Personalized Entity Recommendation: A Heterogeneous Information Network Approach

Xiao Yu, Xiang Ren, Yizhou Sun†, Quanquan Gu, Bradley Sturt, Urvashi Khandelwal, Brandon Norick, Jiawei Han

University of Illinois, at Urbana Champaign

†Northeastern University

02/26/2014
Booming Age of Heterogeneous Information Networks

Information network with multi-typed entities and relationships

Social Media

E-Commerce

Biology

Healthcare

Global Economy
Hybrid Collaborative Filtering with Networks

• Utilizing network relationship information can enhance the recommendation quality

• However, most of the previous studies only use single type of relationship between users or items (e.g., social network Ma, WSDM’11, trust relationship Ester, KDD’10, service membership Yuan, RecSys’11)
The Heterogeneous Information Network View of Recommender System

- Avatar
- Aliens
- Titanic
- Revolution
- Romance
- Zoe Saldana
- Adventure
- Leonardo Dicaprio
- Kate Winslet
- James Cameron
Relationship heterogeneity alleviates data sparsity

Collaborative filtering methods suffer from data sparsity issue

- A small number of users and items have a large number of ratings
- Most users and items have a small number of ratings

- Heterogeneous relationships complement each other
- Users and items with limited feedback can be connected to the network by different types of paths
  - Connect new users or items (cold start) in the information network
Relationship heterogeneity based personalized recommendation models

Different users may have different behaviors or preferences

Two levels of personalization

Data level
• Most recommendation methods use one model for all users and rely on personal feedback to achieve personalization

Model level
• With different entity relationships, we can learn personalized models for different users to further distinguish their differences
Preference Propagation-Based Latent Features

Generate $L$ different meta-path (path types) connecting users and items

Propagate user implicit feedback along each meta-path

Calculate latent features for users and items for each meta-path with NMF related method
Recommendation Models

**Observation 1:** Different meta-paths may have different importance

Global Recommendation Model

\[
\hat{r}(u_i, e_j) = \sum_{q=1}^{L} \theta_q \cdot \hat{U}_i(q) \hat{V}_j(q)^T
\]  
(1)

**Observation 2:** Different users may require different models

Personalized Recommendation Model

\[
\hat{r}_p(u_i, e_j) = \sum_{k=1}^{c} \text{sim}(C_k, u_i) \sum_{q=1}^{L} \theta_q^{\{k\}} \cdot \hat{U}_i(q) \hat{V}_j(q)^T
\]  
(2)
Parameter Estimation

- Bayesian personalized ranking (Rendle UAI’09)
- Objective function

\[
\min_{\Theta} - \sum_{u_i \in U} \sum_{(e_a > e_b) \in R_i} \ln \sigma(\hat{r}(u_i, e_a) - \hat{r}(u_i, e_b)) + \frac{\lambda}{2} \|\Theta\|_2^2
\]

for each correctly ranked item pair
i.e., \( u_i \) gave feedback to \( e_a \) but not \( e_b \)

\[
\sigma(x) = \frac{1}{1+e^{-x}}.
\]

Learning Personalized Recommendation Model

- Soft cluster users with NMF + k-means
- For each user cluster, learn one model with Eq. (3)
- Generate personalized model for each user on the fly with Eq. (2)
Experiment Setup

• Datasets

<table>
<thead>
<tr>
<th>Name</th>
<th>#Items</th>
<th>#Users</th>
<th>#Ratings</th>
<th>#Entities</th>
<th>#Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>IM100K</td>
<td>943</td>
<td>1360</td>
<td>89,626</td>
<td>60,905</td>
<td>146,013</td>
</tr>
<tr>
<td>Yelp</td>
<td>11,537</td>
<td>43,873</td>
<td>229,907</td>
<td>285,317</td>
<td>570,634</td>
</tr>
</tbody>
</table>

• Comparison methods:
  • **Popularity**: recommend the most popular items to users
  • **Co-click**: conditional probabilities between items
  • **NMF**: non-negative matrix factorization on user feedback
  • **Hybrid-SVM**: use Rank-SVM with plain features (utilize both user feedback and information network)
### Performance Comparison

<table>
<thead>
<tr>
<th>Method</th>
<th>IM100K</th>
<th></th>
<th></th>
<th></th>
<th>Yelp</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prec1</td>
<td>Prec5</td>
<td>Prec10</td>
<td>MRR</td>
<td>Prec1</td>
<td>Prec5</td>
<td>Prec10</td>
<td>MRR</td>
</tr>
<tr>
<td>Popularity</td>
<td>0.0731</td>
<td>0.0513</td>
<td>0.0489</td>
<td>0.1923</td>
<td>0.00747</td>
<td>0.00825</td>
<td>0.00780</td>
<td>0.0228</td>
</tr>
<tr>
<td>Co-Click</td>
<td>0.0668</td>
<td>0.0558</td>
<td>0.0538</td>
<td>0.2041</td>
<td>0.0147</td>
<td>0.0126</td>
<td>0.01132</td>
<td>0.0371</td>
</tr>
<tr>
<td>NMF</td>
<td>0.2064</td>
<td>0.1661</td>
<td>0.1491</td>
<td>0.4938</td>
<td>0.0162</td>
<td>0.0131</td>
<td>0.0110</td>
<td>0.0382</td>
</tr>
<tr>
<td>Hybrid-SVM</td>
<td>0.2087</td>
<td>0.1441</td>
<td>0.1241</td>
<td>0.4493</td>
<td>0.0122</td>
<td>0.0121</td>
<td>0.0110</td>
<td>0.0337</td>
</tr>
<tr>
<td>HeteRec-g</td>
<td>0.2094</td>
<td>0.1791</td>
<td>0.1614</td>
<td>0.5249</td>
<td>0.0165</td>
<td>0.0144</td>
<td>0.0129</td>
<td>0.0422</td>
</tr>
<tr>
<td>HeteRec-l</td>
<td><strong>0.2121</strong></td>
<td><strong>0.1932</strong></td>
<td>0.1681</td>
<td><strong>0.5530</strong></td>
<td><strong>0.0213</strong></td>
<td><strong>0.0171</strong></td>
<td><strong>0.0150</strong></td>
<td><strong>0.0513</strong></td>
</tr>
</tbody>
</table>

HeteRec personalized recommendation (HeteRec-p) provides the best recommendation results.
Performance under Different Scenarios

(a) Performance Change with User Feedback Number
(b) Performance Change with User Feedback Popularity

HeteRec–p consistently outperform other methods in different scenarios.
Better recommendation results if users provide more feedback.
Better recommendation for users who like less popular items.
Contributions

• Propose latent representations for users and items by propagating user preferences along different meta-paths

• Employ Bayesian ranking optimization technique to correctly evaluate recommendation models

• Further improve recommendation quality by considering user differences at model level and define personalized recommendation models
  • Two levels of personalization