Read Me

The code is written in python 3.4.4

Suggest to run in PyCharm, but it will also work in command line.

Only standard python library is used. Math, collections, time, csv and multiprocessing.

The code is going to run mean rating, distance based and Resnick approach. The neighbourhood policy is set to no neighbourhood size(default 943, valid range 1-943) and 200 minimum overlap(default 1, valid range 1-1682). These two parameters can only change one each time. For changing neighbourhood size, the overlap should be 1. For changing minimum overlap, the neighbourhood size should be 943. The parameter for distance based approach can be changed at line 308 and 309. The parameter for Resnick approach can be changed at line 367 and 368.

It will take about 10min for the whole procedure.

The csv file will just generate in the same folder as the code file.

The assignment zip file contains:

- code ‘rs.py’
- report ‘RecSys Project Report.pdf’
- three of the experiment result ‘mean_rating.csv’, ‘rmse_distance_multi_n20.csv’ and ‘rmse_pearson_multi_n120.csv’
- details contains mean, median, standard deviation, min, max value of each user and item for task 1, ‘user_detail.csv’, ‘item_detail.csv’
- read me file