CAP- 5610 Machine Learning

Homework 5

Name: Nabila Shahnaz Khan

NID: 5067496

Machine Learning for Recommender Systems

Task 3:

c) Compute the average MAE and RMSE of the Probabilistic Matrix Factorization (PMF), User based Collaborative Filtering, Item based Collaborative Filtering, under the 5-folds cross-validation.

Answer: The average MAE and RMSE of the Probabilistic Matrix Factorization (PMF), User based Collaborative Filtering (UBCF) and Item based Collaborative Filtering (IBCF) under the 5-folds cross-validation is given below:

Model	Average MAE	Average RMSE
Probabilistic Matrix Factorization (PMF)	0.7781	1.0083
User-based Collaborative Filtering (UBCF)	0.7444	0.9685
Item-based Collaborative Filtering (IBCF)	0.7210	0.9348

d) Compare the average (mean) performances of User-based collaborative filtering, item-based collaborative filtering, PMF with respect to RMSE and MAE. Which ML model is the best in the movie rating data?

Answer: Item-based Collaborative Filtering has the lowest RMSE and MAE, User-based Collaborative Filtering (UBCF) has the second lowest RMSD and MAE and Probabilistic Matrix Factorization (PMF) has the highest RMSD and MAE. As we know, the lower the value of RMSD and MAE, the better is the performance of a model in ranking. So in this case, Item-based Collaborative Filtering (IBCF) model is the best in movie rating.

e) Examine how the cosine, MSD (Mean Squared Difference), and Pearson similarities impact the performances of User based Collaborative Filtering and Item based Collaborative Filtering. Plot your results. Is the impact of the three metrics on User based Collaborative Filtering consistent with the impact of the three metrics on Item based Collaborative Filtering?

Answer: The impact of cosine, MSD (Mean Squared Difference), and Pearson similarities on the performances of User-based Collaborative Filtering is given below in Figure 1. As we can see, for User-based Collaborative Filtering, MSD performs the best. Cosine similarity performs slightly better than Pearson similarity.

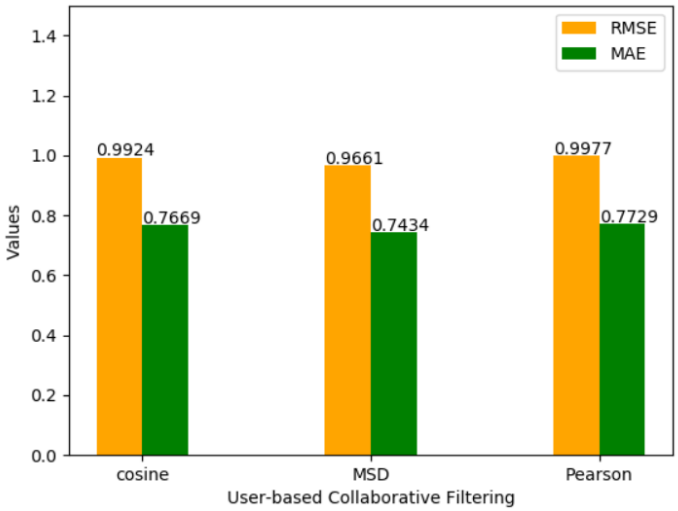

Figure 1: Impact of cosine, MSD and Pearson similarities on User-based Collaborative Filtering

The impact of cosine, MSD (Mean Squared Difference), and Pearson similarities on the performances of Item-based Collaborative Filtering is given below in Figure 2. For Item-based Collaborative Filtering, MSD performs best again. But this time, Pearson similarity performs slightly better than cosine similarity.

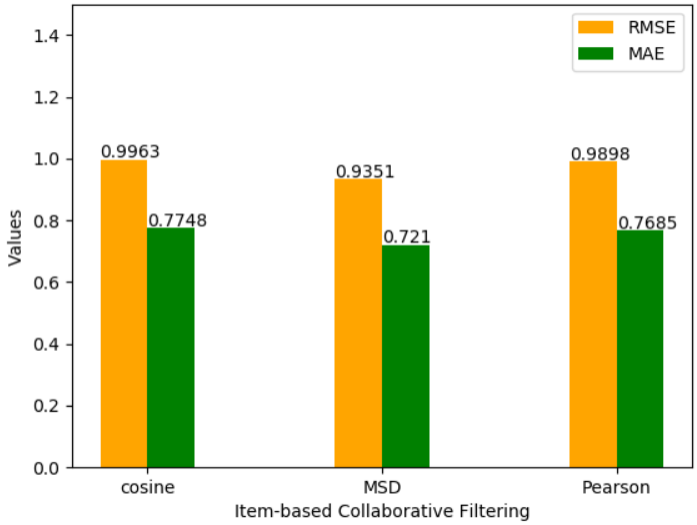

Figure 2: Impact of cosine, MSD, and Pearson similarities on Item-based Collaborative Filtering

In summary, MSD performs best for both User-based Collaborative Filtering and Item-based Collaborative Filtering. But the impact of Pearson and cosine similarities are not consistent on the both models.

f) Examine how the number of neighbors impacts the performances of User based Collaborative Filtering and Item based Collaborative Filtering? Plot your results.

Answer: The impact of the number of neighbors on the RMSE and MAE performance metrics of both User-based Collaborative Filtering and Item-based Collaborative Filtering are shown in Figure 3 and Figure 4 respectively.

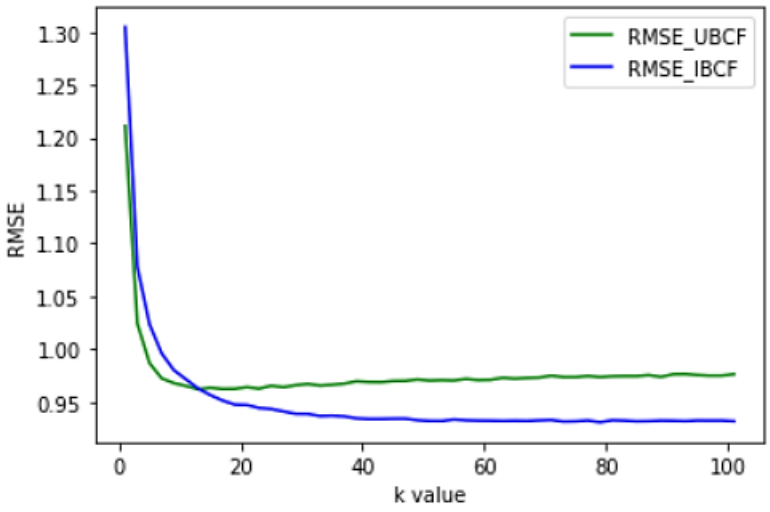

Figure 3: Impact of K-value on RMSE of UBCF and IBCF Models

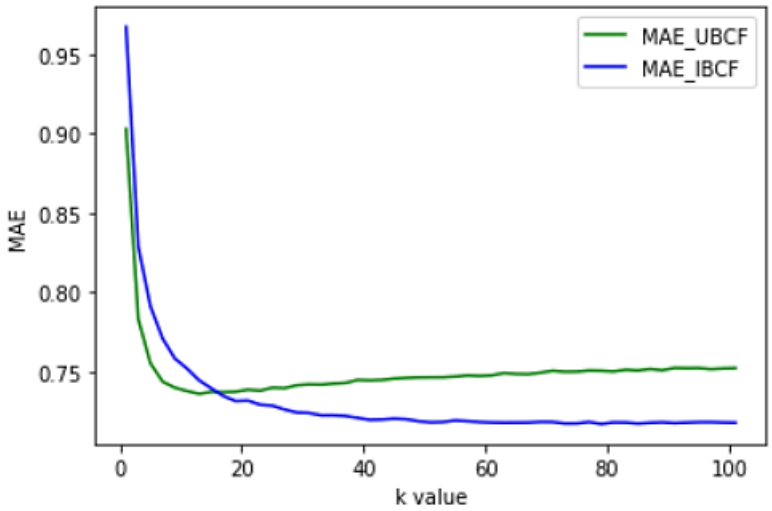

Figure 4: Impact of K-value on MAE of UBCF and IBCF Models

The MAE and RMSE values are calculated for both the User-based Collaborative Filtering (UBCF) model and the Item-based Collaborative Filtering (IBCF) models at K= {1, 3, 5, 7 , ……., 99, 101}. Both the RMSE and MAE values start out very high at K=1 and drops down sharply with increasing number of neighbors. The RMSE and MAE curve with respect to the number of neighbors looks almost similar for both models. However, the UBCF reaches its minimum error value at a lower K-value compared to the IBCF. Also, in-case of IBCF, both RMSE and MAE value reaches saturation at a minimum value and doesn't increase

after that. But in-case of UBCF, both RMSE and MAE values increase slightly after reaching the lowest value and reaches saturation at that slightly increased value.

g) Identify the best number of neighbor (denoted by K) for User/Item based collaborative filtering in terms of RMSE. Is the best K of User based collaborative filtering the same with the best K of Item based collaborative filtering?

Answer: The User-based Collaborative Filtering model reaches its lowest RMSE value at K = 6 while the Item-based Collaborative Filtering model reaches its lowest RMSE at K = 39. So, the best K of UBCF and IBCF models are not the same. The UBCF model reaches its best accuracy at a lower K-value compared to the IBCF model.

Code Link:
https://github.com/NabilaKhan/CAP-5610-Machine-Learning-/blob/main/CAP_5610_HW5.ipynb