Martyna Stasiak id 156071

# INFORMATION RETRIEVAL – SHORT EXERCISES V – COLLABORATIVE FILTERING AND ADWORDS

I. Given the below user-item rating matrix, predict rating of user U7 for item I4:

| | I1 | I2 | I3 | I4 | sim(U7,U·) |
|---|---|---|---|---|---|
| U1 | 5 | 4 | 4 | 4 | 0.0 |
| U2 | 5 | 3 | 7 | 3 | 1.0 |
| U3 | 4 | 3 | 2 | 3 | -0.5 |
| U4 | 6 | 4 | 5 | 4 | 0.5 |
| U5 | 3 | 4 | 2 | 4 | -1.0 |
| U6 | 4 | 3 | 5 | 3 | 1.0 |
| U7 | 4 | 3 | 5 | ? | |

Average

$\frac{5+3+7}{3} = 5$

$\frac{4+3+5}{3} = 4$

$\bar{r}_{U7} = \frac{4+3+5}{3} = 4$

a) Employ user-based CF with k=2 and either simple average or weighted average?

Answer: U7(I4) = $r_{u7,i4} = \frac{3+3}{2} = 3$

here simple average and weighted average give the same result since both k=2 users have the same similarity with user 7.

b) Employ user-based CF with k=2 and modify U7's average rating by the weighted modification of its nearest neighbors averages: $r_{a,i} = \bar{r}_a + \frac{\sum (r_{u,i} - \bar{r}_u) \cdot sim(a,u)}{\sum sim(a,u)}$

Answer: U7(I4) = $r_{u7,i4} = 4 + \frac{1 \cdot (3-5) + 1 \cdot (3-4)}{2} = 4 + \frac{-2-1}{2} = 4 - \frac{3}{2} = \frac{5}{2} = 2{,}5$

c) Which item should be analyzed to predict the rating when using item-based CF with k=1? What would be the predicted rating?

they have the same ratings user-wise, so they have the highest possible cosine similarity = 1.

Answer: item - 12 and prediction 3

II. Four advertisers A, B, C, and D with a daily budget of $2 bid for the following keywords ($1 each): A: w, x; B: x, z; C: x, y; D: y, z. Use a simplified version of BALANCE to select the ads for the following query stream (in the case of a tie use the following order for breaking it A > B > C > D):

| query stream | x | y | w | z | z | w | y | x |
|---|---|---|---|---|---|---|---|---|
| BALANCE | A | C | A | ? B | ? D | ? — | ? C | ? B |
| | ~~A:2~~ | ~~A:1~~ | A:1 | B:2 | B:1 | ~~B:1~~ | ~~B:1~~ | B:1 |
| | B:2 | ~~B:2~~ | ~~B:2~~ | ~~C:1~~ | ~~C:1~~ | ~~C:1~~ | C:1 | ~~D:1~~ |
| | C:2 | C:2 | ~~C:1~~ | D:2 | D:2 | ~~D:1~~ | D:1 | |
| | ~~D:2~~ | D:2 | ~~D:2~~ | | | | | |

no possible option using BALANCE here

A: w, x
B: x, z
C: x, y
D: y, z

competitive ratio = 7/8 -> 87,5%