Predictability Limits in Session-based Next Item Recommendation

Priit Järv

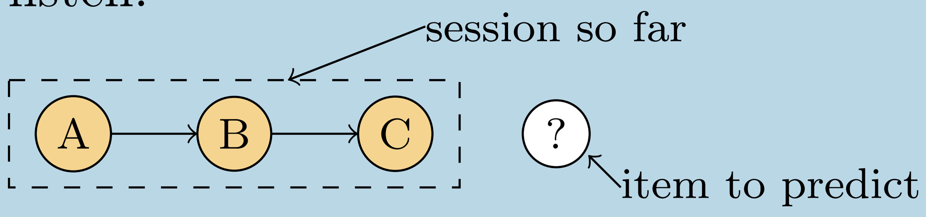
Tallinn University of Technology

priit@whitedb.org

European Union
European Regional Development Fund
Investing in your future

Session-based recommendation

Sequential 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 rate - 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), emphasising the ranking position of the correct item.

References

[1] David Ben-Shimon, Alexander Tsikinovsky, Michael Friedmann, Bracha Shapira, Lior Rokach, and Johannes Hoerle. RecSys Challenge 2015 and the YOOCHOOSE Dataset. *9th ACM Conference on Recommender Systems, Vienna, Austria, 2015*

[2] Malte Ludewig and Dietmar Jannach. Evaluation of session-based recommendation algorithms. *User Model. User-Adapt. Interact. 28, 4-5* (2018)

[3] Brian McFee and Gert R. G. Lanckriet. 2012. Hypergraph Models of Playlist Dialects. *13th International Society for Music Information Retrieval Conference, Porto, Portugal, 2012*

[4] Retailrocket. Retailrocket recommender system dataset. *Kaggle* (2017)

[5] Gavin Smith, Romain Wieser, James Goulding, and Duncan Barrack. A refined limit on the predictability of human mobility. *IEEE International Conference on Pervasive Computing and Communications, Budapest, Hungary, 2014*

[6] Roberto Turrin, Massimo Quadrana, Andrea Condorelli, Roberto Pagano, and Paolo Cremonesi. 30Music Listening and Playlists Dataset. *9th ACM Conference on Recommender Systems, Vienna, Austria, 2015*

[7] Eva Zangerle, Martin Pichl, Wolfgang Gassler, and Günther Specht. #nowplaying Music Dataset: Extracting Listening Behavior from Twitter. *First International Workshop on Internet-Scale Multimedia Management, Orlando, Florida, USA, 2014*

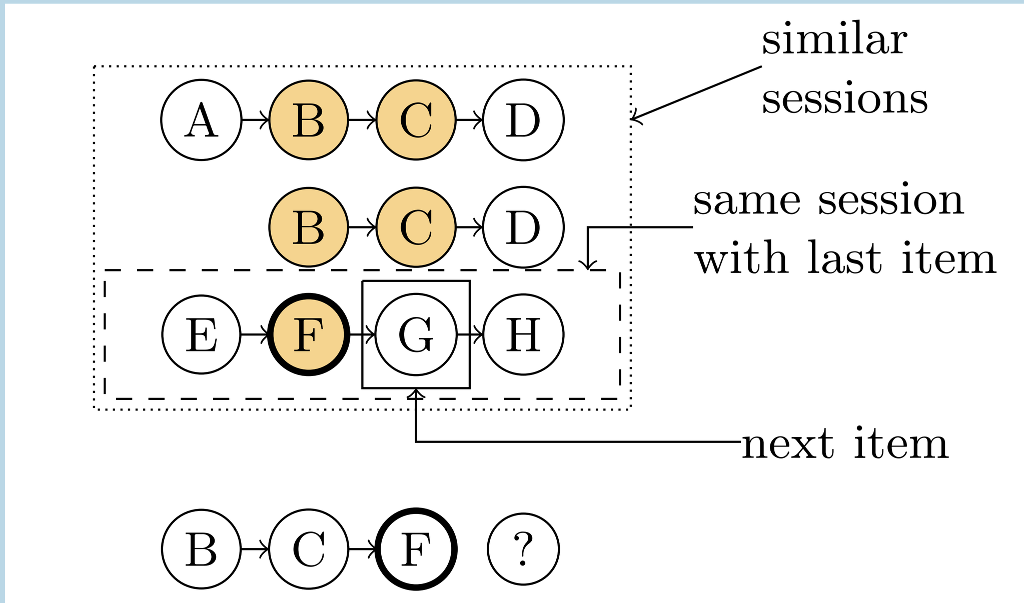
Further information

The source code of the experiments and the data access instructions are available online:

https://github.com/priitj/recsys19

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].

Algorithms:

MC - Markov chain.
AR - Association rules.
SR - Sequential rules.
IKNN - Item-based k-nearest neighbors.
SKNN - Session-based k-nearest neighbors.

Table: predictability limits due to algorithm design

	MC	SR	AR, IKNN	SKNN
AOTM	0.015*	0.042*	0.079	0.92
NOWPLAYING	0.17*	0.28	0.40	0.90
30MUSIC	0.29*	0.36*	0.48	0.91
TMALL	0.25*	0.40	0.45	0.90
RETAILR	0.36*	0.44*	0.57*	0.80
RSC15	0.79	0.86	0.90	0.97
CLEF	0.96	0.98	0.99	0.995

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].

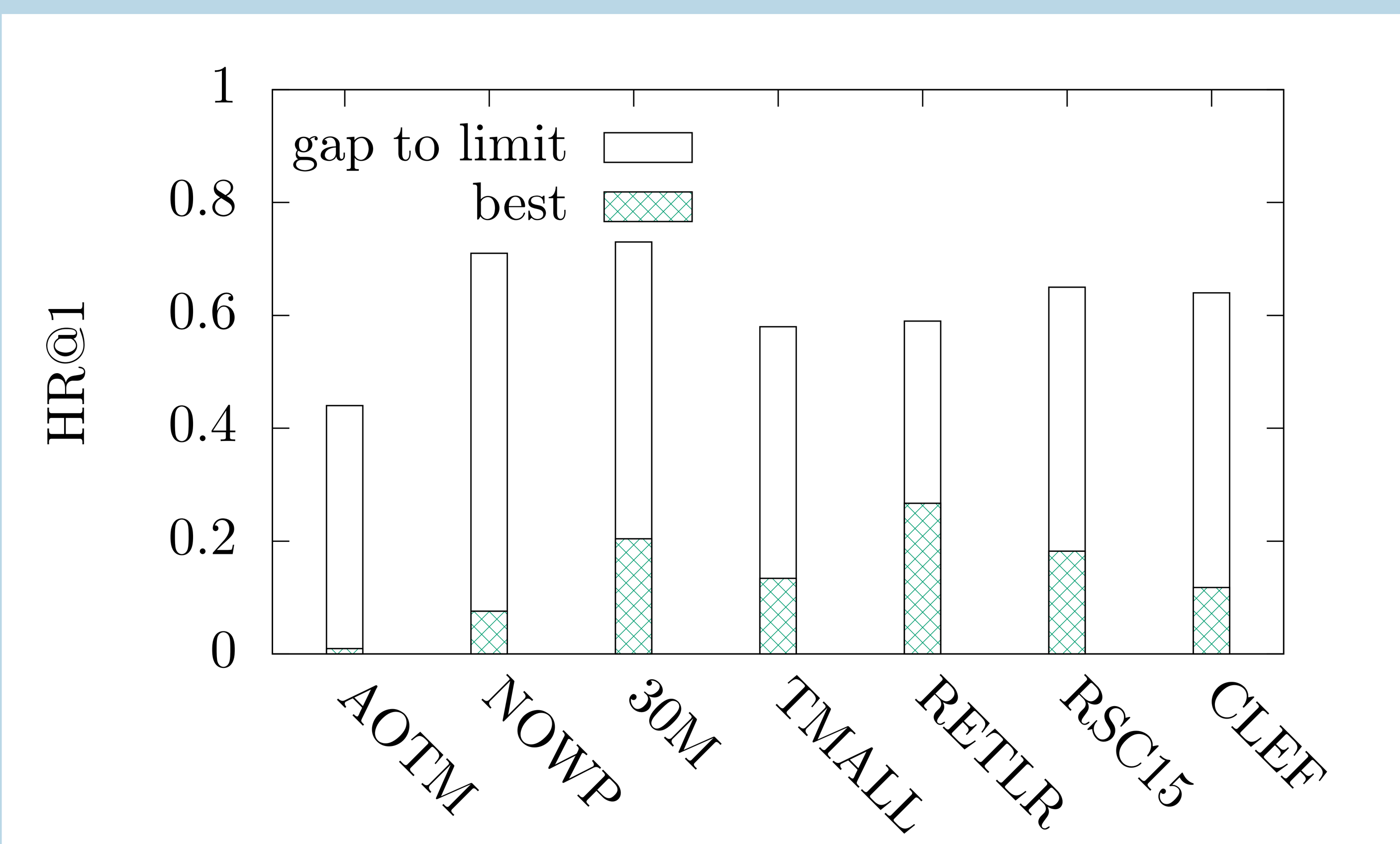
Key observations:

- MC, SR and AR/IKNN 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 Π^{max} 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 $n > 1$.