Sequence Recommender with Item and User Embeddings

Lara Thompson
Principle Data Scientist @ Salesforce
lara.thompson@salesforce.com

April 7, 2023

Abstract

The quality of recommendations in certain domains depend heavily on the sequence of interactions between users and items. User tastes and abilities evolve in time; for items that rely on a skill level, such as bouldering, a good sequence of items (boulder problem) can be key in skill development. In this paper, I use a transformer model that attentively accounts for sequence ordering that jointly trains user and item embeddings and incorporates additional features easily. Trained using final item masking has a heldout prediction accuracy nearing 25% for BC boulder problems (compared to using collaborative filtering). This is on held out users for whom their embeddings aren't trained; the predictions for such users is nonetheless improved by adding user embeddings.

1 Introduction

Content-based recommenders search for similar items to those a user already likes; collaborative filtering suggests items liked by users with similar tastes. In some domains, the sequence of items is important, e.g. we would not recommend a phone after the case for that phone is purchased. User preferences can slowly evolve, e.g. in fashion or reading tastes.

Sequential recommendations aim to personalize recommendations while capturing the current context with recent items. Temporal recommendations account for the actual time elapsed between interactions rather than just the sequence. There's a growing trend in research addressing "session-based" recommendations: these consider the sequence of actions within a single session, generally without any supplemental user information. A multi session recommender considers both session interactions and longer term interactions.

In a skill-based domain such as bouldering, skill and strength acquisition is evident from the sequence of climbs: a climber may at first only be able to do "easy"/"straightforward" climbs before they progress to harder and harder ones. Superposed is a slower evolution of user tastes: from climbs with obvious large holds ("jugs"), to problems with harder holds, either tiny edges ("crimps") or "slopers", or climbs with neither ("slab"). For a recommender, capturing these separate dimensions of user tastes and skills is challenging.

In bouldering, each session can have a different focus. A "project" session involves warm up "easy" climbs, followed by harder projects (which may or may not be climbed successfully, aka "sent"), ending with cool down "easy" climbs again. Alternatively, there are training sessions where a climber seeks out a circuit of problems in a specific style to improve their skills and strength in that style. To complicate data collection, most climbers will only log new sends – that is, they'll only log a boulder problem the first time they climb it – and some climbers won't log warm up and cool down climbs at all.

A bouldering recommender can ideally account for session focus, skill and strength progression as well as evolving style preferences. This paper uses the transformer architecture and analyze what features it can incorporate to improve its sequence recommendations. The sendage dataset (described later) used is relatively small; model parsimony is important.

2 Prior Literature

Early sequence recommenders combined matrix factorization with Markov decision processes to capture long term user preferences and short term sequence patterns (Rendle et al., 2010)

GRU4Rec /+ (Hidasi et al., 2015; Hidasi and Karatzoglou, 2017)

SAS4Rec (Kang and McAuley, 2018)

SSE-PT (Wu et al., 2020)

BERT4Rec (Sun et al., 2019)

Transformers4Rec (de Souza Pereira Moreira et al., 2021)

Paralleling: (Mikolov et al., 2013), GPT-2 (Radford et al., 2019), BERT (Devlin et al., 2019), using Transformers (Vaswani et al., 2017)

3 Data

All models were trained on data from the climbing website sendage.com created and maintained by Jamie Chong, a developer and climber based in Vancouver. It's similar to the larger more well known site 8a.nu, but the climb information is far more heavily currated creating a much cleaner dataset with accurate location data for climbs and practically no duplicates.

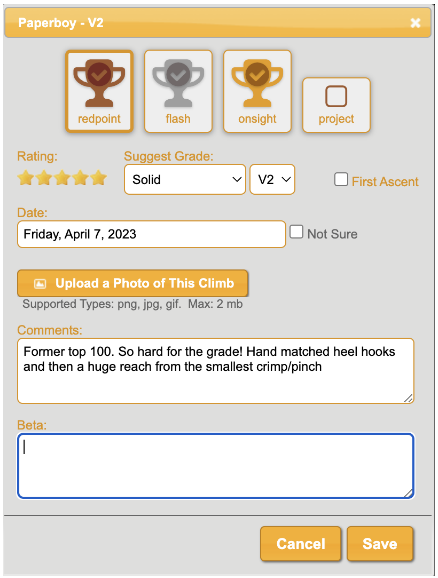

Figure 1: For each send, the boulderer logs the date, any comments and/or "beta" used (the holds, moves or sequence to climb the boulder); they can rate the problem 1-5 stars and give a "feels-like" grade. The date defaults to the current date, so we can expect a few days error in some logged climbs if users do not manually set the actual date the problem was climbed. The default grade is the mean grade of previous logged sends; this biases the user heavily to agree to already proposed grades.

I limited my analysis to boulders, as opposed to sport climbs and trad climbs, because they are more heavily tracked on the site. The most popular boulder problem, Superfly, has nearly 700 "sends" vs 240 ascents of the most popular sport route, Rug Munchers; 45 boulders have more logged sends than Rug Munchers. The most popular trad route, Diedre, has only been logged 186 times, despite there being a queue of parties waiting for Diedre when the weather clears.

To ensure enough data for boulder problems and boulderers alike, I filtered down to a self consistent set of problems with at least 5 sends and boulderers that sent at least 5 boulders. For BC, that reduces to 2844 boulder problems, 1069 boulderers that logged 68,336 sends; if I include much of the world, that only increases to 5367 problems, 1428 boulderers with 116,997 sends (the site is mostly used by BC climbers).

Figure 1 shows a sendage window to log a climb. The mean rating is – with – fraction rated; – / – / – fraction have a date, comments or beta, respectively. The mean range of assigned grades to a particular problem is less than 1 V grade; the maximum range is 2 V grades[1]. The most popular grade logged is V4 with a linear drop up to V11 and a few harder sends up to V14; likewise, there are fewer V0-V3 boulders logged: many climbers don't log the "easy" problems.

[1]Bouldering grades in North America follow the "V scale" that was created by John Sherman in the 1980s. A beginner shouldn't be surprised if even V0 feels impossible; there are currently four boulders with the proposed grade V17 and only one of these has seen repeats to confirm the grade.

Likely to be very detailed if the datasets are new or unfamiliar to the community, or if familiar datasets are being used in new ways. Includes prior work on them, statistics, and a collection protocol.

4 Model

Flesh out your own approach, perhaps amplifying themes from the 'Prior lit' section.

5 Methods

The experimental approach, including descriptions of metrics, baseline models, etc. Details about hyperparameters, optimization choices, etc., are probably best given in appendices, unless they are central to the arguments.

Explicitly define the metrics, even the common ones (or at least reference them). Be clear about how the data is split for assessment.

6 Results

A no-nonsense report of what happened.

7 Analysis

Discussion of what the results mean, what they don't mean, where they can be improved, etc. These sections vary a lot depending on the nature of the paper. (For papers reporting on experiments with multiple datasets, it can be good to repeats Methods/Results/Analysis in separate (sub)sections for each dataset.)

8 Conclusion

Quickly summarize what the paper did, and then chart out possible future directions that anyone might pursue.

Known Project Limitations

For this section, imagine that your reader is a well-intentioned NLP practitioner who is seeking to make use of your data, models, or findings as part of a separate scholarly project, deployed system, or some other kind of real-world intervention. What should such a person know about your work? Especially important here are limitations and biases that might affect this person, their findings, their experiment participants, or the users of their product or service. The idea is that what you say here will be taken into consideration but this well-intentioned user, leading to better outcomes for everyone.

Authorship Statement

I worked on this project solo including the literature review, the iterative implementation and experimentation, result gathering and final writing up. I had many fruitful conversations with Alistair Fraser, also taking

the course, and Michael Nathe about my design choices and result analysis.

References

Gabriel de Souza Pereira Moreira, Sara Rabhi, Jeong Min Lee, Ronay Ak, and Even Oldridge. 2021. Transformers4rec: Bridging the gap between nlp and sequential / session-based recommendation. In *Proceedings of the 15th ACM Conference on Recommender Systems*, RecSys '21, page 143–153, New York, NY, USA. Association for Computing Machinery. `https://github.com/NVIDIA-Merlin/Transformers4Rec`.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT 2019*.

Balázs Hidasi and Alexandros Karatzoglou. 2017. Recurrent neural networks with top-k gains for session-based recommendations.

Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2015. Session-based recommendations with recurrent neural networks.

Wang-Cheng Kang and Julian McAuley. 2018. Self-attentive sequential recommendation. pages 197–206.

Tomas Mikolov, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. In *International Conference on Learning Representations*.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners.

Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2010. Factorizing personalized markov chains for next-basket recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, page 811–820, New York, NY, USA. Association for Computing Machinery.

Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, CIKM '19, page 1441–1450, New York, NY, USA. Association for Computing Machinery.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Liwei Wu, Shuqing Li, Cho-Jui Hsieh, and James Sharpnack. 2020. Sse-pt: Sequential recommendation via personalized transformer. In *Proceedings of the 14th ACM Conference on Recommender Systems*, RecSys '20, page 328–337, New York, NY, USA. Association for Computing Machinery.