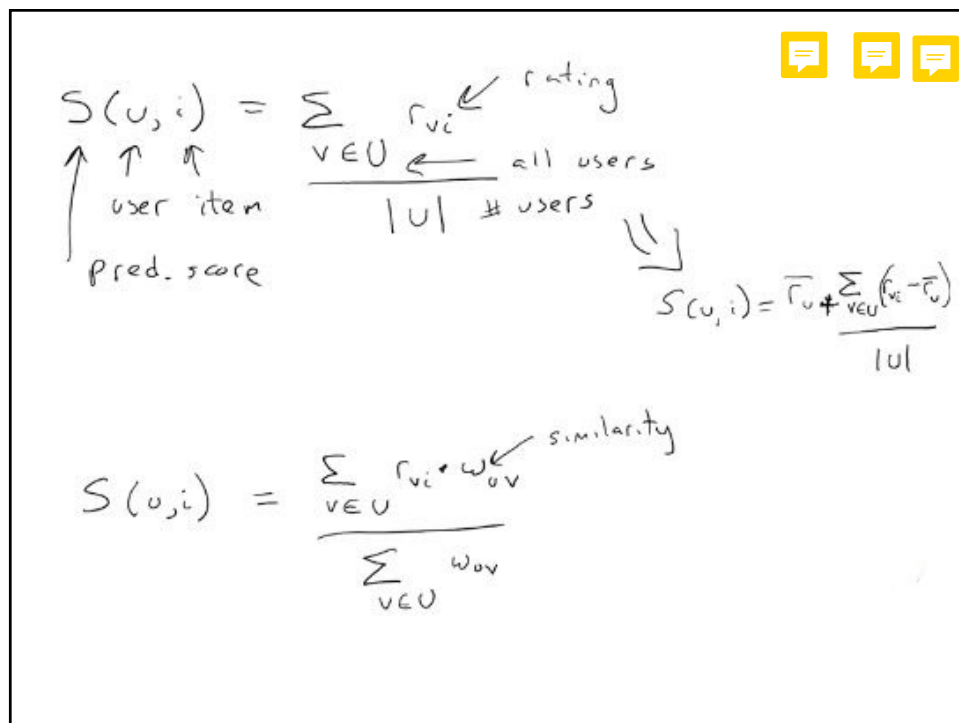


## User-User Collaborative Filtering

### Reference for UUCF

- An Algorithmic Framework for Collaborative Filtering by Herlocker, Konstan, Borchers, and Riedl (Proc. SIGIR 1999)

The following are screenshots the handwritten notes



Handwritten notes showing two formulas for predicting a score  $S(u, i)$ .

**Formula 1:**

$$S(u, i) = \frac{\sum_{v \in U} r_{vi}}{|U|}$$

Annotations for Formula 1:

- $S(u, i)$ : pred. score
- $u$ : user
- $i$ : item
- $r_{vi}$ : rating
- $v \in U$ : all users
- $|U|$ : # users

**Formula 2:**

$$S(u, i) = \bar{r}_u + \frac{\sum_{v \in U} (r_{vi} - \bar{r}_u)}{|U|}$$

**Formula 3:**

$$S(u, i) = \frac{\sum_{v \in U} r_{vi} \cdot w_{uv}}{\sum_{v \in U} w_{uv}}$$

Annotation for Formula 3:

- $w_{uv}$ : similarity

$$S(u,i) = \bar{r}_u + \frac{\sum_{v \in U} (r_{vi} - \bar{r}_v) * w_{uv}}{\sum_{v \in U} w_{uv}}$$

$u \neq v$   
 limit size  
 min similarity Reg.

Pearson Correl

$$w_{uv} = \frac{\sum_{i \in I} (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sigma_u \sigma_v}$$

• small overlap  
 • binary

$$\frac{\sum_{v \in U}}{\sum_{v \in U}}$$

## Common Characteristics

- Collection of Ratings
- Measure of Inter-User Agreement
  - Correlation, Vector Cosine
- Personalized Recommendations/Predictions
  - Weighted Combinations of Others' Ratings
- Tweaks to make things work right ...
  - Neighborhood limitations
  - Normalization
  - Dealing with limited co-ratings

## Let's Formalize This ...

- Given a set of items  $I$ , and a set of users  $U$ , and a sparse matrix of ratings  $R$ ,

We compute the prediction  $s(u,i)$  as follows:

- For all users  $v \neq u$ , compute  $w_{uv}$ 
  - similarity metric (e.g., Pearson correlation)
- Select a neighborhood of users  $V \subset U$  with highest  $w_{uv}$ 
  - may limit neighborhood to top-k neighbors
  - may limit neighborhood to  $\text{sim} > \text{sim\_threshold}$
  - may use  $\text{sim}$  or  $|\text{sim}|$  (risks of negative correlations)
  - may limit neighborhood to people who rated  $i$  (if single-use)
- Compute prediction:

$$s(u, i) = \bar{r}_u + \frac{\sum_{v \in V} (r_{vi} - \bar{r}_v) * w_{uv}}{\sum_{v \in V} w_{uv}}$$

## Implementation Issues

- Given  $m = |U|$  users and  $n = |I|$  items:
  - Computation can be a Bottleneck
    - Correlation between two users is  $O(n)$
    - All correlations for a user is  $O(mn)$
    - All pairwise correlations is  $O(m^2n)$
    - Recommendations at least  $O(mn)$
  - Lots of ways to make more practical
    - More persistent neighborhoods ( $m \rightarrow k$ )
    - Cached or incremental correlations

## Core Assumptions/Limitations

- Why does this work?
  - Let's break it down ...
- Assumption: Our past agreement predicts our future agreement
  - Base Assumption #1: Our tastes are either individually stable or move in sync with each other
  - Base Assumption #2: Our system is scoped within a domain of agreement

## User-User Collaborative Filtering