Recommender Systems

Recommendation system is an information filtering technique, which provides users with information, which he/she may be interested in.

Examples:

"Getting Information off the internet is like taking a drink from a fire hydrant" - Mitchell Kapor

- Information Overload
- User Experience
- Revenues

Recommender systems help in addressing the information overload problem by retrieving the information desired by the user based on his or similar users' preferences and interests.

In General, two types of recommender system.

1. Personalized

Registered customers

John visits an online store to buy an accessory for his Blackberry and conducts an online product search. Since John is a registered customer, the recommendations engine draws up information of his previous purchases – a Blackberry phone bought a few months ago – as soon as he logs in and queries for accessories. The engine also draws up in-store trends and recognizes that a majority of customers that buy the same/ similar Blackberry also purchase a particular kind of accessory. The engine collates information about John and the collective and recommends the same accessory to John. An online store with advanced social commerce features can also display comments based on the buying behavior of John's social circle such as 'Your friend George bought this' or 'Your friend Jane reviewed this product' to influence his decision. If John adds the accessory to his online shopping cart, the engine will continue to offer real-time recommendations of products that complement his Blackberry and/or the new accessory. Thus the engine is constantly aware of John's digital actions and refines its recommendations to suit him.

2. Non-Personalized

New customers

Taking the context of the above example let us say John is a new visitor to the online store, seeking to make the same type of purchase. Despite having no information about John the engine can offer recommendations about collective preferences in the form of 'Best Sellers'. As John begins to browse a few pages, the engine determines John's preferences and leverages this information to offer recommendations that may interest him.

- Personalized

- Non-Personalized

Data Acquisition

1. Explicit Data
    - Customer Ratings
    - Feedback
    - Demographics
    - Physiographics
    - Ephemeral Needs

2. Implicit Data
    - Purchase History
    - Click or Browse History

3. Product Information
    - Product Taxonomy
    - Product Attributes
    - Product Descriptions

Techniques of Recommendation System

1. **Collaborative Filtering** method finds a subset of users who have similar tastes and preferences to the target user and use this subset for offering recommendations.

Basic Assumptions :

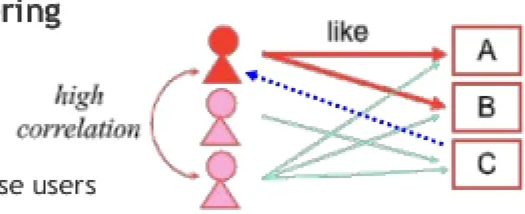
- Users with similar interests have common preferences.
- Sufficiently large number of user preferences are available.

Main Approaches :

- User Based
- Item Based

User-Based Collaborative Filtering

- Use user-item rating matrix
- Make user-to-user correlations
- Find highly correlated users
- Recommend items preferred by those users



Pearson Correlation :

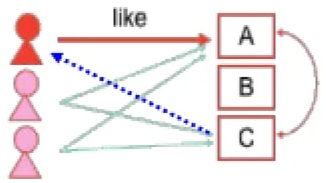
$$userSim(u,n) = \frac{\sum_{i \subset CR_{u,n}} (r_{ui} - \bar{r}_u)(r_{ni} - \bar{r}_n)}{\sqrt{\sum_{i \subset CR_{u,n}} (r_{ui} - \bar{r}_u)^2} \sqrt{\sum_{i \subset CR_{u,n}} (r_{ni} - \bar{r}_n)^2}}$$

Prediction Function :

$$pred(u,i) = \bar{r}_u + \frac{\sum_{n \subset neighbors(u)} userSim(u,n) \cdot (r_{ni} - \bar{r}_n)}{\sum_{n \subset neighbors(u)} userSim(u,n)}$$

Item Based Collaborative Filtering

- Use user-item ratings matrix
- Make item-to-item correlations
- Find items that are highly correlated
- Recommend items with highest correlation



Similarity Metric :

$$itemSim(i,j) = \frac{\sum_{u \subset RB_{i,j}} (r_{ui} - \bar{r}_u)(r_{uj} - \bar{r}_u)}{\sqrt{\sum_{u \subset RB_{i,j}} (r_{ui} - \bar{r}_u)^2} \sqrt{\sum_{u \subset RB_{i,j}} (r_{uj} - \bar{r}_u)^2}}$$

Prediction Function :

$$pred(u,i) = \frac{\sum_{j \in ratedItems(u)} itemSim(i,j) \cdot r_{ui}}{\sum_{j \in ratedItems(u)} itemSim(i,j)}$$

2. **Content Based Systems** recommend items similar to those a user has liked (browsed/purchased) in the past.

OR

Recommendations are based on the content of items rather on other user's opinion.

User Profiles: Create user profiles to describe the types of items that user prefers.

e.g. User1 likes sci-fi, action and comedy.

Recommendation on the basis of keywords are also classified under content based. e.g. Letizia

e.g. IMDB, Last.fm(scrobbler)

Content Based Systems Cont'd...

Advantages:

- No need for data on other users. No cold start and sparsity.
- Able to recommend users with unique taste.
- Able to recommend new and unpopular items.
- Can provide explanation for recommendation.

Limitations:

- Data should be in structured format.
- Unable to use quality judgements from other users.

Related Videos

1. A given time period, we count for each pair of videos v_i, v_j how often they were co-watched.
2. We denote this co-visitation count by c_ij.
3. We define **r**elated score of video v_j to v_i as : $r(v_i, v_j) = \frac{c_{ij}}{f(v_i, v_j)}$
4. $f(v_i, v_j)$ is a normalization function that denote the global popularity
5. Pick the set of related videos R_i for a given seed video v_i as the top N candidate videos ranked by their scores $r(v_i, v_j)$.

Generating Recommendation Candidates

1. A given seed set S (e.g. the videos user watched)
2. Each video v_i in the seed set consider its related videos R_i
3. Denote the union of these related video sets as C_1:

$$C_1(S) = \bigcup_{v_i \in C_S} R_i$$

4. Distance of n from any video in the seed set:

$$C_n(S) = \bigcup_{v_i \in C_{n-1}} R_i$$

Generating Recommendation Candidates

- The final candidate set C_{final} :

$$C_{final} = (\bigcup_{i=0}^{N} C_i) - S$$

- Due to the high branching factor of the related videos graph we found expanding over a small distance yielded broad and diverse recommendations even for users with a small seed set.

Ranking

Using a linear combination of three kinds of signals(a.video quality b.user specifility c.diversification),we generate a ranked list of the candidate videos.