A Comprehensive Introduction to Recommendation Algorithms

Derrick Lam
Informatics Seminar
October 11, 2010

Overview

- Introduction
- Classification
- Algorithms
- Challenges
- Extensions

Introduction

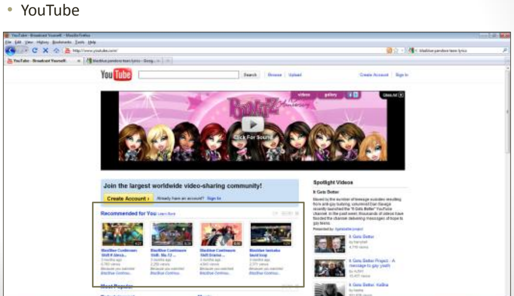
- Suggest items of interest to users
  - Products
  - People
  - Tasks
  - And more!
- Why is it needed?
  - Increase sales
  - Find related items of interest
  - Expand social networks

Introduction – Examples

- Amazon

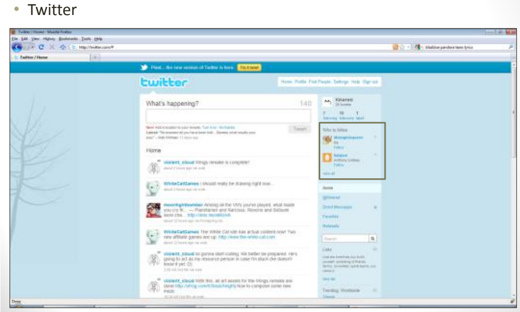
Introduction – Examples

- YouTube



Introduction – Examples

- Twitter



Classification

- What is recommended?
  - Items / Products
    - Items to buy
    - News articles to read
  - People / Users
    - People to meet
    - People to work with
  - Tasks
    - Tasks to perform

Classification

- How is the information collected?
  - Explicit – provided directly from the user
    - More accurate, but introduces work / frustration
    - Examples
      - Ratings made by the user
      - Interests entered in a user profile
      - Goals detailed by the user
  - Implicit – inferred from observations
    - Less intrusive, but less accurate
    - Examples
      - Interests inferred from items frequently viewed
      - Goals inferred from activities performed

Classification

- What is the basis of the recommendation?
  - Item-based – what items are similar to the ones liked?
  - User-based– what do similar users like?
- What is the underlying approach?
  - Association Rules
  - Collaborative Filtering
  - Cluster Based
  - Mixed (Hybrid)

Algorithms – Association Rules*

- What items are frequently viewed / purchased together?
- An association rule: $X \Rightarrow \mathrm{Y}$
  - $X, Y \subseteq P$ where P is the set of all products in the system
  - $X \cap Y = \emptyset$
  - The presence of an item in *X* in a transaction indicates a *strong likelihood* that an item in *Y* is also in the transaction
- Quality of a rule
  - $support\ s = \frac{\#\ transactions\ w\backslash\ X \cup Y}{\#\ transactions}$
  - $confidence\ c = \frac{\#\ transactions\ w\backslash\ X \cup Y}{\#\ transactions\ w\backslash\ X}$

* As described by Sarwar et al. in "Analysis of Recommendation Algorithms for E-Commerce"

Algorithms – Association Rules

- Recommendation Steps
  - Discover rules satisfying some minimum *s* and *c*
  - Find rules "supported" by the user (i.e. purchased all products on l.h.s. of rule)
  - Present top-*n* rules to the user, sorted by *c*

Algorithms – Collaborative Filtering*

- What do similar users (neighbors) like?
- Steps:
  - Neighborhood formation – how to identify similar users
    - Pearson correlation

$$corr_{ab} = \frac{\sum_i (r_{ai} - \bar{r_a})(r_{bi} - \bar{r_b})}{\sqrt{\sum_i (r_{ai} - \bar{r_a})^2 \sum_i (r_{bi} - \bar{r_b})^2}}$$

    - Cosine similarity

$$\cos(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|_2 * \|\vec{b}\|_2}$$

  - Recommendation generation – finding the top-n items to recommend
    - Most-frequent items
    - Best-rated items

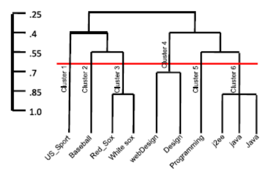
* As described by Sarwar et al. in "Analysis of Recommendation Algorithms for E-Commerce"

Algorithms – Cluster Based

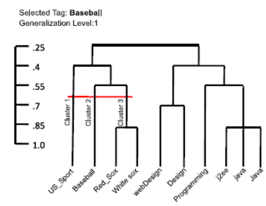
- Group items based on tags, keywords, features, etc.
- Context-Dependent Hierarchical Agglomerative Clustering* steps:
  - Assign every tag to a singleton cluster
  - Combine all tags in one hierarchical cluster
    - Combine clusters
    - Lower similarity threshold
  - Identify the user context
  - Divide a branch of the tree into separate clusters

* Approach by Shepitsen et al. in "Personalized Recommendation in Social Tagging Systems using Hierarchical Clustering"

Algorithms – Cluster Based



Standard Hierarchical Agglomerative Clustering



Context-Dependent Hierarchical Agglomerative Clustering

Algorithms – Cluster Based

- Recommendation steps
  - Calculate cosine similarity between query tag and resources: $S(q,r)$
  - Calculate relevance to user
    - Calculate the user's interest in each cluster
      $$uc\_w(u,c) = \frac{|a =< u,r,t >\in A: r \in R, t \in c|}{|a =< u,r,t >\in A: r \in R, t \in T|}$$
    - Calculate each resource's closest clusters
      $$rc\_w(r,c) = \frac{|a =< u,r,t >\in A: u \in U, t \in c|}{|a =< u,r,t >\in A: u \in U, t \in T|}$$
    - Infer the user's interest in each resource
      $$I(u,r) = \sum_{c\in C} uc\_w(u,c) * rc\_w(r,c)$$
  - Calculate personalized rank score: $S^*(q,r) = S(q,r) * I(u,r)$

Algorithms – Hybrid

- Types*
  - Combining results from separate recommenders
  - Combining attributes from different approaches
- Example – Content-Boosted CF**
  - Collaborative filtering weakness – need many ratings from users for good recommendations
  - Use "pseudo" ratings based on item content when explicit rating not available

* As described by Adomavicius et al. in "Toward the Next Generation of Recommender System: a Survey of the State of the Art and Possible Extensions"
** Approach by Melville et al. in "Content-Boosted Collaborative Filtering for Improved Recommendations"

Challenges

- New User / Item Problem
- Sparsity
- Scalability
- Computational Cost
- Collusion

Extensions

- Trust and Reputation
- Multi-faceted Ratings
- Machine Learning

References

- Adomavicius, G., Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: a Survey of the State-of-the-Art and Possible Extensions. In *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734-749.
- Sarwar, B., Karypis, G., Konstan, J., Riedl, J. (2000). Analysis of Recommendation Algorithms for E-commerce. In *Proceedings of the 2nd ACM Conference on Electronic Commerce*, pp. 158-167.
- Shepitsen, A., Gemmell, J., Mobasher, B., Burke, R. (2008). Personalized Recommendation in Social Tagging Systems Using Hierarchical Clustering. In *Proceedings of the 2008 ACM Conference on Recommender Systems*, pp. 259-266.
- Melville, P., Mooney, R., Nagarajon, R. (2002). Content-boosted Collaborative Filtering for Improved Recommendations. In Eighteenth National Conference on Artificial Intelligence, pp. 187-192.

Questions