Data Science

Assignment #4

|  |  |
| --- | --- |
| Course name | Data science |
| Assignment | recommender |
| Major | Computer |
| Student ID | 2011004028 |
| Name | Seonghun Lee |

1. Summary of my algorithm

In this program, it can predict the item’s ratings which is not yet rated. I used collaborative filtering method and cosine similarity measure. First, this program make rating dictionary and according to this dictionary, find the rating value of test item set.

There are two main functions in this program. First, distance measure algorithm. I employ cosine similarity measure because it is recommended in recommender system and indeed it leads to better results than others measurement. The other function is predict rating function which is using collaborative filtering based algorithm and user based algorithm. This function is really important because it reflects all result and evaluation. It evaluate the average ratings of neighbors.

This program result RMSE is almost 9 in average of five test data.

1. Detailed description of my codes (for each functions)

|  |
| --- |
| Cosine similarity function |
| * This function evaluate cosine similarity value * I used this formula above |
| Predict rate function |
| * This function evaluate the average ratings from neighbor’s ratings |
| Neighbor\_set function |
| * This function find the neighbors from user (argument) * Return neighbor lists of user * Neighbor is decided from cosine similarity measure |
| Main function |
| * Input and output file format * Make watched array which contain watched items id for each user      * Rate\_dict is a dictionary which contains all of train data, it means all ratings values * Neighbor array show the neighbor of each user      * Predictset contain which user item pair should be predicted      * Maintain output format |

1. Instructions for compiling my codes at TA’s computer (e.g. screenshot)

|  |
| --- |
| 1. Put the recommender.py and u#.base, u#.test files into same folder |
|  |
| 1. Execute command prompt and go to this folder |
|  |
| 1. Execute recommender.py with command line arguments |
|  |
| 1. Check the output file in your folder |
|  |

1. Any other specification of my implementations and testing

|  |
| --- |
| Test result |
| * Result>> * RMSE #1 is 0.9033303 * RMSE #2 is 0.8998229 * RMSE #3 is 0.8968445 * RMSE #4 is 0.8962383 * RMSE #5 is 0.8992751 * Average RMSE is almost 0.9 |