Grey's Anatomy Episode Recommender

By Melanie Ackerman

Abstract:

The medical drama, Grey's Anatomy, is on its 18th season and will reach its 400th episode by its conclusion in May of this year. The show has maintained a wide fanbase and remains a top show on Netflix, with viewers watching older seasons as well as newer. As a fan myself, I assume a sizable share of fans rewatch favorite episodes, or even the entire series more than once. Before streaming services took over, sometimes shows would release DVD sets of select episodes (e.g. holiday episodes). In this spirit, I created a recommender based on episode topic to help fans find groups of similar episodes. I use natural language processing, topic modeling, and cosine similarity to produce this episode recommender and deploy it as a web app on Streamlit.

Design:

This project uses topic modeling and cosine similarity to produce episode recommendations in a web app. The text was scraped from the fan website, https://greysanatomy.fandom.com/wiki/Grey%27s_Anatomy_Episodes, which includes full episode summaries for nearly every episode that has aired (a handful of episodes were dropped from the analysis as they were missing full summaries). I conducted text pre-processing and used TFIDF word vectors and a NMF model to produce 13 topics. With my knowledge of the show, I defined these categories. I write the episode-topic matrix to a SQLite database table, then read the table into a Pandas data frame. I find cosine similarity using topic scores as vectors and produce recommendations for the 10 most similar episodes for each episode. The user can select an episode from a dropdown menu and receive the recommendations.

Data Description:

I scraped the fan website, https://greysanatomy.fandom.com/wiki/Grey%27s_Anatomy_Episodes , for episode summaries. As opposed to full episode transcripts, using summaries in theory provide enough information for topic modeling with about a third of the words included in a full transcript.

Algorithms:

I use TFIDF word vectors and an NMF model that produced 13 topics. I set max_df=0.5 and min_df=.01 to avoid common and extremely rare words from confusing the topic categories. I find cosine similarity using topic scores as vectors.

Tools:

Python reference, BeautifulSoup for web-scraping
NumPy and Pandas for data manipulation
SQL for data storage
Scikit-learn for topic modeling and cosine similarity

Tableau for visualization
Streamlit for web app deployment

Communication:

The web app can be found at this link:
https://share.streamlit.io/mackerman4/greys_recommender/main
I produced slides containing the following visuals (including a video of the web app).