A novel approach for venue recommendation using cross-domain techniques

Pablo Sánchez, Alejandro Bellogín

pablo.sanchezp@uam.es, alejandro.bellogin@uam.es

Venue Recommendation: Traditional Evaluation

- Two common approaches: consider each city as an independent dataset (a) or every check-in of many cities as one dataset (b) [1, 2].

- Option a: it allows to isolate behavior on one city, but no external information can be exploited.
- Option b: by training once, many different cities can be evaluated, but no control about dominant cities is possible.

Venue Recommendation as Cross-Domain

- We propose to consider each city as an independent domain, using one target domain (test) and many source domains (training).
- Best options to learn and transfer knowledge? Our proposals: use most popular cities (more data) or closest cities (more overlap).

Experiments and Results

- Dataset: 33M Foursquare check-ins. Temporal split: 6 months for training, 1 month for test.
- Recommenders: closest venues (AvgDis), hybrid (PGN), UB, IB, HKV, MF with geographical information (IRenMF).
- Results using NDCG@5.

<table>
<thead>
<tr>
<th>City</th>
<th>AvgDis</th>
<th>PGN</th>
<th>UB</th>
<th>IB</th>
<th>HKV</th>
<th>IRenMF</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-CD</td>
<td>0.001</td>
<td>0.068</td>
<td>0.073</td>
<td>0.057</td>
<td>0.071</td>
<td>0.079</td>
</tr>
<tr>
<td>△(%)</td>
<td>-9.7</td>
<td>1.6</td>
<td>0.3</td>
<td>-3.2</td>
<td>2.0</td>
<td>-14.8</td>
</tr>
<tr>
<td>P-CD</td>
<td>0.001</td>
<td>0.068</td>
<td>0.073</td>
<td>0.059</td>
<td>0.068</td>
<td>0.052</td>
</tr>
<tr>
<td>△(%)</td>
<td>-0.1</td>
<td>0.9</td>
<td>0.4</td>
<td>0.0</td>
<td>-1.4</td>
<td>-24.7</td>
</tr>
<tr>
<td>Nearest Cross-Domain (N-CD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most-Popular Cross-Domain (P-CD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conclusions and Future Work

- Using Cross-Domain techniques in venue recommendation improves the performance of many recommenders.

- Selecting the cities by proximity is a good strategy to improve the results, confirming that better data is more useful than more data. “Everything is related to everything else, but near things are more related than distant things” [3].

- Future: explore different ways to select cities and exploit categorical information.

References


Source code available at: https://bitbucket.org/PabloSanchezP/TempCDSeqEval