

CSE 640 - Collaborative Filtering

Assignment - I

Abhishek Maiti (2016005)

There are 2 scripts. ItemBased.py does item based recommendation and UserBased.py does the user based recommendation.

For Item based Recommendation, the mean absolute error was coming to be 2.99.

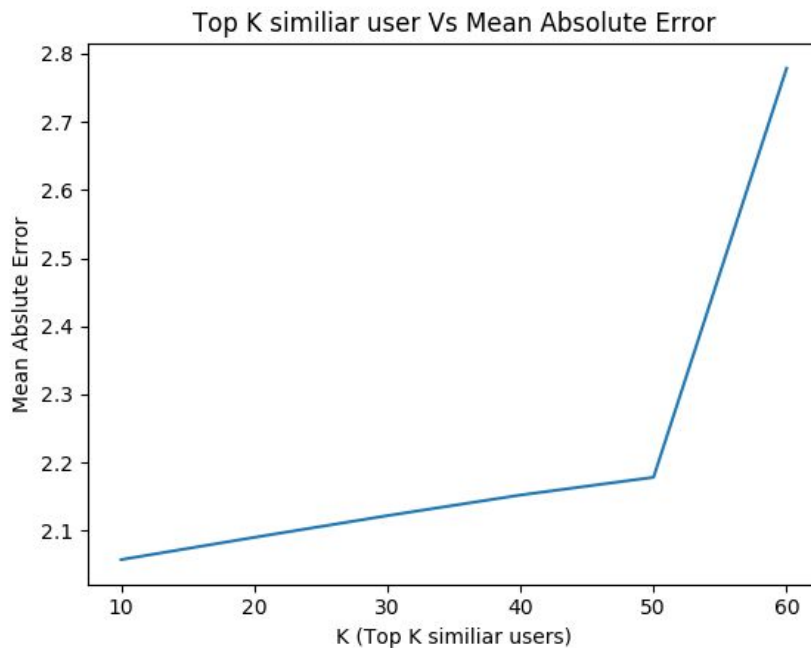
The individual errors for each train test split i.e. taking u1.base as train set and u1.test as the test set and so till 5 are:

[3.052293747300706, 2.9734238512428126, 2.956413131893428, 2.976672538637002, 3.0244707641477735]

For the User-based recommendation, the mean absolute error for the prediction as is as follows.

K (Taking top K similar users)	Mean Absolute Error
Taking all of them	2.778
10	2.057
20	2.090
30	2.122
40	2.152
50	2.178

The plot for MAE Vs k is as follows. K = 60 signifies taking all the users to find the predicted rating. Other Ks signifies taking top K similar users to find the rating.



CONCLUSION

As we can see user based recommendation system outperforms the item based recommendation. That is why this is more preferred in the industry. Also in the user based model, the error increases as we include dissimilar users to influence our ratings.

The individual errors for each split for all values of k are as follows.

K = 10

[2.014271864795014, 2.0749984114210016, 2.0923631383289982, 2.041943345749709, 2.063952631459724]

K = 20

[2.055141000480232, 2.1207821520377705, 2.111157999614918, 2.0789801193461117, 2.0847333433418904]

K = 30

[2.087976696986358, 2.159274743385389, 2.132746915223271, 2.114113495739921, 2.116107246609002]

K = 40

[2.124725445253917, 2.1915741411063445, 2.1583119055920705,
2.144208089568555, 2.142793668495481]

K = 50

[2.154622065893669, 2.2192690358294342, 2.1833970074186877,
2.168992738071769, 2.16540045771405]

K = TRAIN SET

[2.787376010322463, 2.808302396421165, 2.7748137964340205,
2.7626085653718224, 2.7607582241448845]