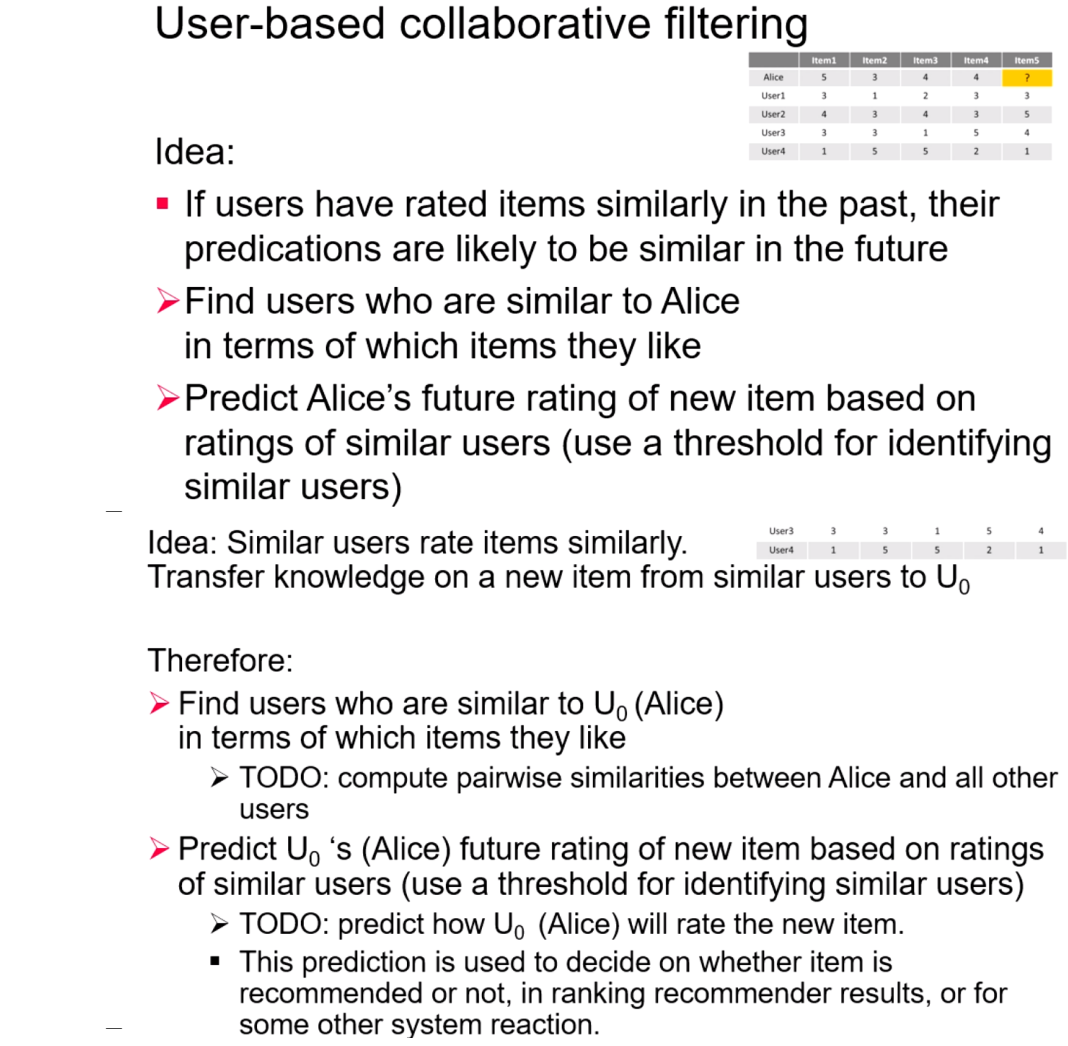
Overview

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



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

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

Possible similarity values between -1 and 1

Interpretation of 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

$$pred(a,p) = \overline{r_a} + \frac{\sum_{b \in N} sim(a,b) * (r_{b,p} - \overline{r_b})}{\sum_{b \in N} sim(a,b)}$$

Idea:

- Set of most similar users (neighbours) N
- Combine their deviation of ratings for the new item in comparison to their average ratings
- … with the 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