Sequential-based recommendation

Session-based recommendation deals with recommending items that fit in a sequence, for example, a next place to visit or a next song to listen.

In session-based recommendation all we know about the user is their current session. The task is to predict the next item that best fits the current session.

Problem

Our paper investigates the following questions in session-based recommendation:

Are there method-specific limitations to performance? The design of simpler algorithms allows us to directly calculate the upper bound on prediction accuracy.

How random or predictable are user actions in different datasets? Are datasets theoretically highly predictable or are there those that are "hopeless"?

Definitions

Predictability is the probability that the recommender correctly predicts the next item. Prediction accuracy metrics assume that the recommender outputs a list of top n items, with the most likely item ranked first.

HR@n Hit ratio - ratio of test cases, where the correct item was predicted among top n results.

MRR@n Mean of the reciprocal rank of the item in top n (0 if rank exceeds n), emphasizing the ranking position of the correct item.

Algorithmic limits

The graph to the right shows some approaches for finding items to recommend. The current session is {B, C, F}, with the bold outline marking the current item and the color fill the session items in training data. The Markov chain method using maximum likelihood could only recommend G. Sequential rule method can also "see" H and association rules would include E. Neighborhood-based methods that find similar sessions can also recommend the item D. The presence or absence of relations in training data determines whether a test case is solvable or not. Finding the ratio of solvable cases gives the limit on predictability.

Datasets:

- RSC15 - online retail, RecSys’15 Challenge[1].
- TMALL - online retail, Tmall.com.
- RETAILR - online shopping, Retail Rocket[4].
- AOTM - music playlists, Art of the Mix[3].
- 30MUSIC - music playlists, last.fm[6].
- NOWPLAYING - "now playing" tweets[7].
- CLEF - news article reads[2].

Table: predictability limits due to algorithm design

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>MC</th>
<th>SR</th>
<th>AR</th>
<th>INN</th>
<th>SKNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOTM</td>
<td>0.015*</td>
<td>0.042*</td>
<td>0.079</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>NOWPLAYING</td>
<td>0.17*</td>
<td>0.28</td>
<td>0.40</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>30MUSIC</td>
<td>0.29*</td>
<td>0.36*</td>
<td>0.48</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>TMALL</td>
<td>0.25*</td>
<td>0.40</td>
<td>0.45</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>RETAILR</td>
<td>0.30*</td>
<td>0.44*</td>
<td>0.57*</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>RSC15</td>
<td>0.79</td>
<td>0.86</td>
<td>0.90</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td>CLEF</td>
<td>0.96</td>
<td>0.98</td>
<td>0.99</td>
<td>0.995</td>
<td></td>
</tr>
</tbody>
</table>

The limits in the table apply to the HR@n metric, for any n. The starred entries are below known benchmark results at HR@20[2].

Key observations:

- MC, SR and AR/INN are severely restricted on sparser music datasets, especially AOTM.
- Limits are less relevant for CLEF. It is a dataset where there are abundant training samples, relative to number of items.
- SKNN (any variant) is much less restricted by design, with the limit above 80% in all cases.

Future work should examine whether it is possible to find predictability limits with more sophisticated recommendation methods - neural networks and matrix factorization.

Predictability limits due to randomness

Habits, trends and other regularities in human behavior suggest that we can anticipate the activities of single users or groups of similar users. However, in session-based recommendation the users are anonymous. Some reasons behind user decisions remain outside the parameters of datasets. Mathematically, we model such unknowns as randomness. The maximum predictability of a sequence \( \pi \) can be derived using the concept of the entropy rate of a sequence[5].

We found the limits to be significantly above the best known results for each dataset[2]. The limit values (between 44% and 73%) apply to the HR@1 metric, so they have direct practical implications in recommendation scenarios where the top predicted item is important. Future work should investigate estimating limits for HR@n where n > 1.

References


Further information

The source code of the experiments and the data access instructions are available online: https://github.com/priitj/recsys19