

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	$sim(U7, U \cdot)$	Average
U1	5	4	4	4	0.0	$(5 + 4 + 4) / 3 = 4.3(3)$
U2	5	3	7	3	1.0	5
U3	4	3	2	3	-0.5	3
U4	6	4	5	4	0.5	5
U5	3	4	2	4	-1.0	3
U6	4	3	5	3	1.0	4
U7	4	3	5	?		4

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

Answer: $U7(I4) = (3 + 3) / 2 = 3$

b) Employ user-based CF with $k=2$ and modify U7's average rating by the weighted modification of its nearest neighbors averages:

Answer: $U7(I4) = 4 + \frac{1.0 * (3 - 4) + 1.0 * (3 - 5)}{1.0 + 1.0} = 4 - 1.5 = 2.5$

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

Answer: item - I2 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	? C