Little is Much: Bridging Cross-Platform Behaviors through Overlapped Crowds

Meng Jiang¹, Peng Cui¹, Nicholas Jing Yuan², Xing Xie², Shiqiang Yang¹
¹ Department of Computer Science and Technology, Tsinghua University, Beijing, China
² Microsoft Research Asia, Beijing, China

Cross-Platform Behavior Prediction
We register the Uber application with our Facebook accounts. So, can we improve the behavior prediction accuracy on Uber with the rich social data?

The key challenge in prediction is sparsity. Knowledge transfer from auxiliary data can alleviate the sparsity problem.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Bridge</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-Source</td>
<td>Non-overlapped</td>
<td>Codebook</td>
</tr>
<tr>
<td></td>
<td>→ User cluster × Item cluster</td>
<td></td>
</tr>
<tr>
<td></td>
<td>→ The same latent representation</td>
<td></td>
</tr>
<tr>
<td>Cross-Domain</td>
<td>Fully overlapped users</td>
<td>CST</td>
</tr>
<tr>
<td></td>
<td>OR Fully overlapped items</td>
<td></td>
</tr>
<tr>
<td></td>
<td>→ User vector or Item vector</td>
<td></td>
</tr>
<tr>
<td></td>
<td>→ The same latent representation</td>
<td></td>
</tr>
<tr>
<td>Cross-Platform</td>
<td>Partially overlapped users</td>
<td>XPTrans</td>
</tr>
<tr>
<td></td>
<td>→ User vector</td>
<td></td>
</tr>
<tr>
<td></td>
<td>→ Different latent representations</td>
<td></td>
</tr>
</tbody>
</table>

XPTrans Framework
Semi-supervised Transfer Learning

Overlapping user similarity across platforms as flexible regularization

Objective Function
Target platform
Auxiliary platform

\[ J = \sum_{i,j} W_{i,j}^{(p)} (R_{i,j}^{(p)} - \sum_{t} \alpha_{t} v_{i,t}^{(p)} v_{j,t}^{(p)})^2 \]

Overlapping user similarity (Pair-wise regularization)

Experiments
Sina Weibo Douban tags/entities movies/books

C: Overlapping users; A and B are NOT.
Q: Can we transfer the auxiliary big data A to improve the performance on sparse data B as good as richer but small data C?

Transfer but the Same Latent Space Size

NO Transfer
Transfer via the Same Latent Space

<table>
<thead>
<tr>
<th>Problem</th>
<th>Method</th>
<th>Weibo tweet entity to Douban movie</th>
<th>Douban book to Weibo social tag</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>MAP</td>
</tr>
<tr>
<td>Cross-Source</td>
<td>CST</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.236</td>
<td>0.805</td>
</tr>
<tr>
<td>Cross-Domain</td>
<td>CST</td>
<td>B</td>
<td>1.439</td>
</tr>
<tr>
<td>Cross-Platform</td>
<td>CST</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.722</td>
<td>0.820</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.38%</td>
<td>+17%</td>
</tr>
</tbody>
</table>

Performance
Different sizes of Latent Space

Acknowledgement
Contact: Meng Jiang
mjjiang89@gmail.com
Dr. Jiang is now a postdoc at UIUC.