

## *User-Based Collaborative Filtering—*

User-Based Collaborative Filtering uses the assumption that similar people will have similar taste, and recommend items by finding similar users to the active user. A specific application of this is the user-based Nearest Neighbor algorithm. This algorithm needs two tasks"

1. Find the K-nearest neighbors (KNN) to the user  $a$ , use a similarity function  $w$  to measure the distance between each pair of users:

$$\text{Similarity}(a, i) = w(a, i) \quad i \in K$$

2. Predict the rating that user  $a$  will give to all items the  $k$  neighbors have consumed but  $a$  has not. We look for the item  $j$  with the best predicted rating.

In other words, we are creating a User-Item Matrix, predicting the ratings on items the active user has not see, based on the other similar users. This technique is also called memory-based.

Pros:

- Easy to implement
- Context independent
- More accurate when compared to other techniques, such as content-based.

Cons:

- Sparsity: the percentage of people who rate items is really low
- Scalability: the more  $K$  neighbors we consider (under a certain threshold), the better my classification should be. Nevertheless, the more users there are in the system, the greater the cost of finding the nearest K neighbors will be.
- Cold-start: new users will have no to little information about them to be compared with other users
- New item: new items will lack of ratings to create a solid ranking

## *Item-Based Collaborative Filtering—*

This method can also be divided into two sub tasks:

1. Calculate similarity among the items

- Cosine-Based Similarity
- Correlation-Based Similarity
- Adjusted Cosine Similarity
- I-Jaccard distance

2. Calculation of Prediction

- Weighted Sum
- Regression

The difference between User-Based and Item-Based is, in this cases, we directly pre-calculate the similarity between the co-rated items, skipping K-neighborhood search.