

# User Collaborative Filtering (CF)

Tuesday, February 27, 2024 1:25 PM

## Rationale behind User CF:

- ① users  $u_1$  and  $u_2$  have similar interest.
- ②  $u_1$  clicked item  $i_1$  and like it
- ③  $u_2$  has NOT been exposed by  $i_1$
- ④ recommend  $i_1$  to  $u_2$

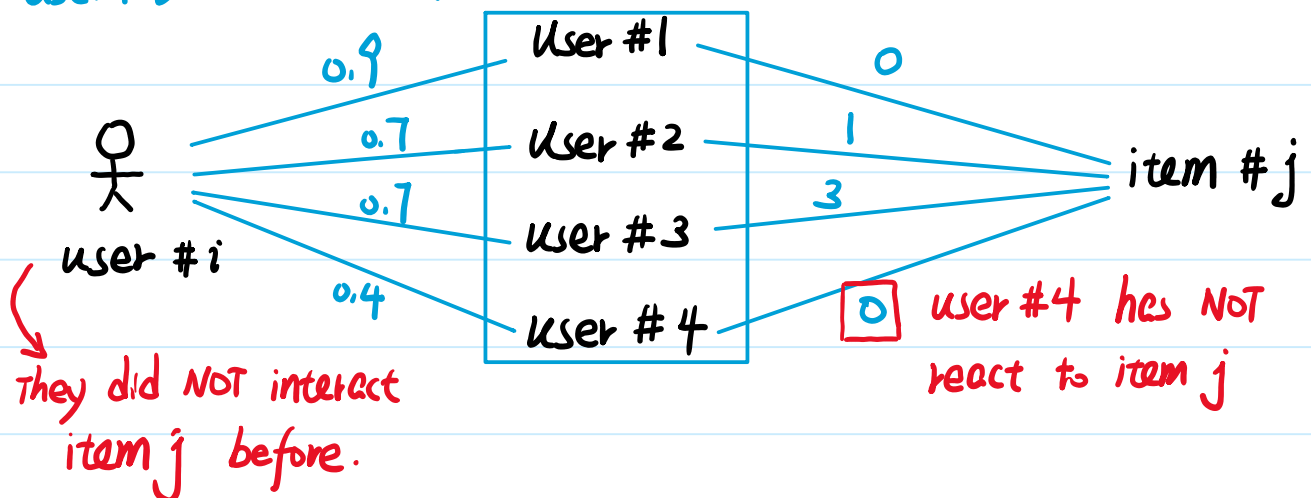
## How to know two users share similar interest?

- ① Their "likes" "collections" "share" lists overlap.
- ② Their following lists overlap a lot.

## User CF implement:

How similar between user #i and others?

How users react to item j



Estimated interest from user #i to item #j:

$$\sum \text{similarity} \cdot \sum \text{like} = 0.9 \times 0 + \dots + 0.4 \times 0 = 2.8$$

## User CF

Tuesday, February 27, 2024 1:25 PM

similarity between users:

- ①  $J_1$ : set of items liked by user  $u_1$
- ②  $J_2$ : set of items liked by user  $u_2$
- ③  $I = J_1 \cap J_2$ : intersection
- ④ similarity:

$$\text{sim}(u_1, u_2) = \frac{|I|}{\sqrt{|J_1|} \cdot \sqrt{|J_2|}} \in [0, 1]$$

→ # of items liked by both users.

limitation:

Every user loves the "popular" item.  
such item cannot help to identify if two users have similar interest.

we should reduce weight of "popular" item

Example:

A student and a professor may both love watching "the Lord of Ring"; but one may love sci-fi but the other likes comedy.

→ a "hot" and popular movie

# User CF

Tuesday, February 27, 2024 1:25 PM

Improved similarity:

$$\text{sim}(u_1, u_2) = \frac{\sum_{l \in I} \frac{1}{\log(1+n_l)}}{\sqrt{|J_1|} \cdot \sqrt{|J_2|}}$$

# of users who like item  $l$ .

larger  $n_l$   
→ smaller weight

Offline calculation:

- ① index from "user → item" (same as item CF)
- ② index from "user → user"

user #  $i$  →

user #1: 0.7
user #2: 0.6
⋮
user # $j$ : 0.4

top  $k$  users who  
are similar to  
user #  $i$

Online recommend:

- ① Given user ID: get top  $k$  users from "user → user" index
- ② For each user in top  $k$ : get last  $N$  items from "user → item" index
- ③ Get like score for  $N \cdot k$  items  
use equation  $\sum \text{sim} \cdot \sum \text{like}$
- ④ Return top 100 items

this is from one retrieval channels;  $M$  channels in parallel.