

Hunter Stusnick
PA2 Movies

URL from CodeClimate: <https://codeclimate.com/github/hdstusnick/movies-2>

URL from Github: <https://github.com/hdstusnick/movies-2>

Algorithm of the prediction method:

For clarity, the passed in user will be called the subject and all other users will be called users. Predict takes a subject and a movie and gets a list of all the users that were similar to the subject. It then checks each of those users to see if they had seen the movie that was passed in. if the user has seen the same movie, it then creates a map of the similarity of the user to the subject => rating the user gave. It then calculates a weighted average of the ratings based on how similar and returns a rounded int as the prediction.

Analysis:

The results from u1 are:

Number of guesses correct = 5209

Number wrong = 14791

Percentage correct 0.26

Average incorrectness of wrong guesses = 1.4

I used my same similarity method from movies part one and I think it could use some improvement. I could also take into consideration the genre when calculating user similarity to add more accuracy.

Benchmarking:

I did not know how to use time.now but the algorithm runs about 10 minutes for 20000 in the base and 80000 in the test. I could have improved this time by better preparing the data when I initially parsed the file as well as selecting users from a smaller group that could be defined by genre instead looking for similar users in the entire data set. On ever expanding input sizes this algorithm will be much too slow.