Collaborative Filtering on Netflix Ratings Dataset

There are two types of filtering for recommender systems.

1. Content-Based filtering
2. Collaborative Filtering

Content-Based filtering recommends the items based on the user's actions or feedback on previous items.

Collaborative Filtering uses similarities between users and items simultaneously to provide recommendations. Collaborative Filtering models can recommend an item to user based on the interests or feedback of a similar user.

Now implemented the Collaborative filtering and ran on the Netflix ratings data.
Used NumPy library in Python for faster processing of the data.

First converted the training data into pivot table of (users x movies) matrix and calculated the $\overline{v}$, user mean rating of the movies he rated, and calculated the mean difference matrix.
Later, using mean difference matrix calculated the Numerator part and Denominator part of the Correlation Weights separately using the APIs of the numpy for multi-dimensional arrays. Later calculated the kappa using the calculated weights. Finally, computed the predictions for the active user on the movies using the above calculated values.

With precomputed predictions and test data, predicted the ratings of the user for the movies. Using the predicted ratings and actual ratings calculated the mean absolute error and root mean squared error. The results are as follows.

Mean Absolute Error (MAE): 0.746
Root Mean Squared Error (RMSE): 0.941

The RMSE is more than MAE then there might be large difference in the error or there might be more outliers in the data.

The time took to do the predictions and calculations on machine with 16GB RAM is 467 seconds. Whereas, the same ran on a machine with more than 32GB RAM it took around 40 seconds only.