CPS842 Project

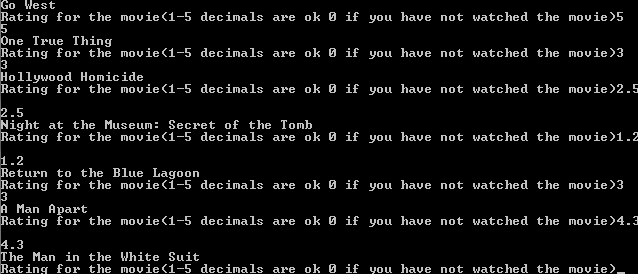
For our project, we decide to do a movie recommender program. We used files from both omdb and movie lens. We used omdb to acquire the proper names for the Movies. Omdb allowed us to do this because it pairs the imdb movie ID and real movie title together. This is necessary because the movie lens files gave imdb movie IDs and so we could not retrieve real movie titles without importing omdb.

With movie lens, we parsed two files. The first file gave us user ratings and we used this as a database. The data base contains 718 unique users, and 100235 movie ratings in total. With this, we have users to compare to. Since there are so many movie ratings, most movies that the user rates will be compared with many other users. This saved us a lot of time because we did not have to create mock profiles to compare to. The second file linked the ratings to the movie, using the imdb movie ID as explained above.

We used the Pearson Similarity Correlation to compute the similarity score between the current user and the users in the data base. We then take all users who have positive similarity scores. We also used the User-Based Collaborative Filtering method to calculate the predicted ratings for movies that the user has not seen yet. If the predicted rating is higher than 3, it will be recommended to the user. The program reads through the movie IDs and finds the real movie title by finding it in the dictionary structure that holds the movie ID linked to the real movie title.

Sample user similarity matrix for one user:

|  |  |  |
| --- | --- | --- |
| userId | movieId | rating |
| 1 | 1 | 5 |
| 1 | 2 | 3 |
| 1 | 10 | 3 |
| 1 | 32 | 4 |
| 1 | 34 | 4 |
| 1 | 47 | 3 |
| 1 | 50 | 4 |
| 1 | 62 | 4 |
| 1 | 150 | 4 |
| 1 | 153 | 3 |
| 1 | 160 | 3 |
| 1 | 161 | 4 |
| 1 | 165 | 4 |
| 1 | 185 | 3 |
| 1 | 208 | 3 |
|  |  |  |

Sample Input Screen:

Sample Recommended Movies: