Sequential and session-based recommendation: Past, present, future

Dietmar Jannach

University of Klagenfurt, Austria

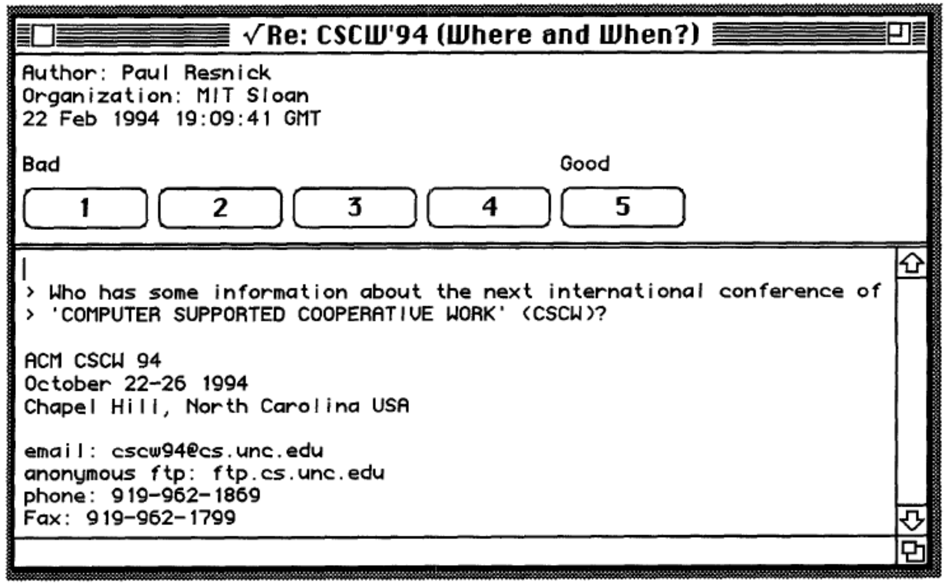
dietmar.jannach@aau.at

Key note at the 6th Workshop on Online Recommender Systems and User Modeling

ACM RecSys 2023, Singapore

History: The GroupLens system

- Presented at CSCW '94
  - A personalized news filtering system
  - Based on explicit ratings in user-item rating matrix
  - Recommendation seen as a "matrix filling" problem
  - Use nearest-neighbors to predict a user's rating for the yet unrated items



Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. 1994. GroupLens: an open architecture for collaborative filtering of netnews. In Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work (CSCW '94). 175-186.

History: Matrix completion

- Many years of algorithm research based this operationalization
  - Developments boosted by the Netflix Prize (2006-2009)
  - Increasingly sophisticated learning algorithms, matrix factorization, ensemble methods

- Shift of focus at the end of Netflix Prize
  - Importance of implicit feedback in practical settings
  - Learning-to-rank instead of pointwise predictions

J. Bennett, and S. Lanning. Proceedings of the KDD Cup Workshop 2007 , page 3--6. New York, ACM, (August 2007)

History: Matrix completion

- Characteristics of plain matrix completion settings
  - Recommendations are based on all available feedback signals of a user
  - Recommendations cover the entire spectrum of past user interests
- Limitations soon became apparent
  - Some older signals may be outdated (interest drift)
    - Development of time-aware recommendation algorithms
  - The user context may matter
    - Development of context-aware recommendation algorithms

Campos, P.G., Díez, F. & Cantador, I. Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols. User Model User-Adap Inter 24, 67–119 (2014).

Adomavicius, G., Tuzhilin, A. (2015). Context-Aware Recommender Systems. In: Ricci, F., Rokach, L., Shapira, B. (eds) Recommender Systems Handbook. Springer, Boston, MA

Sequential patterns

- Time-aware and context-aware methods:
  - Still based on matrix completion
- In many domains, however: sequential patterns exit, e.g.:
  - Buying accessories for a just purchased “main” item in e-commerce
  - Listening to a series of songs from the same artist
  - Watching related videos on a streaming service
  - Visiting a close-by Point-of-Interest when traveling
- Recommender systems should take these patterns into account
  - And maybe focus on the last few user interactions only

Sequential patterns: History

- Early approaches (mid 2000s), e.g.,
  - Next web-page recommendation
    - Based on sequential pattern mining
  - Music (playlist) recommendation
    - Based on item similarity
  - E-commerce
    - Based on Markov Decision Processes
    - (Or simply: Customers who bought … also bought)

B. Mobasher, H. Dai, T. Luo, and M. Nakagawa. Using sequential and non-sequential patterns in predictive web usage mining tasks. In Proceedings of IEEE International Conference on Data Mining, ICDM '02, pages 669–672, **2002**.

R. Ragno, C. J. C. Burges, and C. Herley. Inferring similarity between music objects with application to playlist generation. In Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval, MIR '05, pages 73–80, **2005**.

G. Shani, D. Heckerman, and R. I. Brafman. An MDP-based recommender system. The Journal of Machine Learning Research, 6:1265–1295, **2005**.

Sequential and session-based recommendation

- In many of these application settings, no long-term user history exists
  - Users may even be anonymous
  - Only available information is obtained from the current session
    - Typically implicit feedback signals: purchases, item views etc.

- Leads to session-based recommendation problems
  - User history is organized in (activity) sessions
  - Sessions consist of sequentially ordered observations

Terminology

- Some inconsistency in the literature
  - Sequential recommendation, session-based recommendation, session-aware recommendation
  - In common: Problem is to predict the next user action
- Here:
  - Session-based recommendation
    - Only anonymous interactions of an ongoing session are known
  - Session-aware recommendation
    - User history organized in sessions with user ID: past sessions of a user are known
  - Sequential recommendation
    - User history sequentially ordered, but not organized in sessions

History: A new era since 2015

- Until 2015: Only a few session-based algorithms proposed
- Enteraing a new era: ACM RecSys 2015 Challenge
  - Focused on session-based recommendation problem
  - Release of the YOOCHOOSE dataset
    - Item view and purchase events, still widely used today
  - Goal was to predict:
    - Which users will buy something?
    - Which items will users buy?
  - Winning strategy 2015
    - Feature engineering, classification, gradient boosting

Peter Romov and Evgeny Sokolov, RecSys Challenge 2015: ensemble learning with categorical features, Proceedings of the 2015 International ACM Recommender Systems Challenge, September 2015 Article No.: 1, Pages 1–4

History: GRU4Rec

- **2016:** The landmark method GRU4Rec was proposed
  - Probably the first “neural” session-based method
    - Based on Recurrent Neural Networks
    - Several technical features incorporated, including Gated Recurrent Units
    - Further optimizations proposed later on
  - Evaluated on YOOCHOOSE data and a proprietary dataset
    - Best baseline: Item-KNN
      - Considers items that are similar to “actual” item
      - Similarity is based on the cosine similarity of the sessions
  - GRU4Rec is still widely used today in academia and industry

B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk. Session-based recommendations with recurrent neural networks. In Proceedings International Conference on Learning Representatinos, ICLR ’16, 201

Since 2016

- Countless neural session-based method proposed every year
  - All sorts of technical architectures explored, e.g., convolution, attention, graph-based approaches
  - Sometimes side information is used, e.g., categories
- A smaller number of session-aware methods also published
  - E.g., based on GRU4REc
- Many sequential models put forward (not our focus here)
  - E.g., SASREc and BERT4Rec

Massimo Quadrana, Alexandros Karatzoglou, Balázs Hidasi, and Paolo Cremonesi. 2017. Personalizing Session-based Recommendations with Hierarchical Recurrent Neural Networks. In Proceedings of the Eleventh ACM Conference on Recommender Systems (RecSys '17).

Wang-Cheng Kang, Julian McAuley (2018). Self-Attentive Sequential Recommendation. In Proceedings of IEEE International Conference on Data Mining (ICDM'18)

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)

Own journey, since 2013

- Music
  - Worked on playlist generation problems
    - Found that a session-based nearest-neighbor method worked well for the top positions in the recommendations
    - High recall obtained when focusing on popular items
- E-commerce:
  - Received rich dataset from Zalando, a European fashion retailer
    - Found that focusing on the last session is highly beneficial
    - Used content features to boost performance
    - Proposed an evaluation protocol

Jannach, D., Lerche, L. and Gdaniec, M.: "Re-ranking recommendations based on predicted short-term interests - A protocol and first experiment". In: Proceedings of the Workshop on Intelligent Techniques for Web Personalization and Recommender Systems (ITWP 2013 at AAAI 2013)

Bonnin, G. and Jannach, D.: "A Comparison of Playlist Generation Strategies for Music Recommendation and a New Baseline Scheme". In: Proceedings of the Workshop on Intelligent Techniques for Web Personalization and Recommender Systems (ITWP 2013 at AAAI 2013).

Own journey, since 2013

- Continued work on Zalando dataset
  - Considered various aspects when recommending
    - **Short-term user intents**: Focus on actions in last session
    - **Reminders**: Studied the value of repeated recommendations
    - **Trends:** Considered global item popularity trends
    - **Discounts:** Ranked items up which are currently on sale
  - Found that:
    - All signals are helpful
    - A session-based nearest-neighbor baseline ranker works best

Jannach, D., Lerche, L. and Jugovac, M.: "Adaptation and Evaluation of Recommendations for Short-term Shopping Goals". In: Proceedings of the 9th ACM Conference on Recommender Systems (RecSys 2015). Vienna, Austria, 2015, pp. 211-218

Jannach, D., Ludewig, M. and Lerche, L.: "Session-based Item Recommendation in E-Commerce: On Short-Term Intents, Reminders, Trends, and Discounts". User-Modeling and User-Adapted Interaction, Vol. 27(3-5). Springer, 2017, pp. 351-392

Lerche, L., Jannach, D. and Ludewig, M.: "On the Value of Reminders within E-Commerce Recommendations". In: Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization, (UMAP 2016). Halifax, Canada, 2016.

2017: GRU4Rec vs. Nearest Neighbors

- Motivation:
  - Find out how session-based nearest neighbors fared against GRU4Rec
    - Using original code and data
- Surprising findings:
  - GRU4Rec actually not better than nearest neighbors, but outperformed
  - Combining GRU4Rec with kNN can be synergistic
  - Subsequent in-depth study:
    - Improved kNN methods proposed, again outperforming improved GRU4Rec

Jannach, D. and Ludewig, M.: "When Recurrent Neural Networks meet the Neighborhood for Session-Based Recommendation". In: Proceedings of the 11th ACM Conference on Recommender Systems (RecSys 2017). Como, Italy, 2017

Ludewig, M. and Jannach, D.: "Evaluation of Session-based Recommendation Algorithms". User-Modeling and User-Adapted Interaction, Vol. 28(4-5). Springer, 2018,

Own journey since 2017 (on this topic)

- Performed several algorithmic comparisons
  - Including six recent deep learning methods as well as conceptually simpler methods
- Again, surprising findings:
  - In almost all cases, the simple methods are favorable in terms of accuracy compared to sophisticated models
- Our conclusion
  - Deep learning methods are certainly promising, but simple methods should be considered as baselines to ensure progress

Ludewig, M., Latifi, S., Mauro, N. and Jannach, D.: "Empirical Analysis of Session-Based Recommendation Algorithms". User Modeling and User-Adapted Interaction, Vol. 31(1). 2021, pp. 149–181

Ludewig, M., Mauro, N., Latifi, S. and Jannach, D.: "Performance Comparison of Neural and Non-Neural Approaches to Session-based Recommendation". In: Proceedings of the 2019 ACM Conference on Recommender Systems (RecSys 2019). Copenhagen, 201

Own journey: Reproducibility and Progress

- Related observations made for classical recommendation settings
  - Almost all neural methods of that time (2019) were outperformed by existing non-neural methods
    - E.g., Matrix factorization, nearest neighbors, linear models
  - Leads to a certain stagnation in the field

- Identified reasons in our research methodology
  - Choice of baselines
  - Limited tuning of baselines, other methodological problems
  - (Limited reproducibility in general)

Ferrari Dacrema, M., Cremonesi, P. and Jannach, D.: "Are We Really Making Much Progress? A Worrying Analysis of Recent Neural Recommendation Approaches". In: Proceedings of the 2019 ACM Conference on Recommender Systems (RecSys 2019). Copenhagen, 2019

Ferrari Dacrema, M., Boglio, S., Cremonesi, P. and Jannach, D.: "A Troubling Analysis of Reproducibility and Progress in Recommender Systems Research". ACM Transactions on Information Systems, Vol. 39(2). 2021

Own journey: Impact?

- Change needs time, methodological anti-patterns still common
- Recent studies:
  - Session-aware recommendation:
    - Every single examined neural method outperformed by extended nearest neighbor approach
  - Streaming session-based recommendation:
    - Nearest neighbor technique outperforms sophisticated models published recently at top venues such as SIGIR and KDD
  - Session-based approaches based on Graph Neural Networks (ongoing)
    - Signs that nearest neighbors are competitive as well

Latifi, S., Mauro, N. and Jannach, D.: "Session-aware Recommendation: A Surprising Quest for the State-of-the-art". Information Sciences, Vol. 573. 2021, pp. 291-315

Latifi, S. and Jannach, D.: "Streaming Session-Based Recommendation: When Graph Neural Networks meet the Neighborhood". In: 16th ACM Conference on Recommender Systems (RecSys '22). 2022

Own journey: Impact?

- Recent studies:
  - Sequential recommendation:
    - Benchmarked BERT4Rec on four datasets against nearest neighbors
    - Mixed results:
      - BERT4Rec works significantly better than kNN on two larger datasets (finally)
      - kNN is better on the small datasets
    - Side observation:
      - The use of sampled metrics may lead to misleading results

Latifi, S., Jannach, D. and Ferraro, A.: "Sequential Recommendation: A Study on Transformers, Nearest Neighbors and Sampled Metrics". Information Sciences, Vol. 609. 2022, pp. 660 – 678

Walid Krichene and Steffen Rendle. 2020. On Sampled Metrics for Item Recommendation. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '20).

Own journey: Impact?

- Visibility and impact
  - Works that show the competitiveness of kNN methods reach some visibility (in terms of citations)
  - Competitive results confirmed by a number of other papers, and applied in industry

- Still, kNN-based baselines largely ignored
  - Reviewed 34 papers from SIGIR, RecSys, WSDM etc.
    - **None of them** uses session-kNN as baselines
  - Weaker baselines like item-KNN however considered

Diksha Garg, Priyanka Gupta, Pankaj Malhotra, Lovekesh Vig, and Gautam Shroff. 2019. Sequence and Time Aware Neighborhood for Session-based Recommendations: STAN. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'19)

Barrie Kersbergen, Olivier Sprangers, and Sebastian Schelter. 2022. Serenade - Low-Latency Session-Based Recommendation in e-Commerce at Scale. In Proceedings of the 2022 International Conference on Management of Data (SIGMOD '22)

The present

- It might feel a bit depressing
  - As the problem is well known, also outside the RecSys field
  - Also given that there was a RepSys Workshop held 10 years ago at RecSys
  - And no signs of a major change/crisis are visible

- Our publication and research culture changes slowly
  - Fast pace of publication, publication pressure
  - Focus on technology, methodology seems secondary

Timothy G. Armstrong, Alistair Moffat, William Webber, and Justin Zobel. 2009. Improvements that don't add up: ad-hoc retrieval results since 1998. In Proceedings of the 18th ACM conference on Information and knowledge management (CIKM '09).

Cremonesi, P. and Jannach, D.: "Progress in recommender systems research: Crisis? What crisis?". AI Magazine, Vol. 42(3). 2021, pp. 43-54

The present

- But there are signs of awareness and change
  - More and more conferences have reproducibility tracks
  - Sharing code and data becomes a review criterion
  - A number of frameworks are available (although care must be taken)
- An increasing number of critical/reflective papers are published
  - See also the program of RecSys ‘23

Balázs Hidasi and Ádám Tibor Czapp, The effect of third party implementations on reproducibility, RecSys ’23

Balázs Hidasi and Ádám Tibor Czapp, Widespread flaws in offline evaluation of recommender systems, RecSys ‘23

Zhu Sun, Di Yu, Hui Fang, Jie Yang, Xinghua Qu, Jie Zhang, and Cong Geng. 2020. Are We Evaluating Rigorously? Benchmarking Recommendation for Reproducible Evaluation and Fair Comparison. In Proceedings of the 14th ACM Conference on Recommender Systems (RecSys '20).

Emanuele Cavenaghi, Gabriele Sottocornola, Fabio Stella, and Markus Zanker. 2023. A Systematic Study on Reproducibility of Reinforcement Learning in Recommendation Systems. **ACM Trans. Recomm. Syst. (TORS)**

The present (and future)

- Overcoming methodological issues over time
  - Requires further increased attention and awareness in the community
  - Better education is needed
    - Undergraduates, PhDs, professors, reviewers, grant committees
  - Example: The need to the hyperparameters of all models on all datasets
    - Barely done in the literature in a systematic way
      - So from most studies, nothing can be concluded
    - Code for preprocessing, baseline algorithms, for hyperparameter tuning almost never shared
    - Topic not mentioned much in textbooks

Bauer, C., Fröbe, M., Jannach, D., Kruschwitz, U., Rosso, P., Spina, D. and Tintarev, N.: "Overcoming Methodological Challenges in Information Retrieval and Recommender Systems through Awareness and Education". In: Dagstuhl Seminar 23031: Frontiers of Information Access Experimentation for Research and Education. 2023

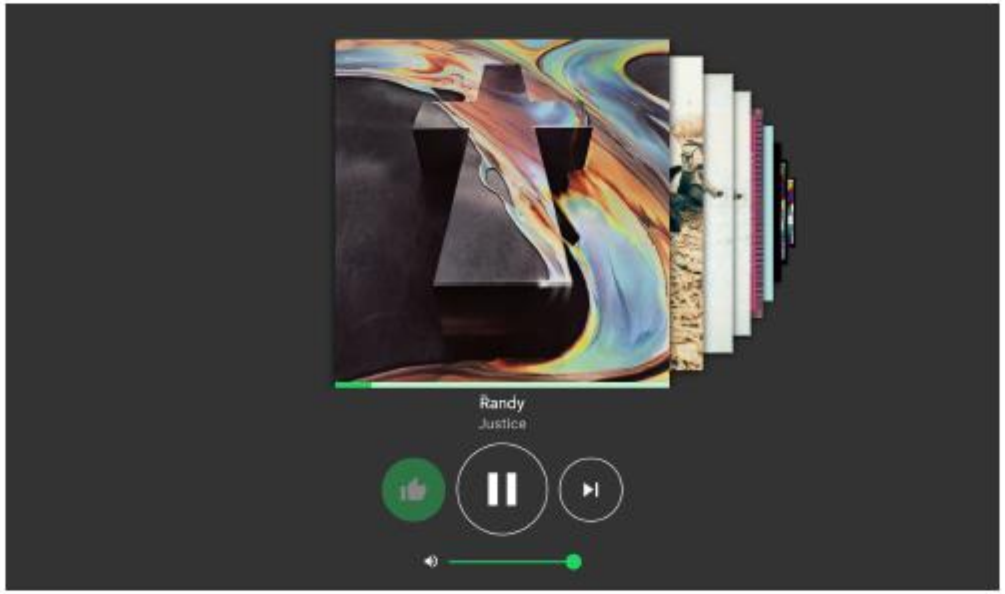
Shehzad, F. and Jannach, D.: "Everyone's a Winner! On Hyperparameter Tuning of Recommendation Models". In: 17th ACM Conference on Recommender Systems. 2023

Where to go from here?

- Putting methodological issues aside:
  - Countless session-based/sequential algorithms are available today
  - As common in RecSys and ML in general: Hyperfocus on offline accuracy
- What are interesting open problems to work on?
  - A subjective selection
    - User studies, longitudinal effects, integration of additional information, real-world studies
  - Most aspects are not limited to session-based algorithms

Kiri L. Wagstaff. 2012. Machine learning that matters. In Proceedings of the 29th International Conference on International Conference on Machine Learning (ICML'12).

User perception of session-based recommendations



Ludewig, M. and Jannach, D.: "User-Centric Evaluation of Session-Based Recommendations for an Automated Radio Station". In: Proceedings of the 2019 ACM Conference on Recommender Systems (RecSys 2019). Copenhagen, 2019

Ludewig, M., Latifi, S., Mauro, N. and Jannach, D.: "Empirical Analysis of Session-Based Recommendation Algorithms". User Modeling and User-Adapted Interaction, Vol. 31(1). 2021, pp. 149–181

User perception study

- It is well known that “being accurate” is not enough
  - High offline accuracy might neither correspond to user nor to business value
    - High hitrate/recall may be achieved be recommending tracks by popular artists, e.g., in the music domain
    - In Point-of-Interest recommendation it is meaningless to recommend the same restaurant that the user goes to every Friday anyway
    - In many domains, discovery (support) is a key purpose of the recommender system
  - Offline evaluation of diversity, novelty etc may be insufficient as well
    - Need to involve the human in the (evaluation) loop
    - **Almost no user studies exist in the literature**

Sean M. McNee, John Riedl, and Joseph A. Konstan. 2006. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06).

User perception study

- A study in the music domain
  - Assess how users perceive the recommendation quality in different dimensions
- Experimental setup:
  - Develop an online application for study participants to interact with
  - Participants select a start track and the application creates and plays a playlist
  - Participants can skip or like tracks, leading to updates of the playlist
  - Participants fill out a questionnaire at the end

User perception study

- Different recommendation algorithms tested
  - Simple association rules AR ("customers who bought")
  - Collocated Artists Greatest Hits (CAGH)
  - GRU4REC: An RNN-based method
  - S-KNN: A session-based nearest neighbor method
  - SPOTIFY: Recommendations were retrieved only through Spotify's API

User interface

User Study Results

- Number of Likes:
  - From 4.48 (Spotify) to 6.48 (AR)
  - Even though AR received the most likes, the recommendations received, on average, the lowest rating scores
  - Reason: Many 1-star ratings for apparently bad recommendations
  - Some insights:
    - Optimizing for likes can be misleading, also avoid bad recommendations

- Popularity of recommendations:
  - Spotify and GRU4REC with the least popular / novel recommendations
  - Popularity highly correlates with number of Likes

User Study Results

- Selected questionnaire results:
  - S-KNN recommendations were often considered a good match for the selected seed tracks
  - AR works poorly in many dimensions
  - Spotify excelled in terms of discovery
  - Intention to reuse:
    - Interestingly, participants stated that they will re-use the Spotify-based system despite the higher novelty and the very low prediction accuracy

Offline Results

- Spotify's algorithm would have very likely been ruled out from an A/B test based on offline accuracy
- Humans in the loop are essential in the evaluation

Algorithm	P@5	R@5	HR@5	MRR@5
S-KNN	**0.271**	**0.044**	0.137	0.077
GRU4REC	0.161	0.028	**0.151**	**0.096**
AR	0.234	0.037	0.135	0.081
CAGH	0.172	0.024	0.052	0.026
SPOTIFY	0.009	0.001	0.002	0.001

Longitudinal effects of session-based recommendations

- Most existing research focused on one-shot evaluation
  - Based on single train-test split
  - Longitudinal effects not assessed
    - E.g., would higher novelty in the recommendations to lead to changed user consumption patterns in the long run?
    - E.g., to what extent do different algorithms have a reinforcing tendency, e.g., to recommend items that are already popular
  - In reality, long term effects are central

Ferraro, A., Jannach, D. and Serra, X.: "Exploring Longitudinal Effects of Session-based Recommendations". In: Proceedings of the 2020 ACM Conference on Recommender Systems (RecSys '20). 2020

Study on longitudinal effects

- Compared several session-based algorithms in a simulation-based approach
  - GRU4Rec, NARM, S-KNN, CAGH
  - Inspired by earlier work on the topic on matrix completion setup

- Simulation loop (sketch)
  - Train model
  - Make recommendations and measure metrics
  - Assume some recommendations are accepted, add new simulated feedback data to training set

Ferraro, A., Jannach, D. and Serra, X.: "Exploring Longitudinal Effects of Session-based Recommendations". In: Proceedings of the 2020 ACM Conference on Recommender Systems (RecSys '20). 2020

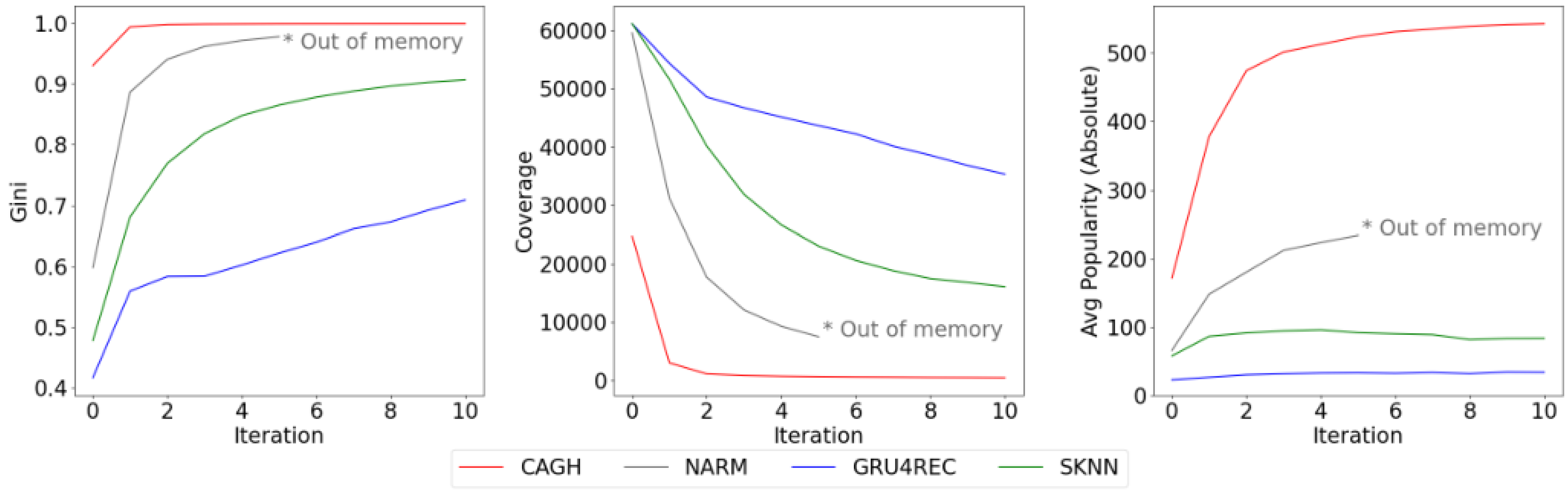
Jannach, D., Lerche, L., Kamehkhosh, I. et al. What recommenders recommend: an analysis of recommendation biases and possible countermeasures. User Model User-Adap Inter 25, 427–491 (2015).

Metrics in longitudinal study

- Different metrics considered
  - Catalog Coverage
  - Gini index (tendency to concentrate on a few items)
  - Average item popularity
- Accuracy results (first round)
  - S-KNN among top-performing models on two data sets

Results of longitudinal study

- Over time, all algorithms deterioriate
  - But to a different extent
  - GRU4Rec exhibits the most "favorable" behavior
  - The other neural NARM model, is worse than S-KNN



Insights (Longitudinal Studies)

- Different algorithms may have very different global effects in the long run
  - These phenomena cannot be derived from offline accuracy results
  - Choice of algorithm should take such longitudinal effects into account

- More research on longitudinal studies required
  - E.g. based on agent-based simulation approaches
  - May also help to assess provider profit vs. consumer relevance trade-off

Zhang, Jingjng and Adomavicius, Gediminas and Gupta, Alok and Ketter, Wolfgang, Consumption and Performance: Understanding Longitudinal Dynamics of Recommender Systems via an Agent-Based Simulation Framework (May 17, 2019). Information Systems Research

Ghanem, N., Leitner, S. and Jannach, D.: "Balancing Consumer and Business Value of Recommender Systems: A Simulation-based Analysis". E-Commerce Research and Applications, Vol. 55. 2022, pp. 101195

Side information and external information in session-based recommendations

- Most research is based on “item views” and no side information
  - And thus predict the next viewed items
- In reality
  - Various types of additional information available:
    - Interaction types, e.g., item view, purchase, add-to-cart, add-to-wishlist
    - Side information, e.g., item meta-data and prices, context information (e.g., time), as well as global trends
- Various avenues for future research

Bauer, J. and Jannach, D.: "Hybrid Session-aware Recommendation with Feature-based Models". User Modeling and User-Adapted Interaction, Vol. online first. 2023

Harte, J., Zorgdrager, W., Louridas, P., Katsifodimos, A., Jannach, D. and Fragkoulis, M.: "Leveraging Large Language Models for Sequential Recommendation". In: 17th ACM Conference on Recommender Systems (Late Breaking Results). 2023

Leveraging heterogenous data

- Own recent work (I)
  - Combining session-based models with feature-based models
    - Use of tabular data and Gradient Boosting
    - Dozens of features engineered
  - Item features, user and session-specific features
    - Item prices, price differences, categories, item recency, item popularity, interaction times, …
  - Two stage-model, leading to substantial performance increase
    - Baseline ranker (e.g., GRU4Rec or S-KNN)
    - Feature-based ranking

Bauer, J. and Jannach, D.: "Hybrid Session-aware Recommendation with Feature-based Models". User Modeling and User-Adapted Interaction, 2023

Leveraging heterogenous data

- Own recent work (II)
  - Leverage Large Language Models (LLMs) for Sequential Recommendations
    - Use OpenAI API
  - Different approaches explored
    - Find semantically related items via LLM encodings
      - based on product names
    - Prompt-based recommendations on own fine-tuned model
    - Initialize BERT4REc with LLM embeddings (based on product names)

Harte, J., Zorgdrager, W., Louridas, P., Katsifodimos, A., Jannach, D. and Fragkoulis, M.: "Leveraging Large Language Models for Sequential Recommendation". In: 17th ACM Conference on Recommender Systems (Late Breaking Results). 2023

Leveraging heterogenous data

- Own recent work (II)
  - LLM-based sequential recommendations
  - Main results:
    - Performance boost for BERT4Rec (10-15%) only by initializing the embeddings with the LLM
    - Retrieving similar items based on LLM embeddings can give good results in some cases
  - Future work:
    - E.g., consider additional information, such as category names

Real-world studies

- Not many results of real-world studies published
- Notable work by Kouki et al.:
  - Research included both offline analyses, a user study, as well as an evaluation in practice (home improvement domain)
  - kNN method good in offline, but not optimal for the given use case
    - Expert study showed that other algorithms more frequently recommended good alternative options
  - A/B test revealed a 15%-18% increase in CTR and revenue, respectively
    - Compared to an existing third-party solution

Pigi Kouki, Ilias Fountalis, Nikolaos Vasiloglou, Xiquan Cui, Edo Liberty, and Khalifeh Al Jadda. 2020. From the lab to production: A case study of session-based recommendations in the home-improvement domain. In Proceedings of the 14th ACM Conference on Recommender Systems (RecSys '20).

Real-world studies

- Work by Kersbergen et al., in the e-commerce domain
- Offline comparison
  - Improved KNN method vs. neural models: Confirmed performance of KNN methods also in this domain
- Online A/B test
  - Proposed architecture to scale KNN methods
    - Up to 600 requests per sec, 6.5 million items
  - Measured “drastically” increased engagement metric
    - Observed cannibalization effect in one configuration – when only most recent item considerd

Barrie Kersbergen, Olivier Sprangers, and Sebastian Schelter. 2022. Serenade - Low-Latency Session-Based Recommendation in e-Commerce at Scale. In Proceedings of the 2022 International Conference on Management of Data (SIGMOD '22).

Summary

- Sequential and session-based recommendation.
  - A still timely and important problem
- A field that is plagued by methodological issues
  - Like other areas of applied ML
  - Slight signs of improvements visible
- Many interesting open questions to be addressed
  - User value and perception, business value, integration of side information, context- and intent-awareness

Thank you!

- Time for questions, contact: dietmar.jannach@aau.at
- Slides: https://tinyurl.com/orsum2023

Improved context- and intent-awareness

- Intent-awareness
  - Users arrive at the site with a certain intent
    - In music, for example, they sometimes would like to explore; and sometimes listen to their favorites
    - In e-commerce, they may be doing research; or they might already be close to a decision
  - Better ways are needed to correctly guess the user intent
    - E.g., based on contextual information or from other signals such as search terms or category navigation events
    - Requires foundational understanding of possible user intents in a given domain

Komal Kapoor, Vikas Kumar, Loren Terveen, Joseph A. Konstan, and Paul Schrater. 2015. "I like to explore sometimes": Adapting to Dynamic User Novelty Preferences. In Proceedings of the 9th ACM Conference on Recommender Systems (RecSys '15).

Rishabh Mehrotra, Mounia Lalmas, Doug Kenney, Thomas Lim-Meng, and Golli Hashemian. 2019. Jointly Leveraging Intent and Interaction Signals to Predict User Satisfaction with Slate Recommendations. In The World Wide Web Conference (WWW '19).

T. Chen, H. Yin, H. Chen, R. Yan, Q. V. H. Nguyen and X. Li, "AIR: Attentional Intention-Aware Recommender Systems," 2019 IEEE 35th International Conference on Data Engineering (ICDE), Macao, China, 2019,

Online learning and updates

- Streaming session-based recommendation
  - Constantly new items come in
  - How to react in real time?
  - How to balance the explore-exploit dilemma?
- News recommendation as a use case
  - Particularly challenging
    - Short life time of items
    - Huge popularity differences
    - Danger of optimizing the "wrong" (short-term) metrics