Rating based Recommendation: From Practice to Theory

Present by Feng Xie

May 30, 2012
Outline

• What is recommender system?
  • Mission
  • History
  • Problems

• What is good recommender system?
  • Experiment Methods
  • Evaluation Metric

• Rating based recommender algorithms

• Two instances
  • YouTube Video Recommendation
  • Rating Frequency based Recommendation

• Conclusions & Future work
What is recommender system?
Information Overload
How to solve information overload?
Search Engine VS. Recommender System

- User will try search engine if
  - they have specific needs
  - they can use keywords to describe needs

- User will try recommender system if
  - they do not know what they want now
  - they can not use keywords to describe needs
Mission

• Help user find item of their interest
• Help item provider deliver their item to right user
• Help website improve user’s loyalty
History I

• Content Filtering
  • An architecture for large scale information systems [1985] (Gifford, D.K)

• Tapestry
  • First collaborative filtering system [1992] (Xerox Palo Alto)
History II

- **GroupLens**
  - First rating based collaborative filtering system [1992]

- **Movielens**
  - First movie recommender system [1997]
  - Provide well-know dataset for researchers
  - MovieLens 100K, MovieLens 1M, MovieLens 10M
ACM Software System Award

- Awarded to an institution or individual(s) recognized for developing a software system that has had a lasting influence, reflected in contributions to concepts, in commercial acceptance, or both. The Software System Award carries a prize of $35,000. Financial support for the Software System Award is provided by IBM.

2010 – GroupLens Collaborative Filtering Recommender Systems

Bergstrom, Peter
Gordon, Lee R
Herlocker, Jonathan L
Iacovou, Neophytos
Konstan, Joseph A
Lam, Shyong (Tony) K.
Maltz, David
McNee, Sean
Miller, Bradley N
Resnick, Paul J
Riedl, John T
Suchak, Mitesh
History III

- Amazon proposed item-based collaborative filtering (Patent is filed in 1998 and issued in 2001)
- Netflix Prize (2006-2009)
  - 1M $ improve accuracy by 10% in term of RMSE
  - Yehuda Koren’s team get prize
- ACM Conference on Recommender System
  - Minneapolis, Minnesota, USA [2007]
Problems

- Top-N Recommendation
- Rating Prediction
Top-N Recommendation

- **Input**

<table>
<thead>
<tr>
<th>user</th>
<th>item</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a</td>
</tr>
<tr>
<td>B</td>
<td>a</td>
</tr>
<tr>
<td>B</td>
<td>b</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

- **Output**
Rating Prediction

- **Input**

<table>
<thead>
<tr>
<th>user</th>
<th>item</th>
<th>rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a</td>
<td>⭐⭐⭐⭐⭐⭐</td>
</tr>
<tr>
<td>B</td>
<td>a</td>
<td>⭐⭐⭐⭐⭐⭐</td>
</tr>
<tr>
<td>B</td>
<td>b</td>
<td>⭐⭐⭐⭐⭐⭐</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

- **Output**

<table>
<thead>
<tr>
<th>user</th>
<th>item</th>
<th>rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>⭐⭐⭐⭐⭐⭐</td>
</tr>
</tbody>
</table>

NSlab@RIIT of TSINGHUA UNIVERSITY
What is good recommender system?
Experiment Methods

- Offline Experiment
  - Train/Test
- User Survey
- Online Experiment
  - AB Testing
Experiment Metrics

• Prediction Accuracy
  • Top-N Recommendation: Precision, Recall
  • Rating Prediction: MAE, RMSE

• Coverage

• Diversity
  • The ability to cover users’ different interests

• Real-time
  • Generate new recommendations when users have new behaviors immediately

• ……
Rating based Recommender Algorithms
The Process

• **Input**
  - List of m users and a list of n Items
  - Each user has a list of items he/she expressed their opinion about (can be a null set)
    - Explicit opinion: a rating score (numerical scale)
    - Implicitly: purchase records

• **Output**
  - Prediction: a numerical value, expressing the predicted likeliness of an item the user hasn’t expressed his/her opinion about
  - Recommendation: a list of N items the active user will like the most (Top-N recommendations)
Practical Prediction Methods

- Baseline Algorithms
  - Per User Average
    - A rating prediction is to take the average of the previous ratings of the target user
  - Per Item Average
    - A rating prediction is to take the average of the previous ratings of all users on the target item

- Shortcoming
  - The former lacks diversity
  - The latter suffers from non-personalization
Collaborative Filtering

- User based
  - Users with similar history selections will share the same future interest

- Item based
  - Users will like items similar to what they consumed before
The Process

- Similarity Measure
  - Cosine
  - Pearson Correlation
- Rating Prediction

\[ p_{u,i} = \bar{r}_u + \frac{\sum_{v \in U(u)} s(v,u)(r_{v,i} - \bar{r}_v)}{\sum_{v \in U(u)} |s(v,u)|} \]

\[ p_{u,i} = \bar{r}_i + \frac{\sum_{j \in I(i)} s(j,i)(r_{u,j} - \bar{r}_j)}{\sum_{j \in I(i)} |s(j,i)|} \]
Existing Problems

- Data Sparsity
  - Difficult to compute similarities
  - RBRA, Grey Forecast model
- Data Correlation
  - Inaccurate similarity measure
  - Orthogonalization, Grey Forecast model
- Rating Prediction
  - Grey Forecast model
  - RBRA
  - Rating Frequency (RF)
Two instances

=YouTube + Rating Frequency
Video Recommendation I

- Video Website Classification
  - User Generated Content (UGC)
    - YouTube, YouKu, TuDou
  - Specialized Video Content (SVC)
    - Hulu, Netflix, QiYi, SOHU
# Video Recommendation II

<table>
<thead>
<tr>
<th></th>
<th>UGC</th>
<th>SVC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video Number</td>
<td>More than 100M</td>
<td>Less than 1M</td>
</tr>
<tr>
<td>Video Length</td>
<td>Short</td>
<td>Long</td>
</tr>
<tr>
<td>Content Metadata</td>
<td>Tag only</td>
<td>Structured</td>
</tr>
<tr>
<td>Lifespan</td>
<td>Short</td>
<td>Long</td>
</tr>
<tr>
<td>Video Quality</td>
<td>Mixed, Substantially duplicate content</td>
<td>Good</td>
</tr>
<tr>
<td>Video Diversity</td>
<td>Variety</td>
<td>Not enough</td>
</tr>
</tbody>
</table>
YouTube Video Recommendation

- Video Suggestion and Discovery for YouTube: Taking Random Walks Through the View Graph (WWW’08)
- The YouTube Video Recommendation System (RecSys’10)
YouTube (WWW’08)

• User-Video Graph
  • Find video $v$ for user $u$
    • $u$ and $v$ have a short path between them
    • $u$ and $v$ have several paths between them
    • $u$ and $v$ have paths that avoid high-degree nodes
  • Solution
    • Adsorption Averaging
    • Adsorption Random Walk
    • Adsorption Linear System
YouTube (RecSys’10)

- User Needs
  - To watch a single video that they found elsewhere
  - To find specific videos around a topic
  - To just be entertained by content that they find interesting
Algorithm Design I

- Similarity Computation

\[ r(v_i, v_j) = \frac{c_{ij}}{f(v_i, v_j)} \]

\[ f(v_i, v_j) = c_i \cdot c_j \]

- Video Graph
Algorithm Design II

- Related Video Set
  \[ C_1(S) = \bigcup_{v_i \in S} R_i \]
  \[ C_n(S) = \bigcup_{v_i \in C_{n-1}} R_i \]

- Final Candidate
  \[ C_{\text{final}} = (\bigcup_{i=0}^{N} C_i) \setminus S \]
Algorithm Design III

- Ranking
  - Video quality
  - User specificity
  - Diversification
- Recommendation
Evaluation

- Comparison
  - Most viewed
  - Top Favorited
  - Top Rated
  - Recommended

- Evaluation Method
  - 21 days A/B online test

- Metric
  - Click Through Rate (CTR)
Result

The graph shows the normalized click-through rate for different categories over a week.

- **Most Viewed**
- **Top Favorited**
- **Top Rated**
- **Recommended**

The y-axis represents the normalized click-through rate, and the x-axis represents the days of the week from Monday to Sunday.
Conclusions

• Hundreds of recommender algorithms are useless in practice due to
  • Algorithm complexity
  • Dataset dependence
• Compromise between accuracy and efficiency
  • User/Item based collaborative filtering
  • Hybrid method
  • Amazon, YouTube, Hulu
• No demand, no product; no practice, no theory
Future Work

- Cross domain / transfer learning
- Social recommendation
- Friends recommendation
- Group recommendation (group discovery / clustering)
- Graph based recommendation
- Privacy preserving
Questions?
Thank you!