	Time consuming. Truth is assumed

Use Cases

Episode Recommendation

1. **Website:**

   Prompts user to enter number of desired recs and favorite episode titles

2. **User:**

   Enters prompts

3. **Website:**

   Returns n episodes with highest similarity

Episode Querying

1. **User:**

   Enters general events/occurrences into search bar and selects desired fiters

2. **Website:**

   Displays a list of episodes that satisfy the query

3. **User:**

   Clicks on an episode to view episode sentiment analytics

General Analysis

1. **User:**

   Selects seasons of interest

2. **Website:**

   Displays general analytics like number of lines spoken by main four and episode rating, by season

3. **User:**

   Interacts with graphs (ex. zoom)

Tool Design

Data Manager

Data Cleaning + Pre-processing

- Cleaned all data to account for merging between IMDb and Kaggle data sources
- Precompute emotional sentiment for each dialogue
- Precompute feature embeddings from BERT on all lines of dialogue.

⇩

Data Manager Utility Tools

- Loads cleaned scripts & metadata and hands it other functions.

Episode Recommender

Episode Recommender Utility Tools

- Generate vector encodings for all episodes based on dialogue, key words, summaries, ratings, sentiment, and # of lines per character
  - Uses pretrained BERT embeddings for text
- Computes pairwise cosine similarities between input episode(s) and all other episodes, and returns top n episodes with the highest average cosine similarity score

⇩

Episode Query Web Page

- Display favorite bar for user input
- Display output table with recommended episodes and metadata: Episode ID, Rating, Title, Writers, & Directors
- Users click on any episode:
  - Can view it's IMDb episode description

Episode Query

Episode Query Utility Tools

- Generate vector encodings for all dialogue + input query, find closest matches via cosine similarity
- Create clickable AgGrid data frame from output
- Filter response by sidebar option (Season, Rating, Speaking Characters)
- Format data for interactive sentiment analysis visualizations

⇩

Episode Query Web Page

- Display search bar for user input, display sidebars for filtering
- Users click on any episode:
  - View it's IMDb episode description
  - Bar chart and sunburst plot depicting average episode sentiments, and average characters' sentiments

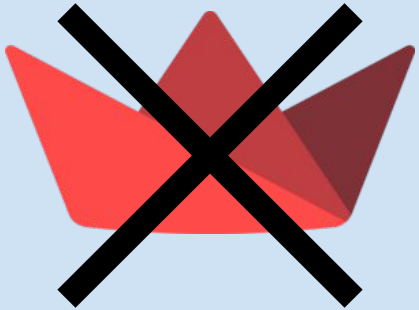
Lessons Learned and Future Work

Lessons Learned:

- Simplicity is not always the best option
- Planning and finalizing structure early on saves time

Future Work:

- Improve recommender algorithm and add additional filter fields
- Add specific dialog search to Episode Query
- Improve sentiment analysis



Website Demo: Local and Hosted