June 8th, 2018

Lab 6 Report

Devon Martin and Shiv Sulkar

Methods Implemented:

Default Voting: This technique gets more accurate with large datasets.
Cosine Similarity: This method works best when many of the ratings deviate from the average.
Cosine Similarity Adjusted: This technique aims to eliminate the issue of different rating schemes in the dataset.

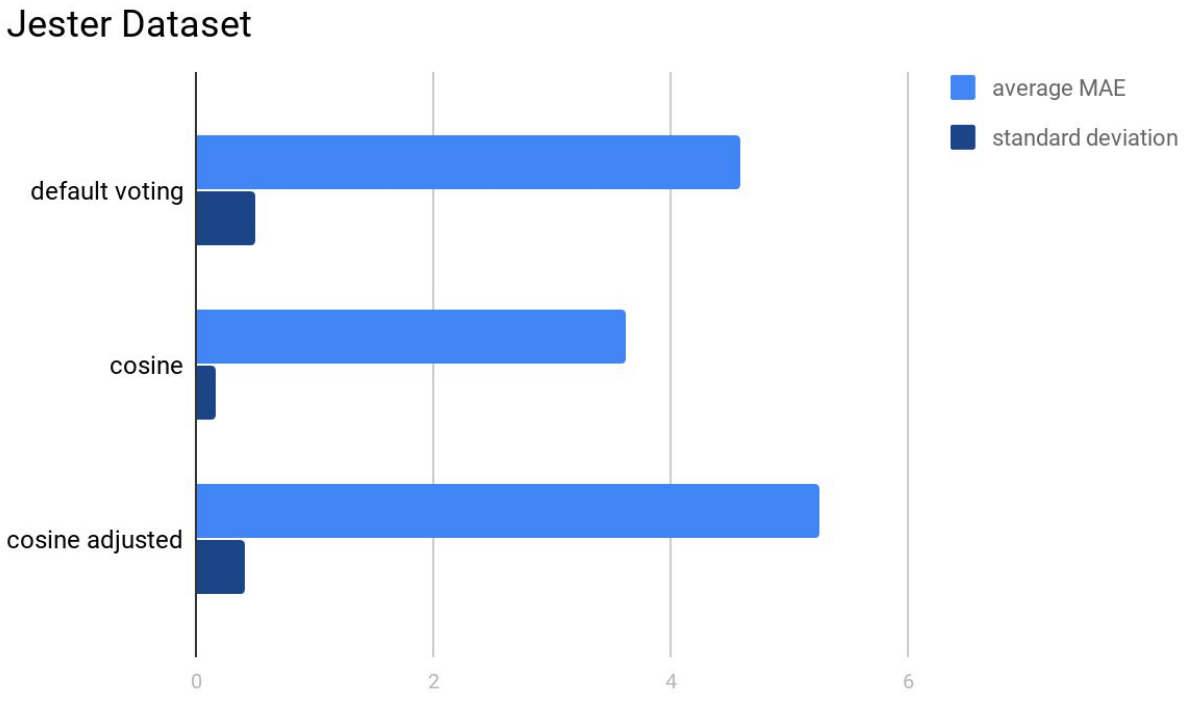
Research Questions:

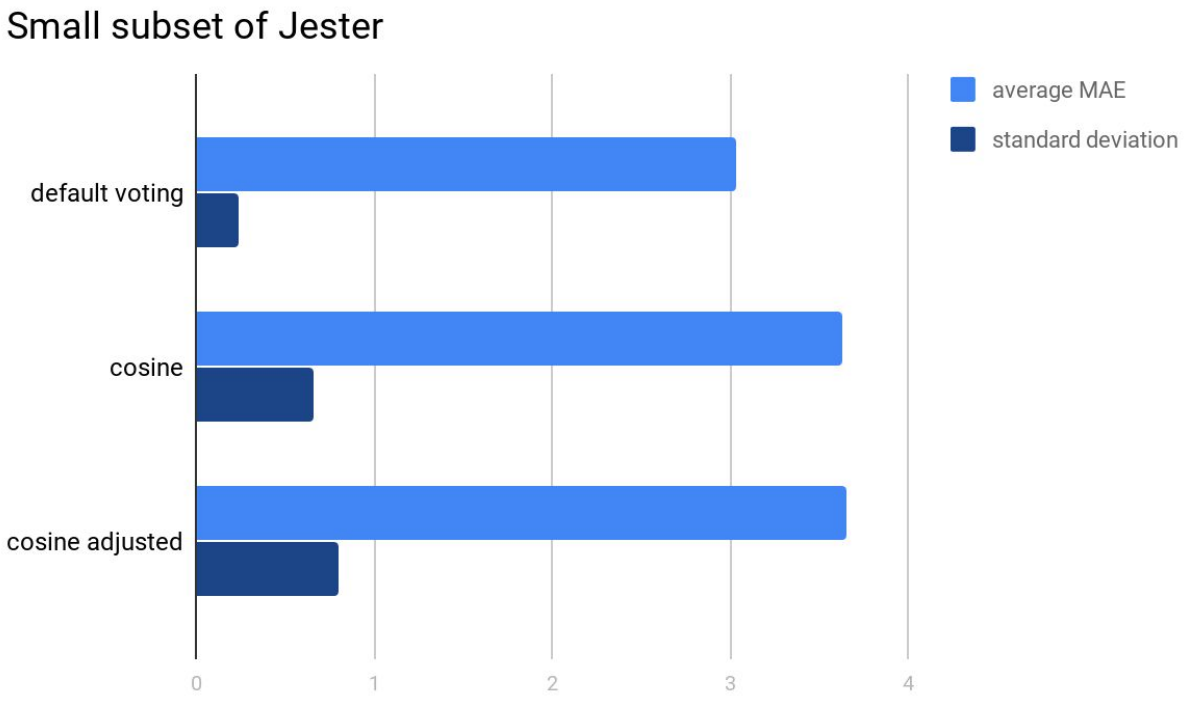
Which method has the highest MAE value?
Which method has the lowest standard deviation?
Which method is the slowest?
Which works best with large/small dataset?

Experiment:

We decided to shrink the jester subset into 3 columns and 10 rows to represent the small subset. We also chose a block that had very few null ratings. Then we ran 30 row tests 2 times for each method on both the large jester dataset and the small one.

Results (we also submitted txt output files):





Conclusion:

We found that the default voting technique worked best with the small subset of jester that had fewer null ratings. We expected it to work best on the large jester set but that was not the case.

This technique also took the longest amount of time to run. The cosine technique worked best with the large dataset. The cosine adjusted technique was the worst on both datasets which may indicate that everyone that participating in the creation of the dataset had the same rating scheme.