<Data Science>

Programming Assignment 4

2011003865 Park jun hyung

1. **Summary of algorithm**

Base dataset and Test dataset are given. In my program, read base dataset and make two double dictionaries. This is structure of dictionaries.

{

‘item\_id1’ : { ‘user\_id1’ : ‘rate’, ‘user\_id2’ : ‘rate’ …},

‘item\_id2’ : { ‘user\_id1’ : ‘rate’, ‘user\_id2’ : ‘rate’ …},

…

}

This dictionary is used when we know item id, then we want to find user and rate.

{

‘user\_id1’ : { ‘item\_id1’ : ‘rate’, ‘item\_id2’ : ‘rate’ …},

‘user\_id2’ : { ‘item\_id1’ : ‘rate’, ‘item\_id2’ : ‘rate’ …},

…

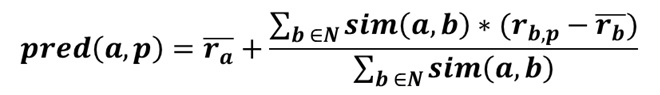
}

This dictionary is used when we know user id, then we want to find item and rate.

Using these dictionaries, program predict a rate that user might give. After making dictionaries, read test data. If test data is 1, 100 (user\_id, item\_id). It means that predict what is the rate of item 100 that user 1 given.

<Algorithm>

1. Find users who evaluate item 100 (using item-user dictionary)
2. Calculate similarity between user 1 and one of users who evaluate item 100 (using user-item dictionary)
3. Calculate prediction rate.



1. If prediction rate is improper, just use average rate of user x
2. **Detailed description of code**

In this program two functions are used.

def sim(x, y, x\_dict, y\_dict):

this function return similarity of user x and y.

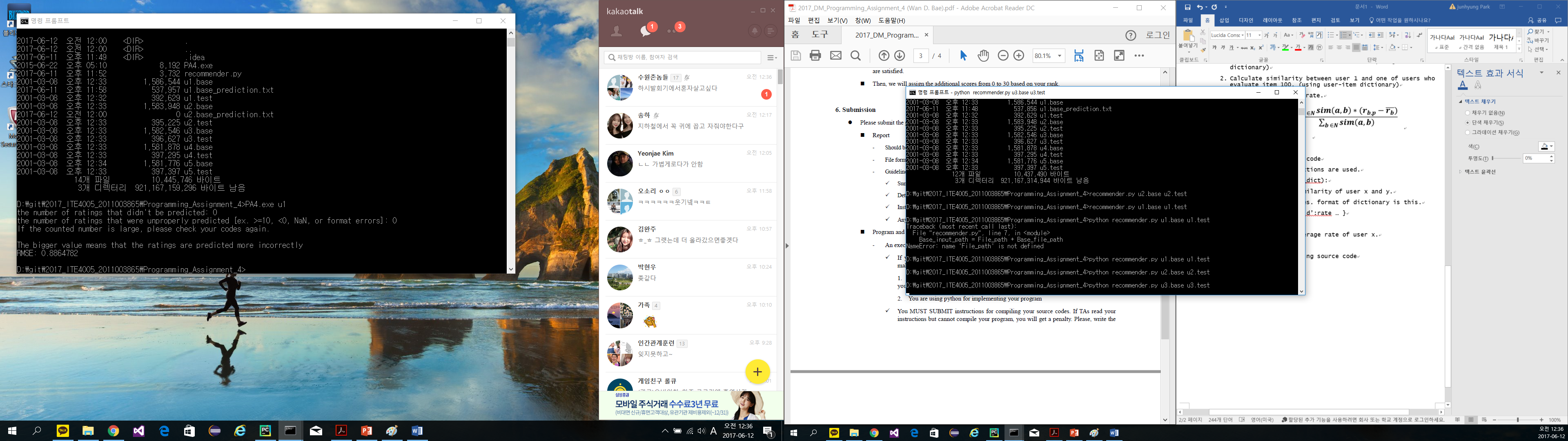
it needs two dictionaries. format of dictionary is this.

{‘item\_id’:rate, ‘item\_id’:rate … }

def average(x, x\_dict):

this function return average rate of user x.

1. **Instructions for compiling source code**



> python recommender.py u1.base u1.test