Lucas Summers (lsumme01@calpoly.edu)
Xiuyuan Qiu (xiqiu@calpoly.edu)

CSC 466 Lab 5 Report

1. Methods Implemented

We implemented four different memory-based collaborative filtering methods:

1. Weighted sum
2. Adjusted weighted sum
3. Weighted N Nearest Neighbors sum
4. Adjusted weighted N nearest Neighbors sum

On top of this, we implemented two similarity measures to be used by all four methods:

- Cosine similarity
- Pearson Correlation

2. Research Questions

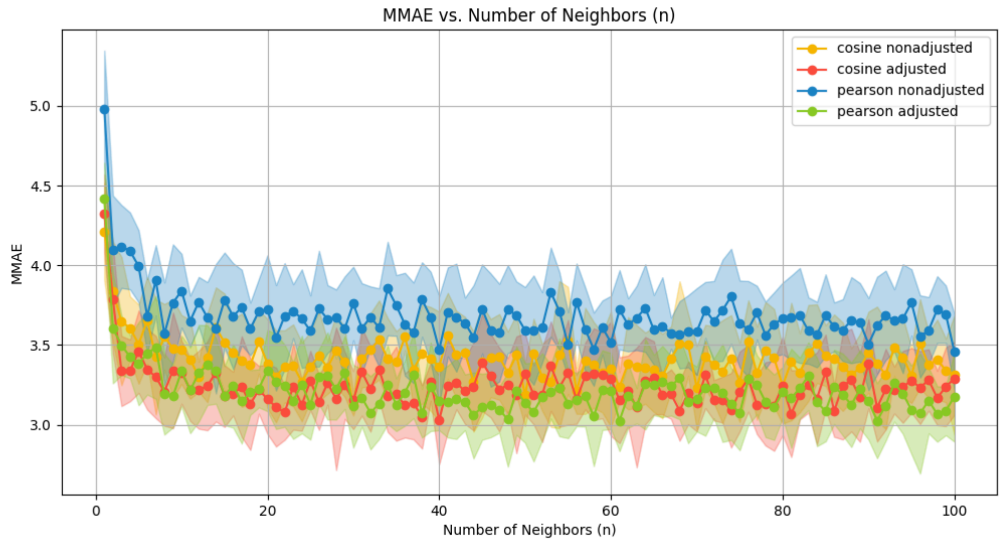
We have multiple questions to study in our experiments:

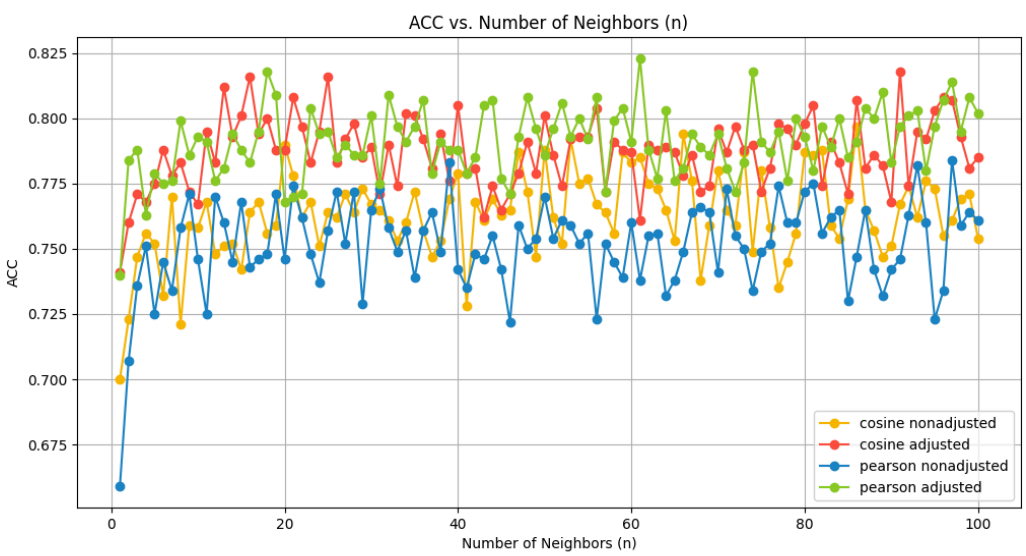
1. What is the best value of **N** for the N nearest neighbors models?
2. Do adjusted weighted sums produce better results than regular weighted sums?
3. What similarity measure produces the best results?
4. Finally, combining all the questions above, *what method produces the best overall results?*

3. Experiments & Results

RQ1: What is the best value of N for the N nearest neighbors methods?

We ran a grid search from N=1 to 100 for weighted and adjusted weighted n nearest neighbors for both Cosine Similarity and Pearson Correlation. For each set of parameters, we ran a random evaluation on 100 points 10 times and measured the results.



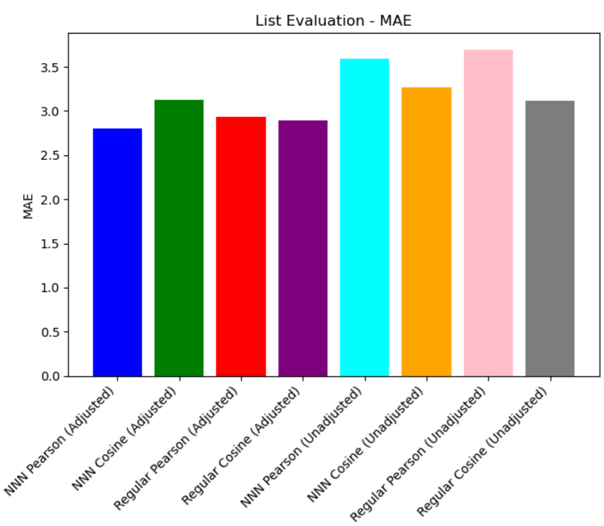


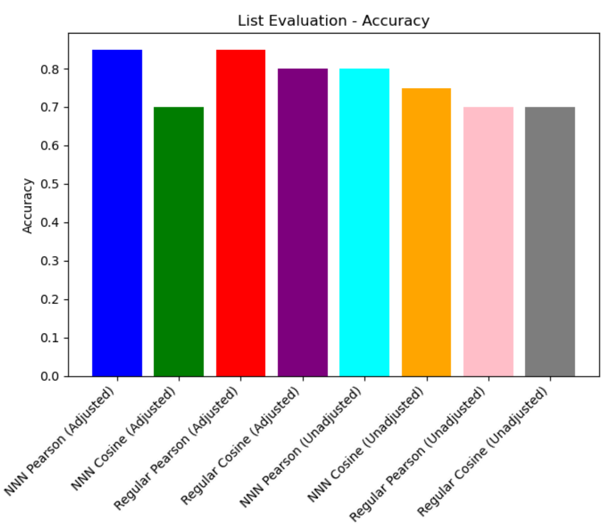
Generally, we found that values of N >= 20 seemed to work the best, and that scores seem to plateau beyond that (or have a near-imperceptible worsening trend). This seems to indicate that looking at all users for predictive purposes will generally be fine, since N nearest just removes users that are less similar to the query (i.e. users with less impact), though perhaps with extremely large userbases, that tail of dissimilar users will start to negatively impact our scores.

RQ2: Do adjusted weighted sums produce better results than regular weighted sums?
The adjusted sums (green and red on the above graph) consistently produce lower MMAE and higher accuracy scores than unadjusted (blue and green), so adjusting the sums seem to produce better results.

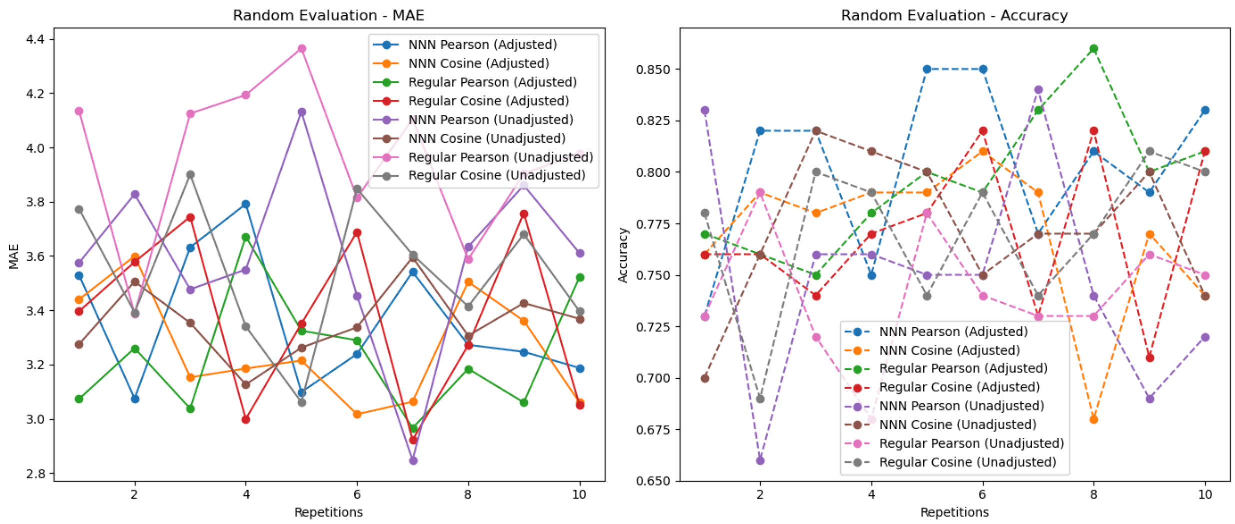
RQ3: What similarity measure produces the best results?
It seems cosine nonadjusted (yellow) performs better than Pearson nonadjusted (blue), but with adjusted weights, both measures perform similarly well.

RQ4: What method produces the best overall results?
We used a list of 50 pre-generated test cases, testing each possible model to show how they directly compare to the same predictions. Using the results of RQ1, we used a value of N=20 for all NNN based models.

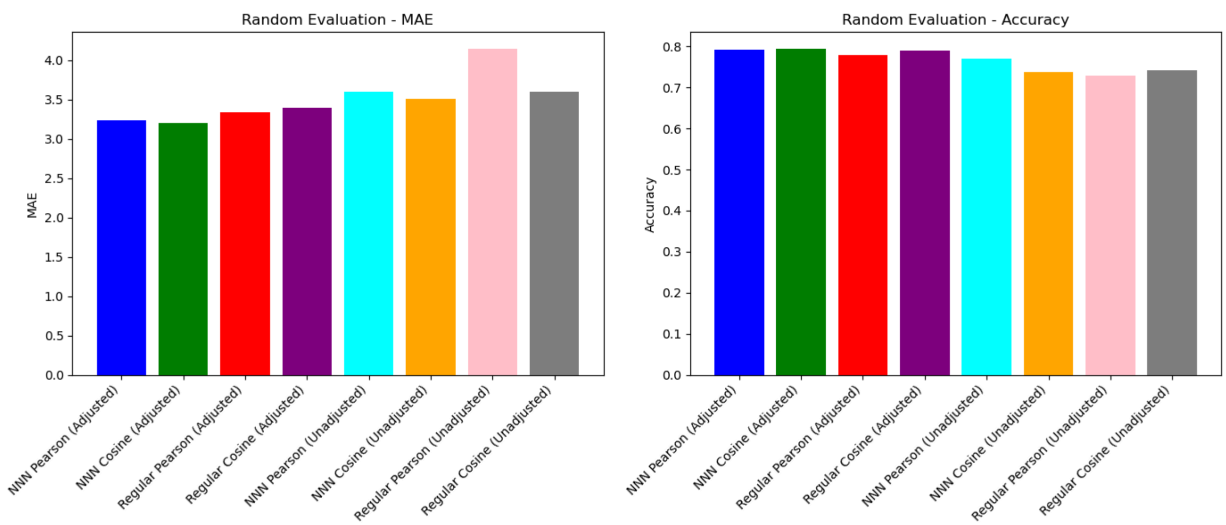




Using list-based evaluation shows that adjusted NNN with Pearson and regular adjusted sum with Pearson have the lowest MAE and highest accuracy out of all the possible models. While accuracy between the two is equal (~85%), the NNN model has a slightly better MAE in this case (~2.8).



We also used random-sample evaluation to confirm our results above. We ran 10 reps for each model, each rep with a sampling size of 100 points. The graph above shows the rep by rep breakdown, but is too noisy to meaningfully interpret.



This shows the average MAE and average accuracy score from the 10 reps shown above. All models are very similar in results, with accuracies only differing by less than 10% and MAE's differing by only about 0.5. The results here are not foolproof because each model is tested on completely different samples, however, it seems like the best model is still adjusted NNN with Pearson adjusted as it seems to consistently provide the best results.

4. Conclusions

Finalizing our results, the best model we found is adjusted weighted NNN sum with Pearson (N=20), with accuracy rates steadily hovering around 80% and MAE staying around 3.0 when tested with both evaluation methods. Overall, through our experiments, we found that generally Pearson is the better similarity measure than Cosine, adjusted sums are better than nonadjusted sums, and NNN weighted sums are better than regular weighted sums.