Graph Heterogeneous Multi-Relational Recommendation

Chong Chen, Min Zhang
Department of Computer Science and Technology, Tsinghua University
Background (1): Users’ Sparse Feedback Information

• RecSys has become a major monetization tool for customer-oriented online services
  • E.g., E-commerce, News Portal, Social Networks, etc.
• Users usually rate or click a small set of items compared to hundreds of millions of items in the system
  • Sparse
  • Positive-unlabeled data
Background (2): Multi-relational Recommendation

- Multi-relational recommendation
  - Heterogeneous interactions between user and item (view, click, cart, purchase).
  - In addition to the target behavior, other behavior data also provides well-structured information and can be used for high-quality recommendation.
Preliminary: Graph Convolutional Networks

- Most existing research on graph convolutional networks are focused on learning representations of nodes in simple undirected graphs.
  \[ E = \sigma(\hat{A}E^{(0)}W) \]

- For capturing high-hop dependencies in the graph, several GCN layers can be stacked:
  \[ E^{(l)} = \sigma(\hat{A}E^{(l-1)}W^{(l)}) \]

- For a relational graph, the GCN formulation is as follows:
  \[ E^{(l)} = \sigma(\hat{A}E^{(l-1)}W^{(l)}_{r}) \]

- This formulation leads to over-parameterization and embeds only nodes in the graph.
Graph Heterogeneous Collaborative Filtering

- Relation-aware GCN propagation layers to jointly embed both representations of nodes and relations in a graph
- Efficient non-sampling learning module to achieve more effective and stable model optimization
- Multi-task prediction
Embedding Propagation Layers

• For node:

$$\mathbf{e}_u^{(l)} = \sigma \left( \sum_{(v,r) \in \mathcal{N}(u)} \frac{1}{\sqrt{|\mathcal{N}_u| |\mathcal{N}_v|}} \mathbf{W}_r^{(l)} \mathbf{e}_v^{(l-1)} \right)$$

$$\mathbf{e}_u^{(l)} = \sigma \left( \sum_{(v,r) \in \mathcal{N}(u)} \frac{1}{\sqrt{|\mathcal{N}_u| |\mathcal{N}_v|}} \mathbf{W}_r^{(l)} \phi(\mathbf{e}_v^{(l-1)}, \mathbf{e}_r^{(l-1)}) \right)$$

where

$$\phi(\mathbf{e}_v, \mathbf{e}_r) = \mathbf{e}_v \odot \mathbf{e}_r$$

• For relation:

$$\mathbf{e}_r^{(l)} = \mathbf{W}_{rel}^{(l)} \mathbf{e}_r^{(l-1)}$$
Multi-task Prediction

- Feature fusion

$$e_u = \sum_{l=0}^{L} \frac{1}{L+1} e^{(l)}_u; \ e_v = \sum_{l=0}^{L} \frac{1}{L+1} e^{(l)}_v; \ e_r = \sum_{l=0}^{L} \frac{1}{L+1} e^{(l)}_r$$

- Prediction

$$\hat{y}_{(k)uv} = e_u^T \cdot \text{diag}(e_{r_k}) \cdot e_v = \sum_{i} e_{u,i}e_{r_k,i}e_{v,i}$$

Efficient Multi-task Learning without Sampling

$$\tilde{L}_k(\Theta) = \sum_{u \in B} \sum_{v \in V^{+}_{(u)}} \left( (c^+_v - c^-_v) \hat{y}^2_{(k)uv} - 2c^+_v \hat{y}_{(k)uv} \right)$$

$$+ \sum_{i=1}^{d} \sum_{j=1}^{d} \left( e_{r_k,i}e_{r_k,j} \left( \sum_{u \in B} e_{u,i}e_{u,j} \right) \left( \sum_{v \in V} c^+_v e_{v,i}e_{v,j} \right) \right)$$
Experimental settings

- **Datasets:**
- **Baselines:**
  - NCF (WWW 17)
  - ENMF (SIGIR 19)
  - LightGCN (SIGIR 20)
  - CMF (WWW 15)
  - MC-BPR (RecSys 16)
  - NMTR (ICDE 19)
  - EHCF (AAAI 19, our previous work)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#User</th>
<th>#Item</th>
<th>#View</th>
<th>#Add-to-cart</th>
<th>#Purchase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beibe</td>
<td>21,716</td>
<td>7,977</td>
<td>2,412,586</td>
<td>642,622</td>
<td>304,576</td>
</tr>
<tr>
<td>Taobao</td>
<td>48,749</td>
<td>39,493</td>
<td>1,548,126</td>
<td>193,747</td>
<td>259,747</td>
</tr>
</tbody>
</table>
### Model Comparisons

- **Performance comparison on three datasets for all methods**
- **Best Baselines:**
  - LightGCN: Single
  - EHCF: Heterogeneous
  - GHCF
- GHCF consistently and significantly outperforms the best baseline

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Beibei</th>
<th>Taobao</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR@10</td>
<td>HR@50</td>
</tr>
<tr>
<td>Single</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPR</td>
<td>0.0437</td>
<td>0.1246</td>
</tr>
<tr>
<td>NCF</td>
<td>0.0441</td>
<td>0.1562</td>
</tr>
<tr>
<td>ENMF</td>
<td>0.0464</td>
<td>0.1637</td>
</tr>
<tr>
<td>LightGCN</td>
<td>0.0451</td>
<td>0.1613</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMF</td>
<td>0.0482</td>
<td>0.1582</td>
</tr>
<tr>
<td>MC-BPR</td>
<td>0.0504</td>
<td>0.1743</td>
</tr>
<tr>
<td>NMTR</td>
<td>0.0524</td>
<td>0.2047</td>
</tr>
<tr>
<td>EHCF</td>
<td>0.1523</td>
<td>0.3316</td>
</tr>
<tr>
<td>GHCF</td>
<td><strong>0.1922</strong></td>
<td><strong>0.3794</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Taobao</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Single</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPR</td>
<td>0.0376</td>
<td>0.0708</td>
<td>0.0871</td>
<td>0.0227</td>
<td>0.0269</td>
<td>0.0305</td>
</tr>
<tr>
<td>NCF</td>
<td>0.0391</td>
<td>0.0728</td>
<td>0.0897</td>
<td>0.0233</td>
<td>0.0281</td>
<td>0.0321</td>
</tr>
<tr>
<td>ENMF</td>
<td>0.0398</td>
<td>0.0743</td>
<td>0.0936</td>
<td>0.0244</td>
<td>0.0298</td>
<td>0.0339</td>
</tr>
<tr>
<td>LightGCN</td>
<td>0.0415</td>
<td>0.0814</td>
<td>0.1025</td>
<td>0.0237</td>
<td>0.0325</td>
<td>0.0359</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMF</td>
<td>0.0483</td>
<td>0.0774</td>
<td>0.1185</td>
<td>0.0252</td>
<td>0.0293</td>
<td>0.0357</td>
</tr>
<tr>
<td>MC-BPR</td>
<td>0.0547</td>
<td>0.0791</td>
<td>0.1264</td>
<td>0.0263</td>
<td>0.0297</td>
<td>0.0361</td>
</tr>
<tr>
<td>NMTR</td>
<td>0.0585</td>
<td>0.0942</td>
<td>0.1368</td>
<td>0.0278</td>
<td>0.0334</td>
<td>0.0394</td>
</tr>
<tr>
<td>EHCF</td>
<td>0.0717</td>
<td>0.1618</td>
<td>0.2211</td>
<td>0.0403</td>
<td>0.0594</td>
<td>0.0690</td>
</tr>
<tr>
<td>GHCF</td>
<td><strong>0.0807</strong></td>
<td><strong>0.1892</strong></td>
<td><strong>0.2599</strong></td>
<td><strong>0.0442</strong></td>
<td><strong>0.0678</strong></td>
<td><strong>0.0792</strong></td>
</tr>
</tbody>
</table>
Handling Cold-Start Issue

- GHCF
- Consistently significantly outperforms the best baseline
- The effectiveness of leveraging auxiliary behavior to alleviate the data sparsity issue and the strong power of our GHCF model
Ablation Study

- **GHCF-P**: The variant model of GHCF which utilizes only purchase data.
- **GHCF-PV**: The variant model of GHCF which utilizes purchase data and view data.
- **GHCF-PC**: The variant model of GHCF which utilizes purchase data and carting data.
Conclusion

- We propose a novel neural model named GHCF for multi-relational recommendation, which uncovers the underlying relationships among heterogeneous user-item interactions and shows multi-task ability to predict various types of user behaviors using one unified model.

- We design relation-aware GCN propagation layers, which jointly embed both representations of nodes (users and items) and relations in a graph to explicitly exploit the collaborative high-hop signals.

- The model consistently outperforms the state-of-the-art recommendation methods, especially for cold-start users.
THANKS

2020.12.19