Heterogeneous Embedding Propagation for Large-scale E-Commerce User Alignment

Vincent W. Zheng, Mo Sha, Yuchen Li, Hongxia Yang, Yuan Fang, Zhenjie Zhang, Kian-Lee Tan, Kevin Chen-Chuan Chang

ICDM 2018 @ Singapore
Outline

- Data and Problem
- Overall Framework
- Proposed Model
- Experiments
- Conclusion
Data and Problem

Data: User Activity Log

<table>
<thead>
<tr>
<th>Time</th>
<th>User</th>
<th>IP</th>
<th>Keywords</th>
<th>Auction</th>
<th>Shop</th>
</tr>
</thead>
<tbody>
<tr>
<td>04/05/2017 16:21</td>
<td>PID1</td>
<td>IP2</td>
<td>toys</td>
<td>-</td>
<td>Shop3</td>
</tr>
<tr>
<td>04/05/2017 22:12</td>
<td>MID3</td>
<td>IP2</td>
<td>lego</td>
<td>Auction1</td>
<td>Shop2</td>
</tr>
</tbody>
</table>

Problem: User ID Linking

To determine if PID1 (a PC identifier) is the same user as MID3 (a mobile device identifier).
Modeling Data as Heterogeneous Interaction Graph

<table>
<thead>
<tr>
<th>Time</th>
<th>User</th>
<th>IP</th>
<th>Keywords</th>
<th>Auction</th>
<th>Shop</th>
</tr>
</thead>
<tbody>
<tr>
<td>04/05/2017 16:21</td>
<td>PID1</td>
<td>IP2</td>
<td>toys</td>
<td>-</td>
<td>Shop3</td>
</tr>
<tr>
<td>04/05/2017 22:12</td>
<td>MID3</td>
<td>IP2</td>
<td>lego</td>
<td>Auction1</td>
<td>Shop2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Exploit interactions among sparse items
Technical challenges: Heterogeneity

Node heterogeneity
- Different types of nodes with various semantics
  - Users
  - IPs
  - Keywords
  - Auctions
  - Shops

Edge features
- Time-based historical access patterns
  - How frequent in past 24 hour?
  - How frequent on Sundays?
  - How frequent in the evenings?
Outline

- Data and Problem
- **Overall Framework**
- Proposed Model
- Experiments
- Conclusion
Overall Framework: Heterogeneous Embedding Propagation (HEP)

Classification loss + Reconstruction loss
Outline

- Data and Problem
- Overall Framework
- Proposed Model
- Experiments
- Conclusion
Semi-supervised learning: some PID-MID pairs \((v_i, u_i)\) are known to be positive or negative \((y_i)\).

\[
P(y_i | v_i, u_i) = \sigma \left( y_i \cdot h_{v_i}^T W h_{u_i} \right)
\]

\[
L_1 = -\frac{1}{m} \sum_{i=1}^{m} \log P(y_i | v_i, u_i)
\]

\(h\) is node embedding
\(W\) is weight matrix
Proposed Model: Reconstruction Loss

Reconstruction loss
Node embeddings are reconstructed from neighbors, aggregated by node type \((c)\).

\[
\tilde{g}_v^{(c)} = \sum_{u \in N_v^{(c)}} \frac{s_{v,u}}{\sum_{u \in N_v^{(c)}} s_{v,u}} h_u
\]

\[
\tilde{g}_v = \text{CONCAT} \left( \tilde{g}_v^{(c_1)}, \ldots, \tilde{g}_v^{(c_{n_1})} \right)
\]

\[
\tilde{h}_v = \sigma \left( W'_{\phi(v)} \tilde{g}_v + b''_{\phi(v)} \right)
\]

\(s_{v,u}\) is learnable edge weight (based on edge feature)
\(\tilde{h}\) is reconstructed node embedding
\(W'\) and \(b''\) is type specific weight/bias
Proposed Model: Reconstruction Loss

Reconstruction loss
Reconstructed embedding ($\tilde{h}$) should be close to the target embedding ($h$).

$$\ell(v, u) = \left[ \gamma + \pi(\tilde{h}_v, h_v) - \pi(\tilde{h}_v, h_u) \right]_+$$

$$L_2 = \frac{1}{|V|} \sum_{v \in V} \sum_{u \sim P_n(u)} \ell(v, u)$$

$\gamma$ is margin (hyperparameter)
$\pi$ is distance function between embedding
$P_n$ is negative sampling distribution
Outline

- Data and Problem
- Overall Framework
- Proposed Model
- **Experiments**
- Conclusion
Experiments: Datasets

- Taobao’s one-week user activity log in a city
- TB-Top: top 10% active users
- TB-Top: random 10% users

<table>
<thead>
<tr>
<th></th>
<th>#record</th>
<th>#PID</th>
<th>#MID</th>
<th>#IP</th>
<th>#shop</th>
<th>#auction</th>
<th>#keyword</th>
<th>#pos</th>
<th>#neg</th>
</tr>
</thead>
<tbody>
<tr>
<td>TB-Top</td>
<td>73,394K</td>
<td>204K</td>
<td>53K</td>
<td>277K</td>
<td>1,125K</td>
<td>3,718K</td>
<td>1,317K</td>
<td>147K</td>
<td>10,611K</td>
</tr>
<tr>
<td>TB-Rnd</td>
<td>31,202K</td>
<td>99K</td>
<td>46K</td>
<td>167K</td>
<td>495K</td>
<td>1,082K</td>
<td>437K</td>
<td>57K</td>
<td>2,363K</td>
</tr>
</tbody>
</table>
Experiments: Baselines

- Validating data model (as a graph)
  - FEM: feature engineering
  - LDA: latent Dirichlet allocation
  - GRU: gated recurrent unit

- Validating technical model (HEP)
  - Metapath2vec: meta-path based embedding
  - EP: embedding propagation
  - HEP-: HEP without edge features
Experiments: Results

<table>
<thead>
<tr>
<th></th>
<th>Preciseion</th>
<th>Recall</th>
<th>F1</th>
<th></th>
<th>Preciseion</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FEM</strong></td>
<td>60.3</td>
<td>3.4</td>
<td>6.4</td>
<td></td>
<td>68.7</td>
<td>1.9</td>
<td>3.7</td>
</tr>
<tr>
<td><strong>LDA</strong></td>
<td>70.4</td>
<td>10.6</td>
<td>18.5</td>
<td></td>
<td>68.3</td>
<td>6.1</td>
<td>11.3</td>
</tr>
<tr>
<td><strong>GRU</strong></td>
<td>51.8</td>
<td>26.2</td>
<td>34.8</td>
<td></td>
<td>52.6</td>
<td>22.1</td>
<td>31.2</td>
</tr>
<tr>
<td>Metapath2vec</td>
<td>1.7</td>
<td>62.9</td>
<td>3.4</td>
<td></td>
<td>2.3</td>
<td>58.7</td>
<td>4.4</td>
</tr>
<tr>
<td>EP</td>
<td>34.3</td>
<td>6.7</td>
<td>11.2</td>
<td></td>
<td>35.0</td>
<td>6.1</td>
<td>10.4</td>
</tr>
<tr>
<td><strong>HEP-</strong></td>
<td>32.9</td>
<td>31.3</td>
<td>32.1</td>
<td></td>
<td>34.7</td>
<td>25.0</td>
<td>29.0</td>
</tr>
<tr>
<td><strong>HEP</strong></td>
<td>36.5</td>
<td>39.2</td>
<td><strong>37.8</strong></td>
<td></td>
<td>44.5</td>
<td>40.5</td>
<td><strong>42.4</strong></td>
</tr>
</tbody>
</table>

Non-graph models (FEM, LDA, GRU): high precision but very low recall

HEP: good balance and highest F1
Outline

- Data and Problem
- Overall Framework
- Proposed Model
- Experiments
- Conclusion
Conclusion

- Heterogeneous interaction graph
  - Able to capture interactions between items
  - Able to mitigate the sparsity issue

- Heterogeneity challenge
  - Node types
  - Edge features

- Heterogeneous embedding propagation
  - Classification loss
  - Reconstruction loss