DAT 301 - Assignment 8: Recommender Systems

- **Name**: Colin Bowers (bowerc3)

- **Date**: Mar 28, 2024

Question 1

1-You want to design a recommendation system for an online bookstore that has been launched recently. The bookstore has over 1 million book titles, but its rating database has only 10,000 ratings. Which of the following would be a better recommendation system?

a) User-based collaborative filtering
b) Item-based collaborative filtering
c) Content-based recommendation.

In One sentence, justify your answer.

ANSWER:

c) Content-based recommendation.

Since the data available in terms of user-item interactions is sparse (only 10,000 ratings for 1 million books) using features from the items themselves (such as titles, authors, genres) would be more suitable in this scenario.

Question 2

2-Suppose the bookstore is using the recommendation system you suggested above. A customer has only rated two books:" Linear Algebra" and" Differential Equations" and both ratings are 5 out of 5 stars. Which of the following books is less likely to be recommended?

a) "Operating Systems"
b) "Crazy Rich Asians"
c) "Convex Optimization"
d) It depends on other users' ratings.

ANSWER:

c) "Crazy Rich Asians"

The customer has given previous high ratings on books related to mathematics. Out of the options given, “Convex Optimization” would most likely be recommended the most since the terms used related closely with math. Close to this is “Operating Systems” which is a technical subject that uses mathematics but not directly related. By converse, the words in the “Crazy Rich Asians” are not highly related to mathematics which would not align to the data we have on the user’s preferences.

Question 3

3-After some years, the bookstore has enough ratings that it starts to use a more advanced recommendation system like the one won the Netflix prize. Suppose the mean rating of books is 3.4 stars. Alice, a faithful customer, has rated 350 books and her average rating is 0.4 stars higher than average users’ ratings. “Animals Farm”, is a book title in the bookstore with 250,000 ratings whose average rating is 0.7 higher than global average.

What would be a baseline estimate of Alice’s rating for Animals Farms?

ANSWER:

From my own research, a significant method used in the winning model of the Netflix Prize competition was the baseline rating estimate, calculated as follows:

$$\textit{Baseline Estimate} = \textit{Global Mean Rating} + \textit{User Bias} + \textit{Item Bias}$$

Where:

1. Global Mean Rating is the average rating across all users and items.
2. User Bias is the difference between the average rating of the user and the global mean rating.
3. Item Bias is the difference between the average rating of the item and the global mean rating.

Therefore, in this scenario we have:

$$\text{Baseline Estimate} = 3.4 + 0.4 + 0.7 = \textbf{4.5 stars}$$

Therefore, the baseline estimate of Alice’s rating for “Animal Farm” would be 4.5 stars.