Recommender Systems

November 13, 2008

If I have 2 millions customers on the web, I should have 2 million stores on the web. (Jeff Bezos)

- converting browsers into buyers
- increasing cross-sell
- building loyalty
- directly selling (amazon, netflix, cdnow, itunes)
- information intermediary (yelp, tripadvisor)

Recommendation Interfaces

- similar item
- email
- text comment
- average rating
- top N

Item-to-Item Correlation

People-to-People Correlation

- collaborative filtering
- use group opinion to recommend information items
  - nearest neighbor technique
- building a database of preferences for products by consumers which have historically similar taste
  - scale of neighborhood network
  - length of information about a customer (think "browsing data")

User Input

- past purchase data
- evaluation data ("like it")
- text
- editor's choice

Finding Recommendations

- organic navigation: helping new and infrequent visitors
- request recommendation list
- inviting customers back: notification services
- building credibility through community: customer comments
- builidng long-term relationship: deep personalization

Collaborative Filtering Algorithm

- distinctions
  - memory-based model: uses entire database to make predictions
  - model-based: estimate or learn a model, then make a prediction
- implicit or explicit voting
  - explicit ranking
  - implicit ranking, purchase behavior, browsing history
- how to treat missing data (missing item is not random decision)

Memory Based Algorithm

- prediction about choice, preference
- average valuation

$$\bar{v}_i = \frac{1}{|I_i|} \sum_{j \in I_i} v_{i,j}$$

- the predicted vote for user a and the basis of some weights $w(a, i)$ are given by:

$$p_{a,j} = v_a + \kappa \sum w(a, i) (v_{i,j} - \bar{v}_i)$$

Data Mining

- m products: $\mathcal{P} = \{P_1, ..., P_m\}$
- transaction $T \subseteq \mathcal{P}$ is a set of products purchased together
- association rule $X \Rightarrow Y$ states that if X is in the transaction T, then Y is likely to be there as well
- support s of $X \Rightarrow Y$:

$$s = \frac{\#\text{ of transactions containing } X \cup Y}{\#\text{ of total transactions}}$$

- confindence c of $X \Rightarrow Y$:

$$s = \frac{\#\text{ of transactions containing } X \cup Y}{\#\text{ of transactions containing } X}$$

Top N Recommendation from Data Mining

- define minimum support and confidence requirements
- find all rules support by the history of the customer, X,
- sort predicted set P_u on the basis of confidence

Collaborative Filtering

- specific algorithm for recommending
- representation
- neighboorhood formation
- recommendation generation

Representation

- customer-product matrix $R = r_{i,j}$: i-th customer has purchased j-th product
- original representation: $m \times n$
  - sparsity with many products, loss of niehgborhood transitivity due to too few common products
  - scalability
- lower (reduced) dimensional representation with meta-products: $k \ll n$

Neighborhood Formation and Similarity

- correlation via Pearson correlation

$$corr_{a,b} = \frac{\sum_i \left(r_{a,i} - \bar{r}_a\right)\left(r_{b,i} - \bar{r}_b\right)}{\sqrt{\sum_i \left(r_{a,i} - \bar{r}_a\right)^2 \sum_i \left(r_{b,i} - \bar{r}_b\right)^2}}$$

- cosine

$$\cos\left(a, b\right) = \frac{a \cdot b}{\|a\|_2 \cdot \|b\|_2}$$

Generation of Recommendation

- most frequent item recommendation
- association rule based recommendation
- data available on movielens.org, netflix.com, grouplens, net perceptions