

Overview

- assuming users A and B are similar
 - similarity score needs to be above threshold
- recommend items which A liked to B

User-based collaborative filtering

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

Idea:

- If users have rated items similarly in the past, their predications are likely to be similar in the future
- Find users who are similar to Alice in terms of which items they like
- Predict Alice's future rating of new item based on ratings of similar users (use a threshold for identifying similar users)

Idea: Similar users rate items similarly.

User3	3	3	1	5	4
User4	1	5	5	2	1

Transfer knowledge on a new item from similar users to U_0

Therefore:

- Find users who are similar to U_0 (Alice) in terms of which items they like
 - TODO: compute pairwise similarities between Alice and all other users
- Predict U_0 's (Alice) future rating of new item based on ratings of similar users (use a threshold for identifying similar users)
 - TODO: predict how U_0 (Alice) will rate the new item.
 - This prediction is used to decide on whether item is recommended or not, in ranking recommender results, or for some other system reaction.

Similarity Score

- cosine similarity of user vectors
 - does not account for different user rating tendencies
 - * some easily 10/10, some 8/10 at max
- cosine similarity of centered user vectors
 - normalize user ratings by each user's average rating value

a, b : users

$r_{a,p}$: rating of user a for item p

\bar{r}_a : average rating of user a across P

P : set of items, rated both by a and b

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

Possible similarity values between -1 and 1

Interpretation of $\text{sim}(a,b)$

- **Pearson correlation** - Correlation of two variables a,b
- Cosine of angle between two centered vectors a,b

Prediction

- Common prediction function for user-based collaborative filtering

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

Idea:

- Set of most similar users (neighbours) N
 - Combine their deviation of ratings for the new item in comparison to their average ratings
 - ... with their similarity to user a
 - ... and add/subtract this value from user a 's average rating.
- fine tuning via
 - more similarity if users agree on controversial items
 - * controversial if high variance in ratings
 - * more weight of those items
 - more weight to ratings of similar users