Steps for User Based Collaborative Filtering – An Example

User-based nearest-neighbor collaborative filtering

- Example
  - A database of ratings of the current user, Alice, and some other users is given:

	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

  - Determine whether Alice will like or dislike *Item5*, which Alice has not yet rated or seen

Measuring user similarity

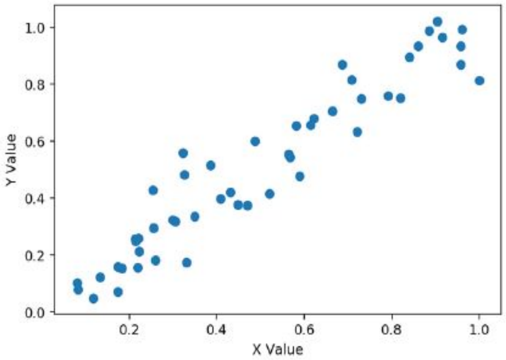
A popular similarity measure in user-based CF: **Pearson correlation**

$$r(x,y) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})}}$$

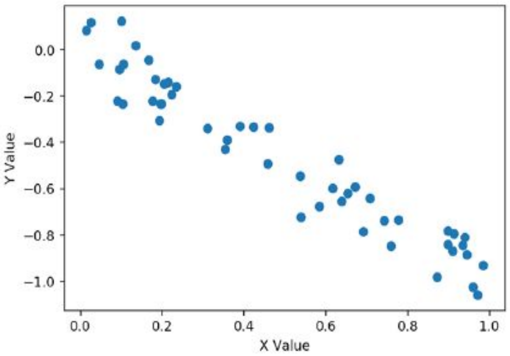
- r represents the Pearson correlation value between 2 numerical arrays x and y
- Value between +1 and -1
- Strength of Linear Relationship

Analytics Vidhya

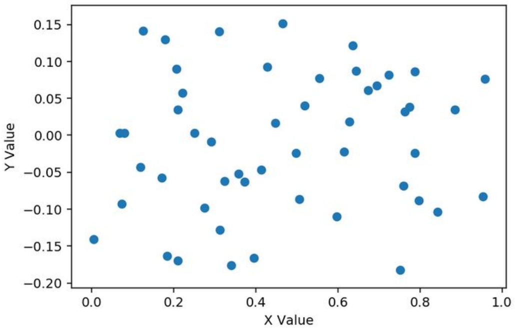
Measuring user similarity



Positive Correlation
(0 < r < 1)



Negative Correlation
(-1 < r < 0)



Zero Correlation
(r = 0)

Analytics Vidhya

Measuring user similarity

- Pearson Correlation between User Rating arrays

a, b : users

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

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

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

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

Measuring user similarity

- A popular similarity measure in user-based CF: **Pearson correlation**

, b : users

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

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

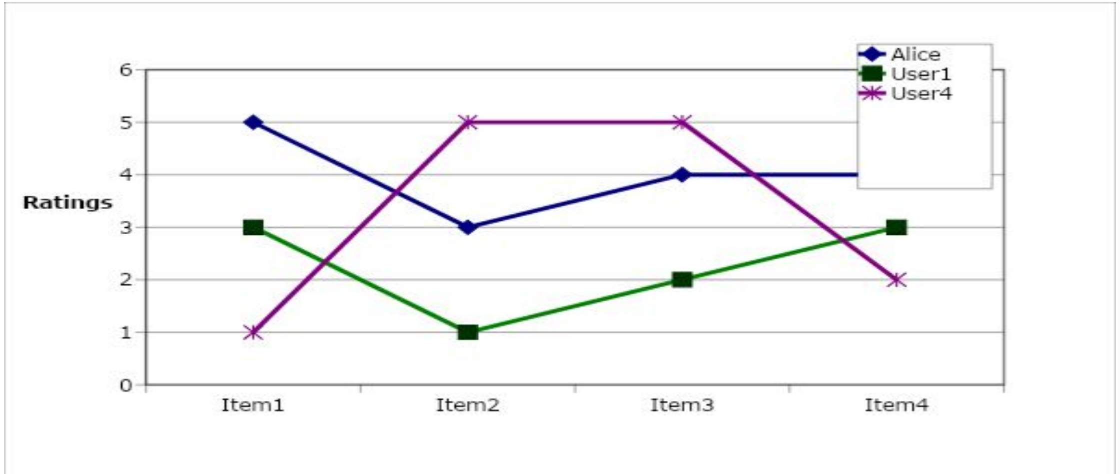
P : set of items, rated both by a and b

	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

sim = 0.85
sim = 0.70
sim = 0.00
sim = -0.79

Measuring user similarity

- Takes differences in rating behavior into account



Choosing Neighbourhood size

	Item1	Item2	Item3	Item4	Item5	
Alice	5	3	4	4	?	
User1	3	1	2	3	3	sim = 0.85
User2	4	3	4	3	5	sim = 0.70
User3	3	3	1	5	4	sim = 0.00
User4	1	5	5	2	1	sim = -0.79

Choosing Neighbors & Making predictions

- A common prediction function:

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

$$pred(Alice, Item\ 5) = \overline{r_{alice}} + \frac{\sum_{b \in N} sim(Alice,b) * (r_{b,p} - \bar{r_b})}{\sum_{b \in N} sim(Alice,b)}$$

$$N = \{User\ 1, User\ 2\}$$

Analytics Vidhya

Choosing Neighbors & Making predictions

	Item1	Item2	Item3	Item4	Item5	
Alice	5	3	4	4	?	
User1	3	1	2	3	3	sim = 0.85
User2	4	3	4	3	5	sim = 0.70
User3	3	3	1	5	4	sim = 0.00
User4	1	5	5	2	1	sim = -0.79

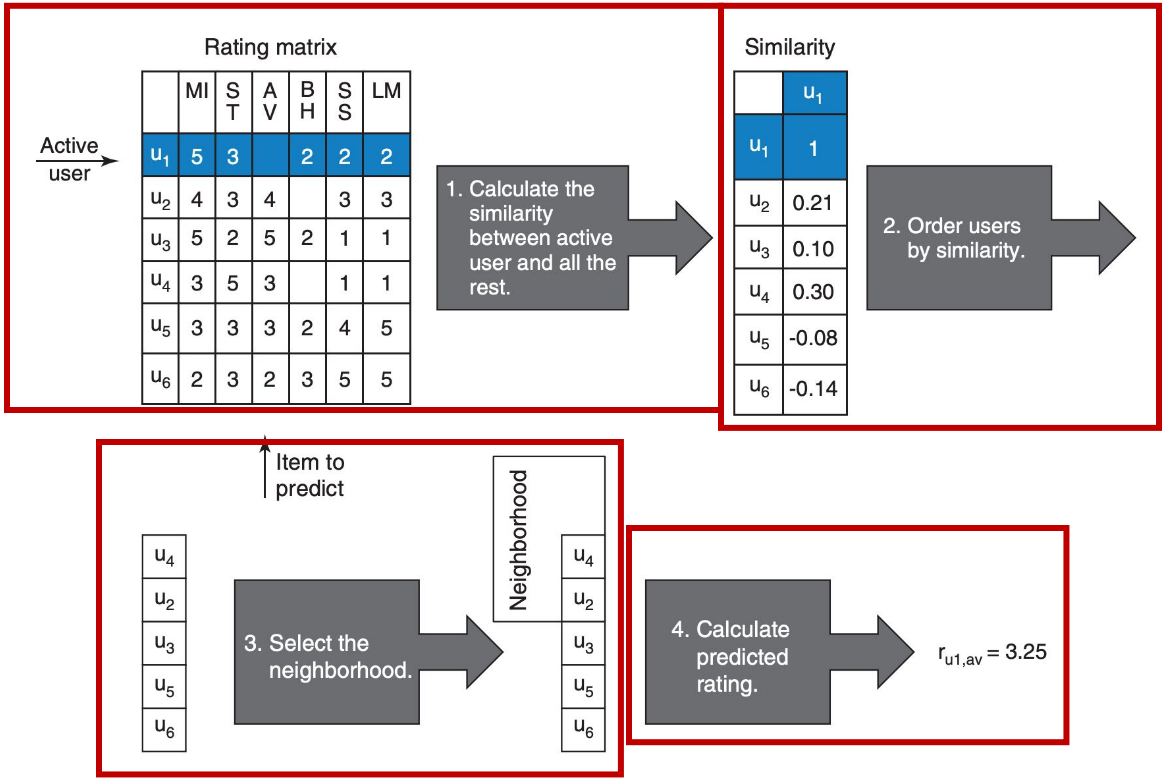
$$pred(Alice, Item\ 5) = \overline{r_{alice}} + \frac{sim(Alice, User\ 1) * \left(r_{User\ 1,item5} - \overline{r_{User\ 1}}\right) + sim(Alice, User\ 2) * \left(r_{User\ 2,item5} - \overline{r_{User\ 2}}\right)}{sim(Alice, User\ 1) + sim(Alice, User\ 2)}$$

$$pred(Alice, Item\ 5) = 3.75 + \frac{0.85 * (3 - 2.4) + 0.70 * (5 - 3.8)}{0.85 + 0.7}$$

$$pred(Alice, Item\ 5) = 4.621$$

Analytics Vidhya

Steps for User-User CF



Analytics Vidhya

Design Decisions

- Neighbourhood Size
    - All users?
    - K users most similar to active user u
- Similarity Function
    - Pearson Correlation
    - Cosine Similarity
    - Spearman Rank Correlation
- Averaging Function
    - Weighted average
    - Simple Average
    - Regression

Improving the metrics/prediction function

- Not all neighbor ratings might be equally "valuable"
  - Agreement on commonly liked items is not so informative as agreement on controversial items
  - **Possible solution**: Give more weight to items that have a higher variance
- Value of number of co-rated items
  - Use "significance weighting", by e.g., linearly reducing the weight when the number of co-rated items is low
- Case amplification
  - Intuition: Give more weight to "very similar" neighbors, i.e., where the similarity value is close to 1.
- Neighborhood selection
  - Use similarity threshold or fixed number of neighbors