Recommender Systems

Debapriyo Majumdar
Indian Statistical Institute Kolkata

Credits to Bing Liu (UIC) and Angshul Majumdar (IIITD) for some slides

Recommender systems

Business

- How to increase revenue?
- How to recommend items customers like?

Customer

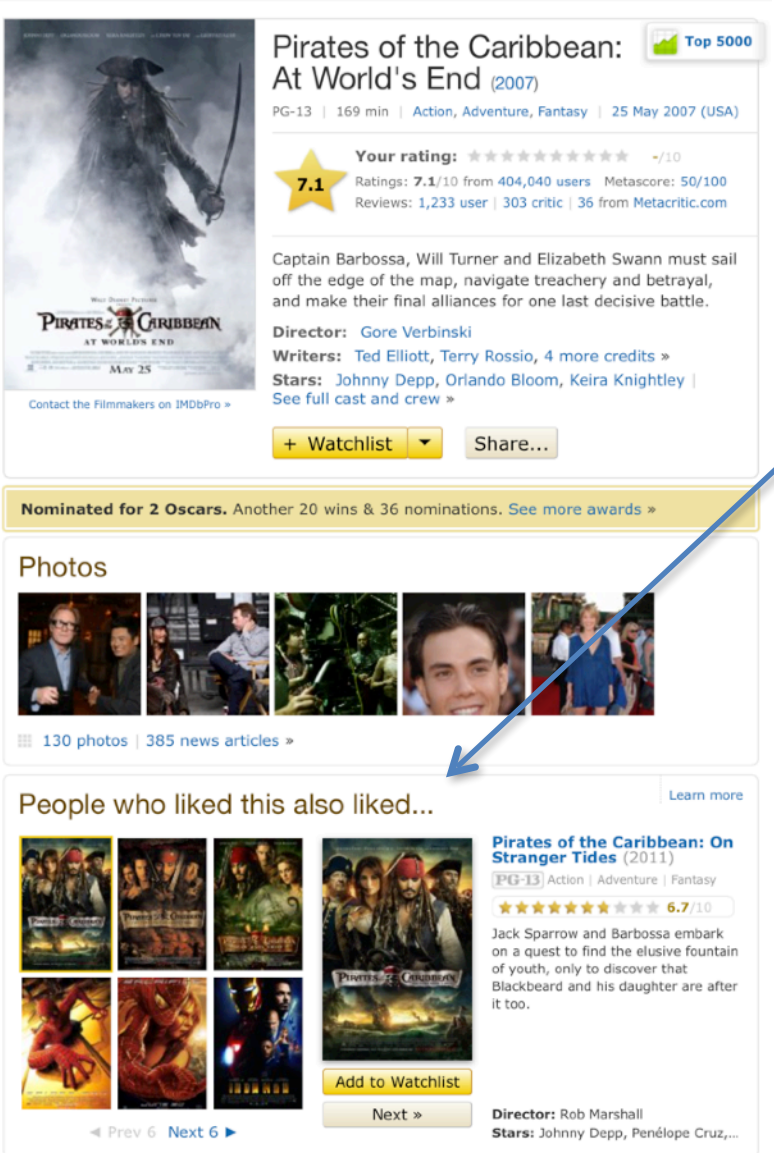
- Too many options.
- How to choose the right one?

Recommender systems

Customers who viewed / bought this product also bought

Since you are looking at this, you may also look at …

Recommender systems



Viewers who liked this movie also liked the other movies

Since you are looking at this page, you may also like…

The Recommendation Problem

- We have a set of users U and a set of items S to be recommended to the users.
- Let p be an utility function that measures the usefulness of item s ($\in S$) to user u ($\in U$), i.e.,

  $p : U \times S \rightarrow R$, where R is a totally ordered set (e.g., non-negative integers or real numbers in a range)
- Objective
  - Learn p based on the past data
  - Use p to predict the utility value of each item s ($\in S$) to each user u ($\in U$)

Two main formulations

- Rating prediction: predict the rating score that a user is likely to give to an item that (s)he has not seen or used before
  - Rating on an unseen movie
  - In this case, the utility of item *s* to user *u* is the rating given to *s* by *u*
- Item prediction: predict a ranked list of items that a user is likely to buy or use

Approaches

Content-based recommendations:

- The user will be recommended items similar to the ones the user preferred in the past

Collaborative filtering (or collaborative recommendations):

- The user will be recommended items that people with similar tastes and preferences liked in the past

Hybrids: Combine collaborative and content-based methods

Content based recommendation

- Will user *u* like item *s*?
- Look at items similar to *s*; does *u* like them?
  - Similarity based on content
  - Example: a movie represented based on features as specific actors, director, genre, subject matter, etc
- The user's interest or preference is also represented by the same set of features (the user profile)
- Candidate item *s* is compared with the user profile of *u* in the same feature space
- Determine if *u* would like *s*, or
- Top *k* similar items are recommended

Collaborative filtering

- Collaborative filtering (CF): more studied and widely used recommendation approach in practice
  - *k*-nearest neighbor
  - association rules based prediction
  - matrix factorization
- Key characteristic: predicts the utility of items for a user based on the items previously rated by other like-minded users (thus, *collaborative*)

k nearest neighbor approach

- No model building
- Utilizes the entire user-item database to generate predictions directly, i.e., there is no model building.
- This approach includes both
  - User-based methods
  - Item-based methods

User based kNN CF

- Let the record (or profile) of the target user be u (represented as a vector), and the record of another user be v.
- The similarity between the target user, u, and a neighbor, v, can be calculated using the **Pearson's correlation coefficient**:

$$\text{sim}(u,v) = \frac{\sum_{i \in S}(r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in S}(r_{u,i} - \bar{r}_u)^2}\sqrt{\sum_{i \in S}(r_{v,i} - \bar{r}_v)^2}}$$

  and compute V, is the set of k similar users based on this similarity

- Compute the rating prediction of item i for target user u

$$p(u,i) = \bar{r}_u + \frac{\sum_{v \in V} \text{sim}(u,v) \times (r_{v,i} - \bar{r}_v)}{\sum_{v \in V} |\,\text{sim}(u,v)\,|}$$

Problems with user based CF

- The problem with the user-based formulation of collaborative filtering is the lack of scalability:
  - it requires the real-time comparison of the target user to all user records in order to generate predictions
- A variation of this approach that remedies this problem is called item-based CF

Item-based CF

- The item-based approach works by comparing items based on their pattern of ratings across users. The similarity of items i and j is computed as follows:

$$\text{sim}(i,j) = \frac{\sum_{u \in U}(r_{u,i} - \bar{r}_i)(r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{i \in U}(r_{u,i} - \bar{r}_i)^2}\sqrt{\sum_{i \in U}(r_{u,j} - \bar{r}_j)^2}}$$

- After computing the similarity between items we select a set of k most similar items and generate a predicted value of user u's rating

$$p(u,i) = \frac{\sum_{j \in J} r_{u,j} \times \text{sim}(i,j)}{\sum_{j \in J} |\text{sim}(i,j)|}$$

where J is the set of k similar items

Association rule-based CF

- Transaction database: users, items
  - User → item: viewed, bought, liked
- Find association rules such as
  - Bought *X,* bought *Y* → Bought *Z*
  - Confidence and support (how strong is this association)
- Rank items based on measures such as confidence, subject to some minimum support
- Further reading: association rule mining