Week 10- Write up

Felipe Rodriguez

Bellevue University

DSC 630 Predictive Analytics

Professor Andrew Hua

November 3, 2023

Week 10 – Write up

This recommender system is created using the guide from the article "How To Build Your First Recommender System Using Python & MovieLens Dataset". The first step of creating this system is importing the data and ensuring that it is loaded correctly. For this system, there are two data frames being created. One contains the titles and genres of the movies, and the other contains the ratings of the movies.

Once that the data has been loaded in, it needs to be joined so that all the information is in one data frame. This data frame will include the movie titles and the ratings. Next, the total amount of rating is calculated to understand the proportion of the ratings. "The rating of a movie is proportional to the total number of ratings it has" (Nair, 2019). The recommender will use the ratings from each user for each movie, to create correlations. With that, the data will need to be pivoted to have all the ratings as values.

With the data pivoted, the ratings can be used to determine correlations between movies. The first example of this system will use the movie title 'Jumanji (1995)'. Please note in order for this recommender system to work, the title needs to be in the same format as the title in the data frame, or else it will not load recommendations. With the correlations added, the number of ratings each movie has will also be a factor. This way a filter can be included to omit ratings that fall below a threshold of total ratings received. Below is the example of Jumanji. This system shows the top 10 movies correlated to Jumanji when the title has more than 100 ratings.

```
# Creates recommendation of highest correlating movies only when there are 100 or more reviews
recc = recommendation[recommendation['Total Ratings']>100].sort_values('Correlation',ascending=False).reset_index()

recc = recc.merge(df,on='title', how='left')
recc.head(11)
```

	title	Correlation	Total Ratings	movieId	genres
0	Jumanji (1995)	1.000000	110	2	Adventure|Children|Fantasy
1	Cliffhanger (1993)	0.581001	101	434	Action|Adventure|Thriller
2	True Lies (1994)	0.493617	178	380	Action|Adventure|Comedy|Romance|Thriller
3	Back to the Future (1985)	0.485140	171	1270	Adventure|Comedy|Sci-Fi
4	Mrs. Doubtfire (1993)	0.480007	144	500	Comedy|Drama
5	Net, The (1995)	0.474888	112	185	Action|Crime|Thriller
6	Trainspotting (1996)	0.464547	102	778	Comedy|Crime|Drama
7	Twister (1996)	0.460929	123	736	Action|Adventure|Romance|Thriller
8	Incredibles, The (2004)	0.460369	125	8961	Action|Adventure|Animation|Children|Comedy
9	Bourne Identity, The (2002)	0.440918	112	5418	Action|Mystery|Thriller
10	Mask, The (1994)	0.440507	157	367	Action|Comedy|Crime|Fantasy

Week 10 – Write up

Reference:

Nair, A. (2020, October 10). *How to build your first recommender system using Python & Movielens Dataset*. Analytics India Magazine. https://analyticsindiamag.com/how-to-build-your-first-recommender-system-using-python-movielens-dataset/