SYSTEMS AND BIOMEDICAL ENGINEERING

Machine Learning

A7-Recommendation System

Date	**21/5/2023**
Name	Yousr Ashraf hejy

Collaborative Filtering (CF) recommender system:

It is a recommender system that predicts a user's preferences for a product based on the reactions by similar users.

The idea behind CF is that people who have similar preferences in the past are likely to have similar preferences in the future. Therefore, if we can find users who have similar preferences to a target user, we can use their preferences to predict what the target user will like.

It works by searching a large group of people and finding a smaller set of users with tastes similar to a particular user.

There are two main types of CF algorithms:
1. User-based CF that calculates the similarity between users based on their rating patterns and recommends films that similar users have rated highly.
2. Item-based CF which calculates the similarity between films based on the rating patterns of users who have rated both films and recommends films that are similar to those that the target user has already rated highly.

Steps to create this system:

1. Load the required libraries and data files
2. Prepare the data by filtering the ratings data frame to keep only the first 200 users then merge the first 200 movies data frame with the filtered ratings data frame to create a new data frame with movie ratings for the first 200 users only
3. Create a pivot table from the new data frame, with user IDs as rows and movie titles as columns, and the movie ratings as values and fill the missing values with zero or the mean value of the column, but I choose to assume it zero cause I cant tell if they like it or not.
4. Calculate the cosine similarity matrix
5. Recommend top 10 similar movies by sorting the similarity scores in descending order and printing the top 10 movies with the highest similarity scores
6. Recommend 3 films to userID 200 by saving the ids of similar users to our user then loop on similar users and print the name of corresponding movies that the user didn't rate as it corresponds to the movie that user didn't watch.

The Used Function in the code:

1. Prepare_data: takes rating and movie data frames and return the combined dataset according to the shape required in the task (movie ratings for the first 200 users only).

	userId	movieId	rating	title
0	1	1	4.0	Toy Story (1995)
1	5	1	4.0	Toy Story (1995)
2	7	1	4.5	Toy Story (1995)
3	15	1	2.5	Toy Story (1995)
4	17	1	4.5	Toy Story (1995)

2. Cf_table: takes the combined data frame as input and returns the pivot table.

userId / title	1	3	4	5	6	7	8	9	11	12	...	191	192	193	194	195	196	197	198	199	200
Ace Ventura: When Nature Calls (1995)	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.5
Amazing Panda Adventure, The (1995)	0.0	0.0	0.0	0.0	4.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
American President, The (1995)	0.0	0.0	0.0	0.0	4.0	0.0	4.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0
Angels and Insects (1995)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
Anne Frank Remembered (1995)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

3. similarity_matrix: takes the pivot table and return the similarity scores data frame between movies.

title	Ace Ventura: When Nature Calls (1995)	Amazing Panda Adventure, The (1995)	American President, The (1995)	Angels and Insects (1995)	Anne Frank Remembered (1995)	Antonia's Line (Antonia) (1995)	Apollo 13 (1995)	Assassins (1995)	Awfully Big Adventure, An (1995)	Babe (1995)	...	Unstrung Heroes (1995)	Up Close and Personal (1996)	Usual Suspects, The (1995)
Ace Ventura: When Nature Calls (1995)	1.000000	0.166578	0.337001	0.000000	0.000000	0.080112	0.388591	0.104858	0.000000	0.342358	...	0.166578	0.092468	0.305973
Amazing Panda Adventure, The (1995)	0.166578	1.000000	0.235012	0.000000	0.000000	0.297044	0.232556	0.155520	0.000000	0.298277	...	0.705882	0.243809	0.063435
American President, The (1995)	0.337001	0.235012	1.000000	0.022456	0.000000	0.000000	0.520100	0.157799	0.000000	0.410753	...	0.100719	0.369623	0.365531
Angels and Insects (1995)	0.000000	0.000000	0.022456	1.000000	0.573539	0.331133	0.131876	0.000000	0.573539	0.208784	...	0.000000	0.203841	0.159923
Anne Frank Remembered (1995)	0.000000	0.000000	0.000000	0.573539	1.000000	0.577350	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000

4. recommender _system_item: returns the data frame of similar movies to a chosen movie by the user after printing the names of similar movies.

```
The similar movie for Toy Story (1995):      The similar movie for Waiting to Exhale (1995):
Braveheart (1995)                            Before and After (1996)
Apollo 13 (1995)                             Mighty Morphin Power Rangers: The Movie (1995)
Usual Suspects, The (1995)                   Three Wishes (1995)
Happy Gilmore (1996)                         Georgia (1995)
Die Hard: With a Vengeance (1995)            Mad Love (1995)
Seven (a.k.a. Se7en) (1995)                  Before the Rain (Pred dozhdot) (1994)
Babe (1995)                                  Reckless (1995)
Waterworld (1995)                            Cure, The (1995)
Batman Forever (1995)                        White Man's Burden (1995)
Taxi Driver (1976)                           Now and Then (1995)
```

5. recommender_system_user: Takes the userID and no. of recommendation desired to get the movie recommendations

```
The recommended movies for to userID of 200:
First Knight (1995)
Usual Suspects, The (1995)
Pocahontas (1995)
```