

Scalability Challenge for User Based Collaborative Filtering

Scalability Challenge

- Scalability - Nearest neighbor require computation that grows with both the number of users and the number of items.

Diagram illustrating the scalability challenge in nearest neighbor computation. The matrix represents the relationship between users and items.

The matrix is labeled with n_{users} on the vertical axis and n_{items} on the horizontal axis.

The matrix structure is as follows:

	item_1	item_2	...	item_19	item_20
user_1	-	5	...	-	1
user_2	4	-	...	-	-
...
...
user_57	-	5	...	2	4
...
...
...
...
...

The matrix size is indicated as $n_{\text{users}} \times n_{\text{items}}$.

Scalability Challenge

- As user rates new items:
 - Rating vector of user will change
 - As a result, similarity with other users will also change
- Finding similar users in advance is difficult!
- Most user-based CF methods find neighbors when predictions are needed
- Faces issues of scalability

Item-based Collaborative Filtering

- Item-based filtering does not solve the scalability problem itself
- Pre-processing approach by Large Scale E-Commerce
 - Calculate all pair-wise item similarities in advance
 - Item similarities are supposed to be more stable than user similarities
- Scalability with respect to memory
 - Up to N^2 pair-wise similarities to be memorized (N = number of items) in theory
 - In practice, this is significantly lower (items with no co-ratings)
 - Further reductions are possible by limiting the neighborhood size