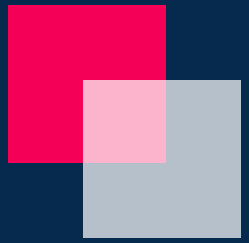


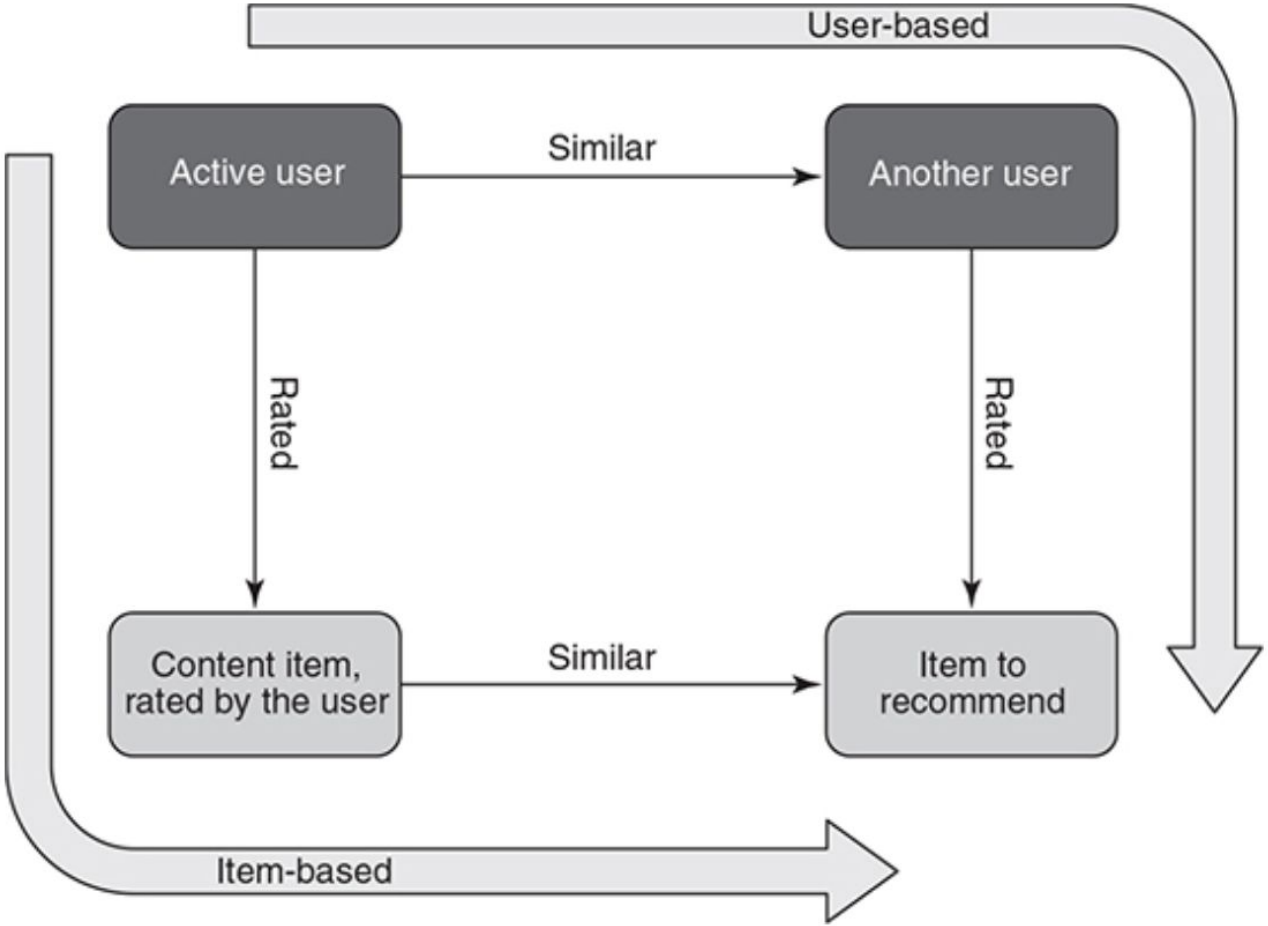
Collaborative Filtering

What is Collaborative Filtering

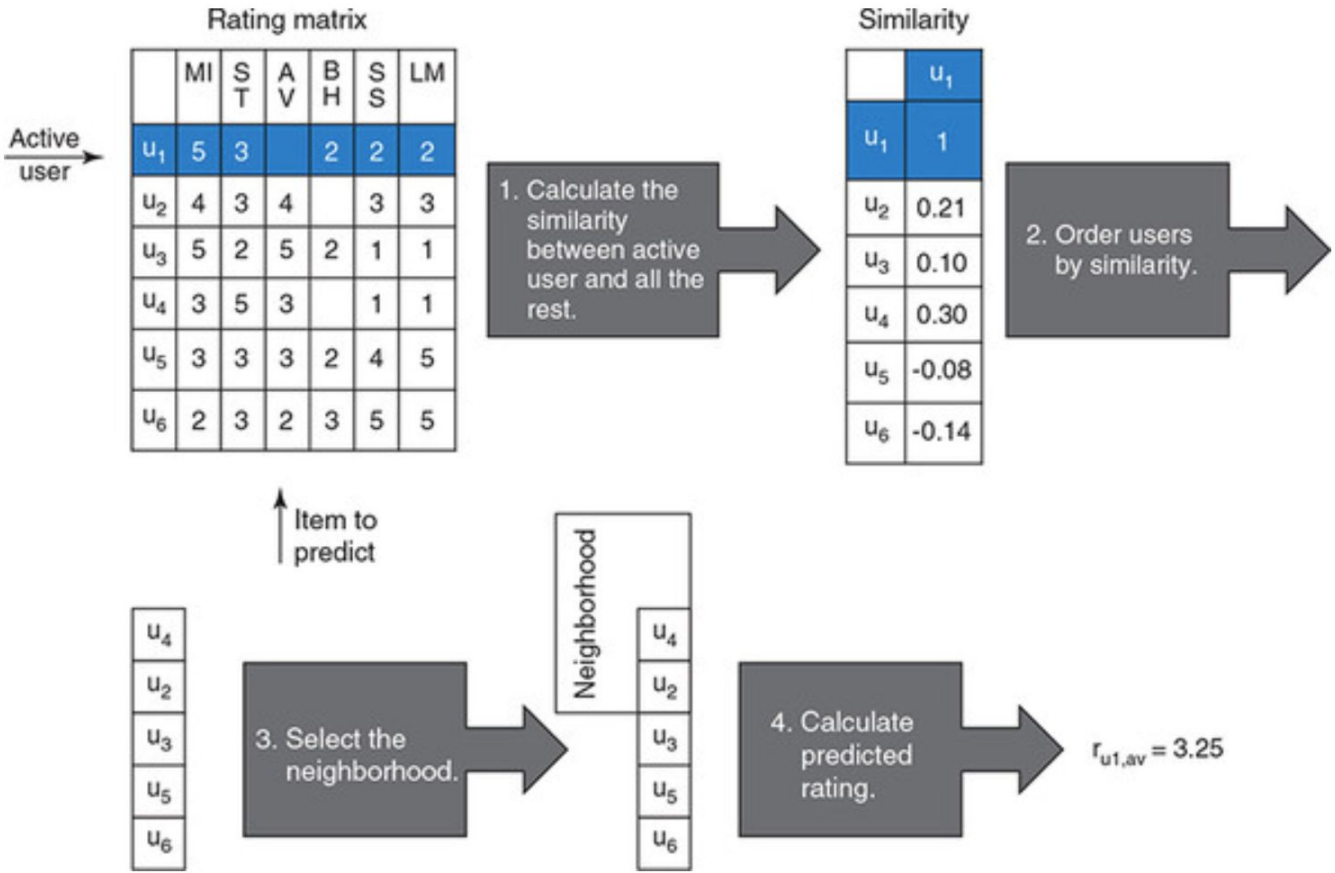
- Recommendations is based on people who like same things as you, but who also like something that you haven’t yet consumed

Two ways to do Collaborative Filtering

- User based filtering
- Item based filtering
- Both the methods are calculated based on the ratings


User-based
Active user
Similar
Another user
Rated
Rated
Content item, rated by the user
Similar
Item to recommend
Item-based

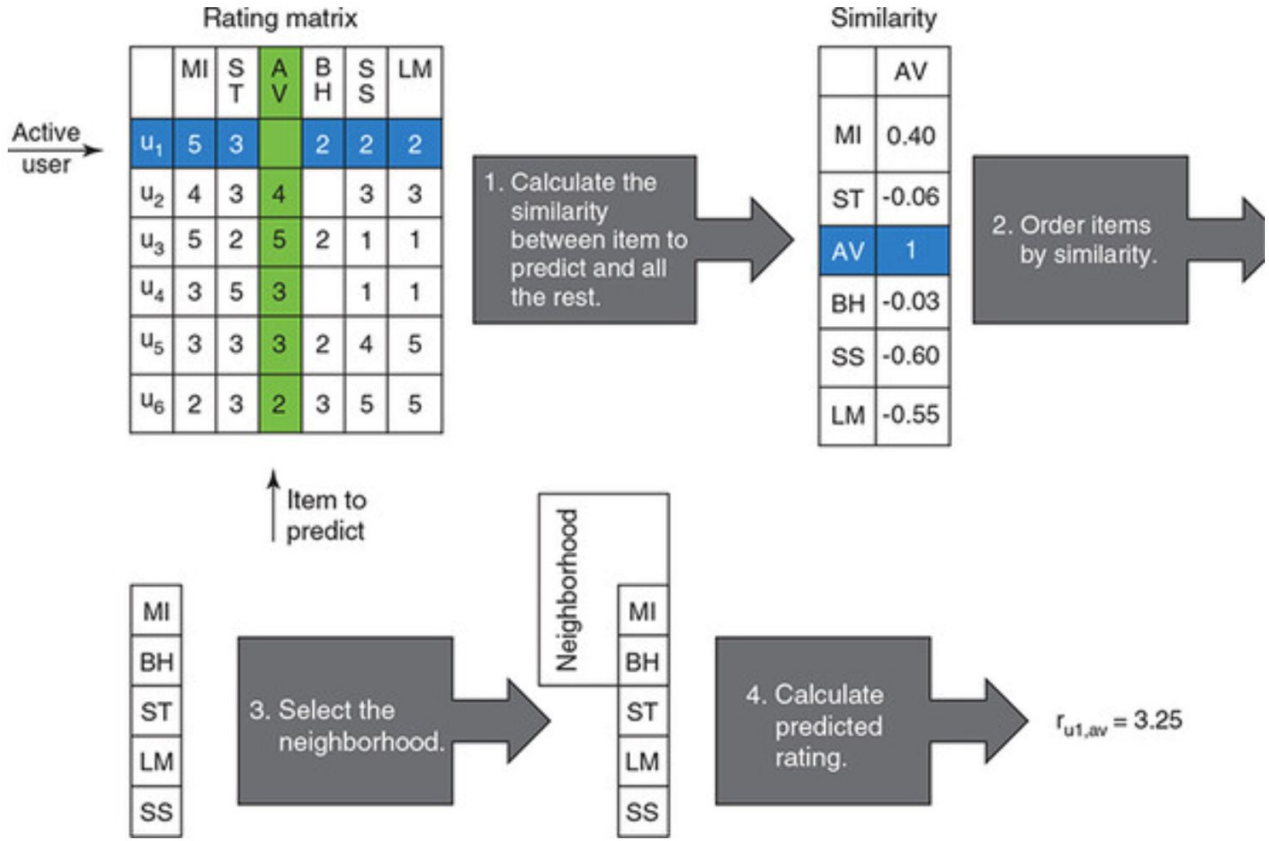
User-based filtering



Challenges of User based Collaborative filtering

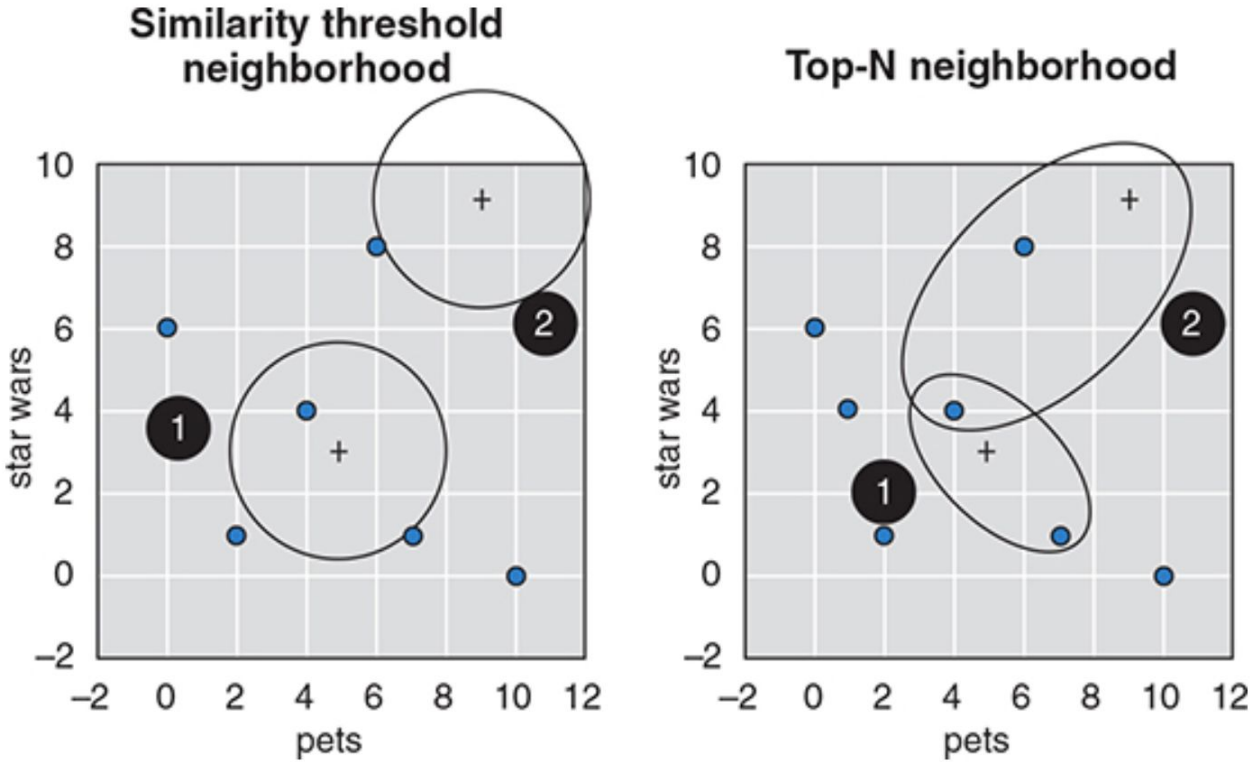
- If users did not rate any content, then there will be no recommendations
- Users who don’t have overlapping tastes with other users won’t receive good recommendations.

Item based filtering



Ways to select Neighborhood

- Clustering - Threshold based or Top-N



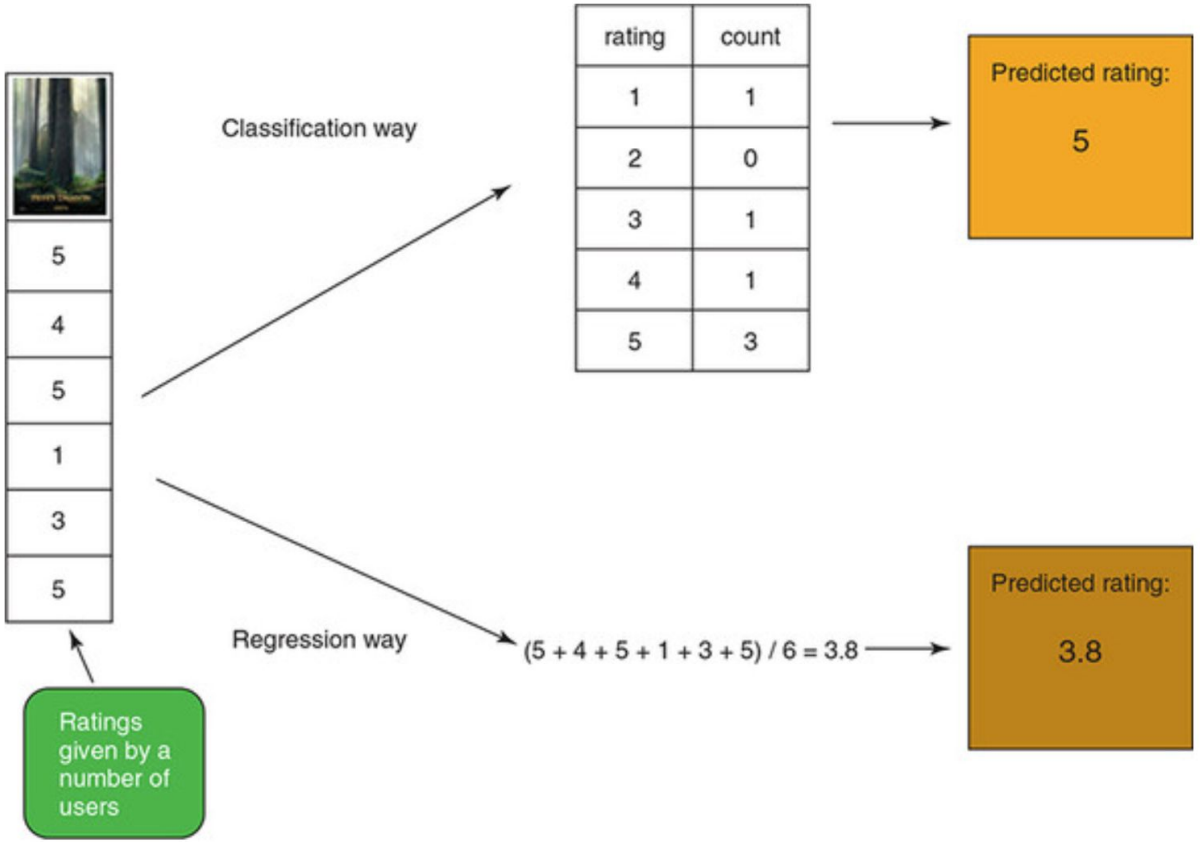
User-user or item-item collaborative filtering?

- User Rating - Adding a new rating can change the system's calculation of user's taste
- Unwise to pre-compute which users are similar
- An average user doesn't have many item ratings
- Items are more stable
- Studies shown that we can pre-calculate similarities for items
- Similar items won't provide the serendipity that similar (user) ratings can provide
- It is easier to explain item based recommendations

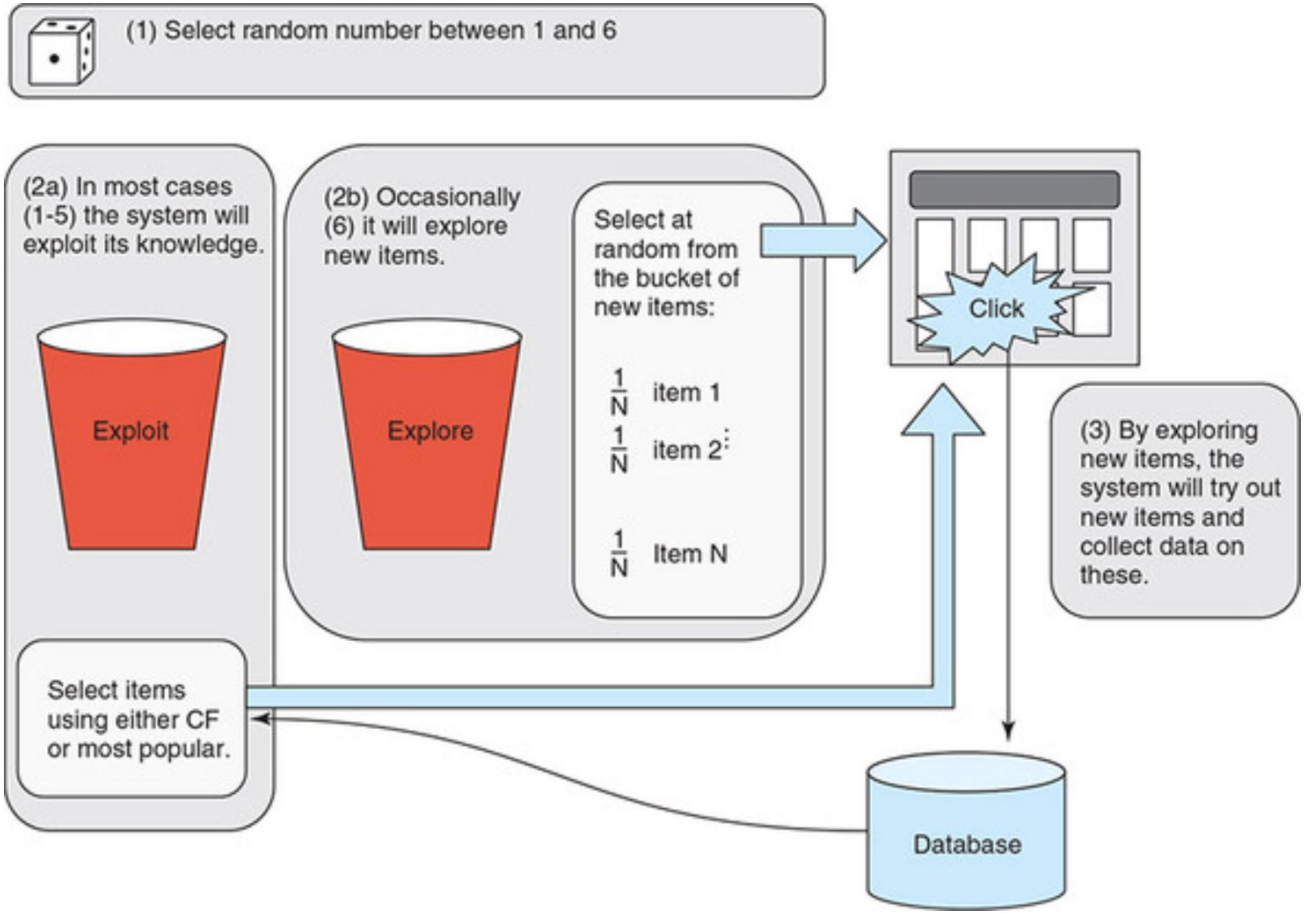
If you’ve many more users than items, then you should go for item based filtering; otherwise, user based filtering is more economical.

Choosing between Top-*N* and threshold is choosing between quantity and quality. Choose the threshold method for quality; Top-*N* for quantity.

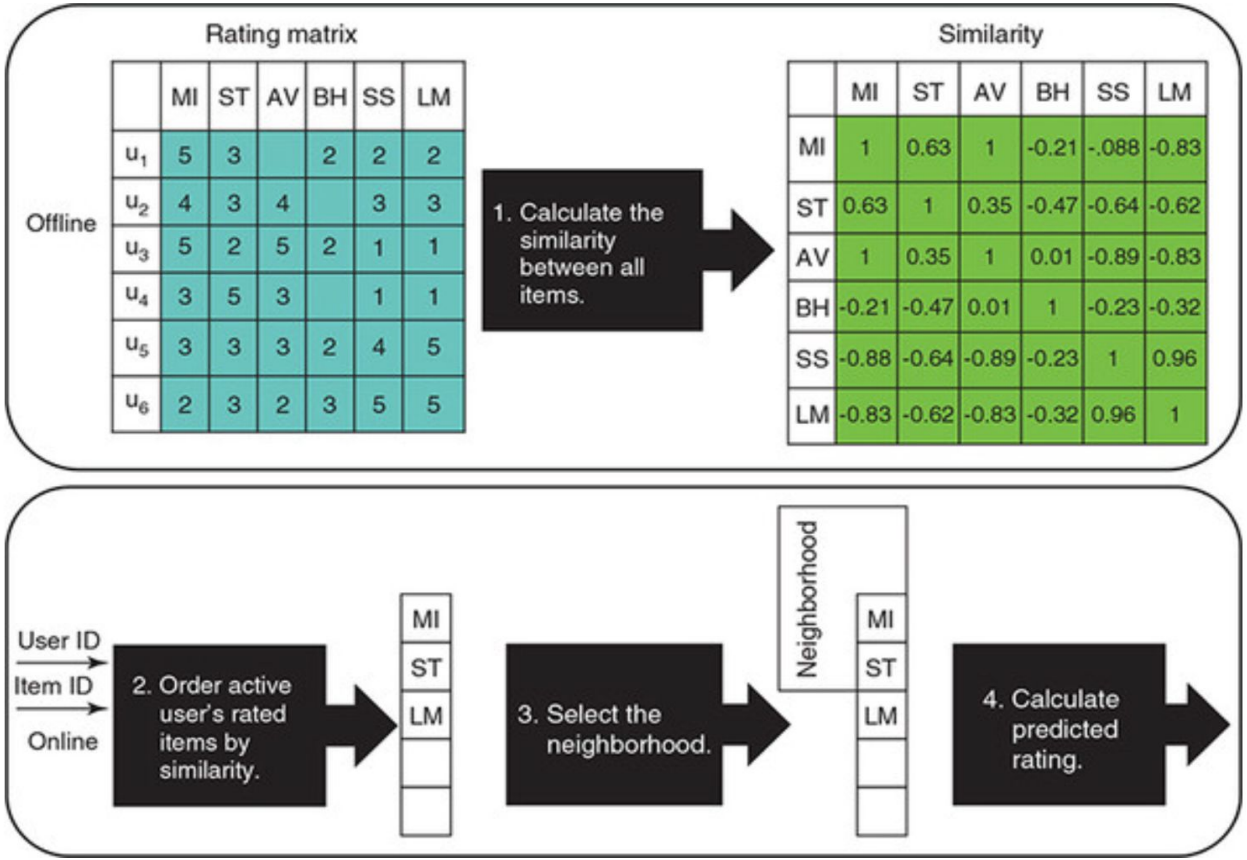
Ways to calculate Predicted ratings



Handling cold start



Offline Vs Online



Pros and Cons

—

- Sparsity
- Gray sheep
- Number of ratings
- Recommend popular items often
- No need of user and item metadata

Summary

- The pipeline of neighborhood filtering can either use user-based filtering, looking at similar users, or item-based filtering, looking at similar items.
- Use user-based filtering if there are more items than users; otherwise, use item-based filtering.
- A similarity matrix makes it possible to quickly look up similar items.
- Using a similarity table enables the system to make neighborhoods using the clustering, Top-N, or threshold procedures.
- The neighborhoods you find let you calculate predictions when you've a small set of similar users.
- Amazon's first stab at a recommender system was item-based collaborative filtering.