

Implementation of Collaborative Filtering Based Recommender System and Experimenting With Tools Available

Group Members:

1. Kshitij Sharma 2012A7PS009H
2. Abhishek Kaushik 2012A7PS056H
3. Rohit Sharma 2012A7PS050H
4. Anudeep Reddy 2012A7PS175H

Part-1 (Implementation):

A user based recommender system has been implemented in C++ language. The input to the program is a file consisting of tuples of "User-Id, movie-Id, rating, timestamp". The algorithm calculates similarity between the inputted user ID and the user IDs in the dataset. The functions used to calculate similarity between users is calculated as follows:

$$sim(a, b) = \frac{\sum_{p \in P} (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P} (r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in P} (r_{b,p} - \bar{r}_b)^2}}$$

Where a and b are users and p is an item, here, a movie ID,

r_{ap} and r_{bp} are the ratings given the users a and b to item p respectively.

$$pred(a, p) = \bar{r}_a + \frac{\sum_{b \in N} sim(a, b) * (r_{b,p} - \bar{r}_b)}{\sum_{b \in N} sim(a, b)}$$

The results obtained are as follows:

```
C:\ Command Prompt
A:\Studies\4-1\Information Retrieval\Assignment\Assignment 2, 3\3\kshitij\Assignment 2>recommender < recInp
1. 1309 5.98711 1 3408 4 978300275
1. 814 5.51281 1 2355 5 978824291
1. 1536 5.31492 1 1197 3 978302268
1. 1467 5.12599 1 1287 5 978302039
1. 1599 5.02855 1 2804 5 978300719
1. 1500 5.01796 1 594 4 978302268
1. 1642 4.90047 1 919 4 978301368
1. 1449 4.88916 1 595 5 978824268
1. 1398 4.81163 1 938 4 978301752
1. 851 4.80138 1 2398 4 978302281
1. 1673 4.75735 1 2918 4 978302124
1. 1629 4.71631 1 1035 5 978301753
16 1 2791 4 978302188
17 1 2687 3 978824268
A:\Studies\4-1\Information Retrieval\Assignment\Assignment 2, 3\3\kshitij\Assignment 2>
19 1 3105 5 978301713
20 1 2797 4 978302039
21 1 2321 3 978302205
22 1 720 3 978300760
23 1 1270 5 978300055
24 1 527 5 978824195
25 1 2340 3 978300103
```

Part – 2 (Experimentation with a tool - GraphLab)

GraphLab is a machine learning package in Python. The dataset is given as input to 4 different algorithms within the package.

The algorithms used and the results are as follows:

1. Default Recommender

	A	B	C	D
1	user_id	item_id	score	rank
2	1	50	5.40625	1
3	1	127	5.22883	2
4	1	98	5.07388	3
5	1	100	5.01678	4
6	1	318	4.97387	5
7	1	64	4.96517	6
8	1	174	4.94754	7
9	1	181	4.92748	8
10	1	313	4.91444	9
11	1	172	4.91096	10
12	1	258	4.8363	11
13	1	56	4.83119	12

2. Item – Similarity Recommender

	A	B	C	D
1	user_id	item_id	score	rank
2	1	1118	5	1
3	1	325	5	2
4	1	1231	5	3
5	1	843	5	4
6	1	1435	5	5
7	1	791	5	6
8	1	725	5	7
9	1	537	5	8
10	1	335	5	9
11	1	1437	5	10
12	1	1617	5	11
13	1	891	5	12

3. Factorization Recommender

	A	B	C	D
1	user_id	item_id	score	rank
2	1	1449	5.23843	1
3	1	1467	4.84048	2
4	1	1500	4.79216	3
5	1	64	4.69894	4
6	1	1398	4.69031	5
7	1	318	4.68157	6
8	1	169	4.61619	7
9	1	483	4.59714	8
10	1	50	4.59122	9
11	1	12	4.58822	10
12	1	1064	4.58528	11
13	1	603	4.57712	12

4. Popularity Recommender

	A	B	C	D
1	user_id	item_id	score	rank
2	1	1653	5	1
3	1	1201	5	2
4	1	1293	5	3
5	1	1122	5	4
6	1	1536	5	5
7	1	1467	5	6
8	1	1189	5	7
9	1	814	5	8
10	1	1500	5	9
11	1	1599	5	10
12	1	1449	4.625	11
13	1	1642	4.5	12