Aidan Duffy

BU MET

CS 677

Final Project Summary

For my final project, I wanted to construct a recommendation system for platforms like YouTube where users provide their data, and in turn, the platform suggests further content that might enjoy. This is called a “content-based recommendation system”, and as an aspiring software developer, this seems like an interesting area to focus on as well as a valuable skill to understand. One area of interest I have in my private life is film. As of this point, I have watched nearly 80 movies so far this year, some re-watches but many are new. I really enjoy websites like “Letterboxd” where users can post reviews, both serious and comical, about their opinions on movies. Because of this, I wanted to design a system that could recommend movies to people!

As I noted in a text file among my source code, I referenced numerous websites that helped provide me with the datasets required to train my program (ie metadata on 45,000 movies), brief conceptual overviews on how recommendation systems work, as well as more in depth technical overviews on the system. I found numerous new functions in the modules and libraries we already use on a weekly basis for our homework assignments, which was really fascinating to me.

In terms of technical progress, I feel that I was able to develop a rudimentary but robust recommendation system given the data set I discovered online. My system can currently factor in plot synopses, genre, and cast and crew members into how it will recommend films to a person. The program is also able to display the most highly rated films in a given percentile of popularity. For instance, films in the tenth percentile only required one IMDb review to be factored in, versus the 90th percentile which required 160 reviews. Therefore, depending on user input, these results can be heavily skewed. For instance, generally, *The Shawshank Redemption* and the *Godfather* are the top rated films, but if we increase or decrease required review count, it fluctuates between those films and Christopher Nolan’s *Interstellar* (the most popular film) as well as small, unknown art house films with almost no reviews. Eventually I would like to factor in a user’s own reviews to more accurately tailor the recommendation. Currently, the user just provides a film title and the system tells them the ten most similar films. I would also like to add filters to this so that a user could opt out of having very unpopular films, older films, or foreign films from their recommendations.

In terms of how to run the file, all you need to do is download the datasets provided by the link in the works cited text file and run the main program through python. The program will prompt you for any required input. When inputting a film name, it is not case sensitive, but please ensure everything is spelled properly! I hope to add functionality in the future where the program is able to detect a small typo and ask “did you mean…?” The dataset files must be placed in a directory called “Data” that is in the same location as the main.py file. I will attempt to move most of the data with the source files, however, the files are quite large.

Overall, I am very excited with the progress I was able to make on this project, and I look forward to continuing to develop it! Thank you for such a great class and have a wonderful summer!