Programming Assignment #4

Build Recommender System

2013011457 컴퓨터전공

Jisoo Song

1. Summary of Algorithm

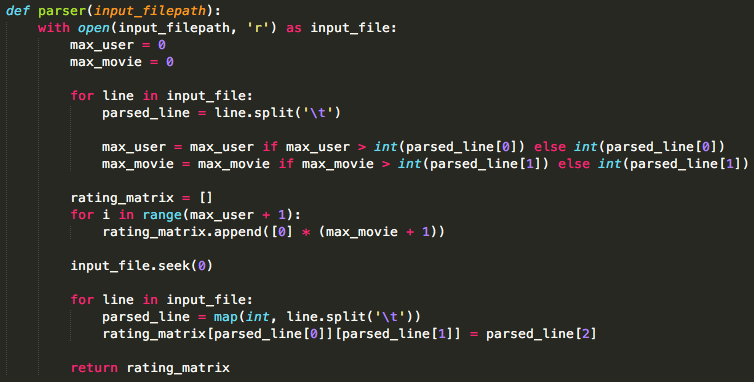
I implemented Movie Recommender System using user-based Collaborate Filtering (CF) algorithm. At first, my program parses the input file and make the matrix named ‘rating\_matrix’. This matrix has the information about what user rated for what movie as some score.

Using this matrix, my program calculates the similarity between each pair of all users. Then it makes the ‘user\_similarity\_matrix’.

After that, my program predicts a rating for unrated movie for each user, using ‘user\_similarity\_matrix’. It filters users who has similarity larger than 0.5 and already rated the target movie. If the filtered users are more than 5 persons, it predicts the rating as an average of ratings of filtered users. If not, it predicts the rating as an average of ‘target user’s rating average for other movies’ and ‘target movie’s rating average of other users’.

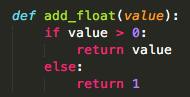
1. Details about each Functions

* Parser



parser function parses the input training data. At first it figures out the max number of user and movie to make a rating matrix. Then it reads input data line by line, and it writes the rating information on the rating matrix. It returns the rating matrix.

* Add\_float



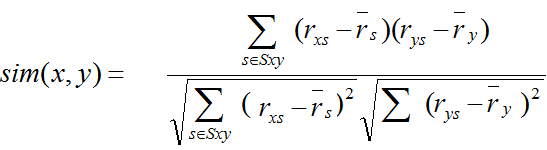
This function is made for handling ‘division by zero’ error. On the divide operation for the other functions, sometimes the divisor is zero and ‘division by zero’ error is occurred. To avoid the error, all divisor on the divide operation must check if it is zero or not. If it’s zero, it returns 1 to make divisor to 1. If it’s not zero, it returns the input value as it is.

* Similarity



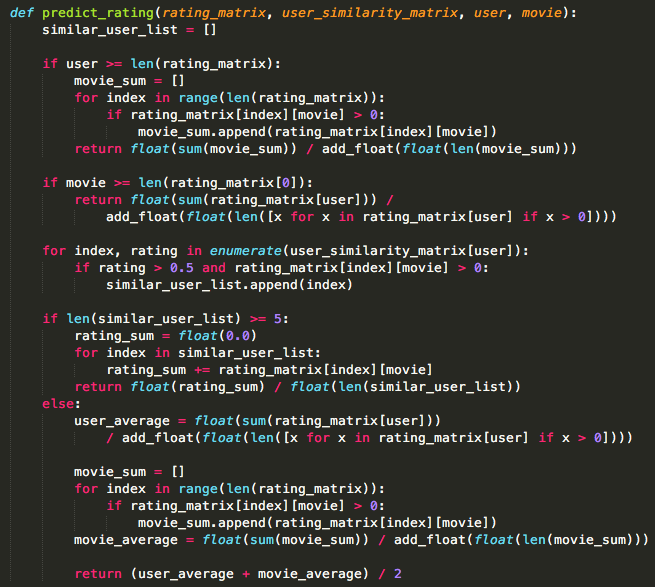
Similarity function calculates user similarity between each pair of all users. First it makes user\_average\_rating\_matrix to save the information about each user’s average rating information. This information will be used in calculating similarity.

Then, it calculates similarity between users based on movie rating. I refer to the following equation in the ppt slide:



and it makes user\_similarity\_matrix to save the similarity information.

* Predict\_rating



Predict\_rating function predicts rating for unrated movie for each user for recommend movies. Based on user\_similarity\_matrix which was made by similarity function, it draws the users who have 0.5 or more similarity and already have rated for the movie which we want to predict.

If the drawn similar users are more than 5 persons, the rating will be predicted as the average of the drawn similar users’ rating average for the target movie. If not, the rating will be predicted as the average of ‘target user’s average rating for the other movies’ and ‘other users’ average rating for the target movie’.

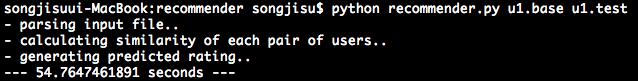
* Main



main python file handles input, test and output files and call the functions for processing. It writes prediction information to the output file.

1. How to Compile the Codes

I implemented this codes using Python 2.7. You can compile this code like this:



My main function’s name is recommender.py. I followed the rules of arguments that written on the notice file. It will take a little bit long time (about a minute) so please wait a minute till processing is over.