



What Do You Like About This Cult Film?

11.20.2019

Emily Hillenbrand

Metis Data Science

Final Project

Seattle, WA

Overview

A cult classic film is a film that was not initially commercially successful but has since developed a cult following. I like them because I like learning about unusual or obscure things that are known to be enjoyable for various reasons.

What do people like about cult films? It depends on the film and the person. When I'm looking for a film recommendation based on a film that I've seen, I don't necessarily want one that's identical, but I might want one with the same level of silliness or weirdness or cleverness.

I want to describe cult films based on the different things that people are saying about them, and then use that to create a recommendation app where the user can specify which topics are important to them. That is, I want the app to recommend films based on a specified film's distribution of selected topics.

Goals

1. Categorize Wikipedia's [list of cult films](#). Identify and name helpful topics that describe them.
2. Create an app that returns a list of recommended films based on similarity to an input film's topic distribution.

Data Sources

Cult Film Info	Acquisition	Description	MVP?
title	scrape	List of cult films from Wikipedia	yes
plot	dataset	Movie Plots dataset from Kaggle	yes
review	scrape	IMDb "most helpful" user review	yes
director	dataset	Movie Plots dataset from Kaggle	no
genre	scrape	Wikipedia (dataset genres are inaccurate)	no
rating	scrape	IMDb	no

Milestones

I. MVP

Create topic models based on a film's plot and reviews, then combine them with an ensemble classifier. Use this to create a function that returns a list of the top 5 recommended films based on similarity to an input film's topic distribution.

II. Iterate

Allow the user to specify which topics are relevant, then return films with the same levels of those topics.

Create an interactive widget that allows the user to modify the topic levels using sliders.

III. Stretch Goals

Create additional features from film title, director, genre, and rating, and see if they improve the app.

See if I can beat the MVP with a neural network and LSTM.

Build a web app.

Known Unknowns

I. Quality of Recommendations

I don't have extensive domain knowledge of cult films, so my ability to judge the quality of output will be limited. This may make optimization difficult, or I might not be able to identify why something isn't working the way I expect.

II. What are people saying?

I'm planning on taking only the "most helpful" user review from IMDb, but this might not give me enough information to create a clear topic model. I may need to collect multiple reviews, possibly even from multiple websites. I may also need to collect professional reviews.