User-Based Collaborative Filtering Algorithms

The plain cosine similarity resulted in the MAE of 0.7894844

Pearson correlation method gave an increase of error to 0.8188. My expectations were Pearson correlation would increase the accuracy, the best possible explanation for increase in the MAE could be due to small sample sizes. Since there are less users who have commonly rated, which reduces the effectiveness of Pearson correlation.

Cosine Similarity	
MAE 5	0.81430649
MAE 10	0.787166
MAE 20	0.769364
Overall	0.7894844

Pearson Correlation	
MAE 5	0.868701
MAE 10	0.792
MAE 20	0.795987
Overall	0.8188

Extensions to the basic user-based collaborative filtering algorithms

Implementing the inverse user frequency slightly reduced the accuracy from the base Pearson correlation. Predicted scores were often outside the range of 1-5 and many of the scores had to be levelled to 1 or 5.

Case amplification with rho = 2.5 resulted in better scores than base Pearson correlation. I tried combining both the techniques but could not get better score than cosine similarity

Pearson IUF	
MAE 5	0.872202
MAE 10	0.79716666
MAE 20	0.798109386
Overall	0.82221966

Pearson Case Amplification	
MAE 5	0.861823
MAE 10	0.791666
MAE 20	0.795987267
Overall	0.816532589

Item-Based Collaborative Filtering Algorithm

Basic implementation of item-based collaborative filtering gave results just behind the cosine similarity (best model so far). On further studying the rankings the ranking was different, but the MAE is similar to cosine similarity.

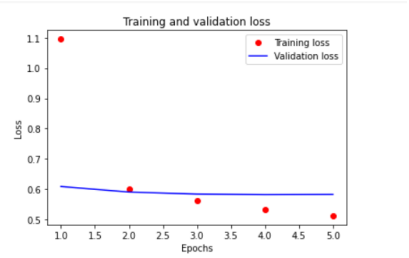
Adj Cosine Item Based	
MAE 5	0.876079
MAE 10	0.790333
MAE 20	0.769364
Overall	0.809556

Implement your own algorithm

I experimented with few different approaches with goal in the mind to beat the cosine similarity algorithm. I happen to try out the deep learning approach with Embedding and Dense layers.

But the accuracy deteriorated greatly due to less data. Deep learning approach requires huge amount of data to be trained on. Thus, resulted in the very less MAE.

I happen try out using the output of all the previous models and sending it into the deep learning algorithm which slightly improved the results with naïve deep learning algorithm. Attached is the Training and Validation loss graph.



Deep Learning	
MAE 5	2.261598
MAE 10	2.315667
MAE 20	2.327096
Overall	2.302783

Personal Deep Learning with Stacking	
MAE 5	2.167812929
MAE 10	2.2401666
MAE 20	2.2241
Overall	2.2096