Assignment 3: Sequential Recommendations

How to run the code:
The program file is a .py file. When running the code, one requires to give the user ID of the interested user from the provided list, as the input to the program. Then it'll provide the top-10 group recommendations in 3 different sequences.

The assumptions made while completing the assignment:

- For Individual recommendation - It's assumed that obtaining **50 similar users** when making the predictions will provide a reasonable recommendation.

Other important things about the code:

- Since it's mentioned to provide the top-10 recommendations in 3 different sequences, the dataset is divided into 3 chunks as follows. First, the dataset is sorted according to the timestamp and divide the dataset into 2 halves. The first half is used to the system as the initial dataset. Then the second half is again divided into 2 halves. In each iteration, a new dataset is concatenated with the previous one.
- When selecting users for the group, a user similar to the initial user and a user dissimilar to the initial user were selected based on the Pearson correlation. So, the group consists of 2 similar users and one dissimilar user. (The initial user is taken as a user input to the program, and it's selected from the users in the initial dataset)
- When recommending movies to the user group, the movies that no one from the group has rated were selected.
- When selecting 50 similar users for making the individual predictions, the top 50 users who have rated for a given movie were selected.
- When calculating the Pearson correlation between 2 users, only the commonly rated movies were selected. This is done as there are many missing values in the dataset. Also, correlation is calculated if the commonly rated movie count is greater than 2. This is done because we can't say 2 users are similar if they have similar ratings for a very small number of movies like 1,2.

Explanations and clarifications of the method Used.

Steps:

- The dataset is divided into 3 sets. (First sort the dataset according to the timestamp and then divide it into 2 halves. Then the second half is again divided into 2 halves.)
- Select users to the user group where the group consists of 2 similar users and one dissimilar user. (Pearson correlation is used)
- A for loop is used to obtain group recommendations in 3 different sequences.

- In the first iteration, the first half of the dataset is used. And then in upcoming iterations, each other dataset is concatenated with the previous dataset.
- Obtained individual recommendations. (Based on Pearson correlation 50 similar users were selected to the given user and according to their ratings, the recommendations are done.
- Then an aggregation function is used to predict the group recommendation.

$$Aggregation\ function = (1 - alpha) * mean\ of\ the\ movie\ ratings\ of\ 3\ users + alpha * standard\ deviation\ of\ the\ movie\ ratings\ of\ 3\ users$$

- In each iteration user satisfaction is measured for the individual user.

$$User\ satisfaction = sum\ of\ user's\ top\ recommended\ movie\ ratings\ if\ the\ movie\ is\ in\ the\ top\ group\ recommendations\ /\ sum\ of\ the\ user's\ top\ recommended\ movie\ ratings$$

- Then the alpha value is updated according to the user satisfaction score.

$$alpha = previous\ iteration\ maximum\ user\ satisfaction - previous\ iteration\ minimum\ user\ satisfaction$$

- In the initial step alpha is taken as 0.

About the used Method

Sequential group recommendations are those that are developed across several iterations for a group of consumers. The aim of this is to enhance group user satisfaction through iterative refinement of recommendations. This enables to retention of knowledge from past interactions to keep the output diverse Every time a sequence is completed, the system receives new data that it uses to help incorporate new information over time, allowing the model to adjust to shifting user behavior patterns or preferences.

The proposed method combines the mean and standard deviation of individual recommendations to generate an aggregate recommendation for the user group. The variability (standard deviation) of individual recommendations and central tendency (mean) of individual recommendations are both acknowledged by this method. The mean takes the average of user ratings. While a higher standard deviation indicates greater diversity in preferences within the group, a lower standard deviation indicates smaller diversity. Depending on how user satisfaction varies, the alpha value dynamically modifies the standard deviation's and central tendency's influence. A higher alpha makes the suggestions more responsive to individual variances because it reflects a wider range of user satisfaction within the current iteration. It enables the model to adjust to the group's varied preferences.

The method is sensitive to changes in user satisfaction because of the usage of alpha values and standard deviation, which can be helpful in capturing the diversity of group preferences. So, the proposed methos is suitable for sequential recommendations as it helps to reduce any

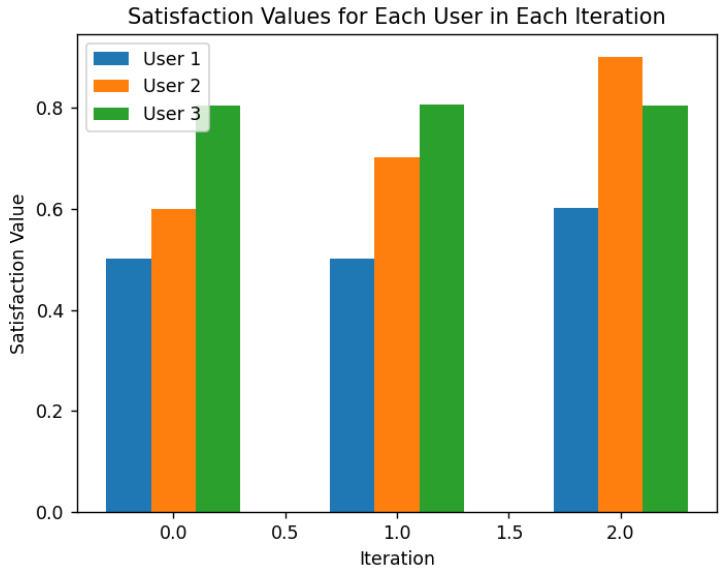
diversification within the user group and increase the user satisfaction of the group members with time.

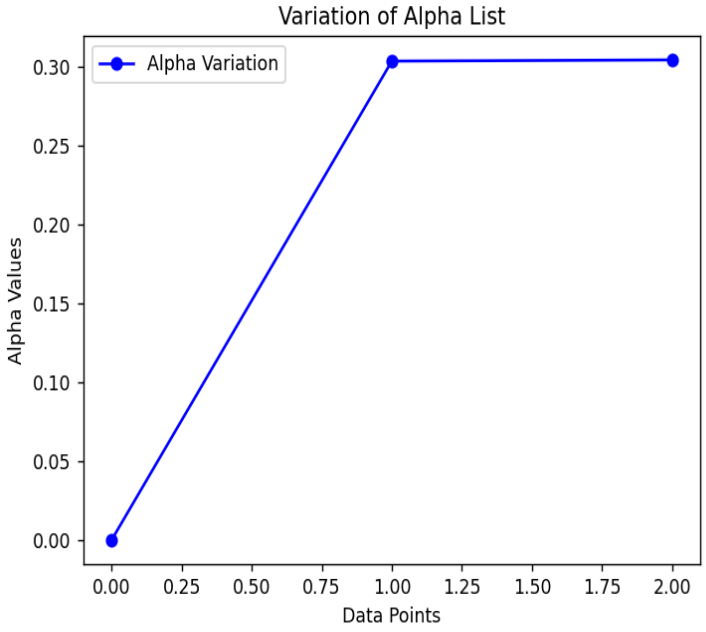
Results

```
10 most suitable movies to user group with aggregated function method in iteration 1 are
275 :  Sense and Sensibility (1995)
134 :  Citizen Kane (1941)
191 :  Amadeus (1984)
223 :  Sling Blade (1996)
9 :  Dead Man Walking (1995)
276 :  Leaving Las Vegas (1995)
169 :  Wrong Trousers, The (1993)
22 :  Braveheart (1995)
427 :  To Kill a Mockingbird (1962)
183 :  Alien (1979)
Group satisfaction in iteration 1 are 0.634852136017289
```

```
10 most suitable movies to user group with aggregated function method in iteration 2 are
313 :  Titanic (1997)
22 :  Braveheart (1995)
183 :  Alien (1979)
427 :  To Kill a Mockingbird (1962)
408 :  Close Shave, A (1995)
480 :  North by Northwest (1959)
134 :  Citizen Kane (1941)
191 :  Amadeus (1984)
223 :  Sling Blade (1996)
169 :  Wrong Trousers, The (1993)
Group satisfaction in iteration 2 are 0.6693843116523119
```

```
10 most suitable movies to user group with aggregated function method in iteration 3 are
313 :  Titanic (1997)
272 :  Good Will Hunting (1997)
480 :  North by Northwest (1959)
427 :  To Kill a Mockingbird (1962)
285 :  Secrets & Lies (1996)
134 :  Citizen Kane (1941)
408 :  Close Shave, A (1995)
316 :  As Good As It Gets (1997)
315 :  Apt Pupil (1998)
183 :  Alien (1979)
Group satisfaction in iteration 3 are 0.7686147363130155
```





Submitted by
iqra.nasir@tuni.fi
piyumika.herathmudiyanselage@tuni.fi