Session-aware Recommendation

Dietmar Jannach
TU Dortmund, Germany
dietmar.jannach@tu-dortmund.de

User modeling and recommender systems workshop, Haifa, 2017

Recommender Systems

- Automated recommendations
  - A pervasive part of our online user experience
  - Recommend us shopping items, movies, music, news, friends, jobs, groups or people to follow, restaurants, hotels…

  - Recommendations are often personalized
  - „User modeling“ is a central task in such systems

User Modeling & Recommendation

- Explicit preference statements
  - Indication of preferred topics (Google News) or ratings
  - Provision of strict criteria (e.g., location for a hotel recommender)
- User models are however often automatically derived by observing the user's behavior
  - Which restaurants have you visited in the past?
  - Which other people do you follow?
  - For which hotels did you write reviews?
  - Which kind of music did you listen to yesterday?

- Recommendation task
  - Find objects (items) that match the user preferences

Outline

- Why a common academic problem abstraction can be insufficient
- Defining Sequence-Aware Recommender Systems
- Case Studies
  - Session-aware Recommendation in E-Commerce
    - Considering long- and short-term user models in e-commerce
    - The role of reminders
  - Session-aware Next-Track Music Recommendation
- Outlook

Matrix Completion

- A common problem abstraction
- Given a matrix
  - where rows are users and columns are items, and
  - a number in a cell indicates a preference statement (e.g., ratings) of a user for a certain item
- Compute values for the missing cells
  - Recommend items that have high predicted values

	Item1	Item2	Item3	Item4	Item5
Alice	5	?	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

Does time matter here?

- Mostly no, researchers typically abstract this aspect
- However, consider the (usual) movie domain:
  - Doesn't your taste change over the years?
  - Doesn't the set of suitable movies depend on your current mood (or, generally, your context)?
- A few works on "time-aware" recommenders exist
  - They can, for example,
    - consider interest drift over longer periods of time and
    - look at the user's behavior at a certain point in time,
    - or simply give less weight to older ratings
  - One found that considers the optimal point in time to recommend (i.e, wait for item to be discounted)

Beyond the movie domain

- Being able to predict which kinds of things a certain user generally likes, is important
- However, assume you visit your favorite online shop, and here's what you looked at or purchased during the last weeks



- Now, you return to the shop and browse these items



What to recommend?

- Some plausible options
  - Only shoes or only watches?
  - Mostly Nike shoes?
  - Maybe also some T-shirts?
- Using the matrix completion formulation
  - One trains a model based only on past actions
  - The context of the user's current shopping intent is considered only in "context-aware" recommenders
  - Without the context:
    - The algorithm will probably most recommend only T-shirts and trousers
    - Might not be what you expect

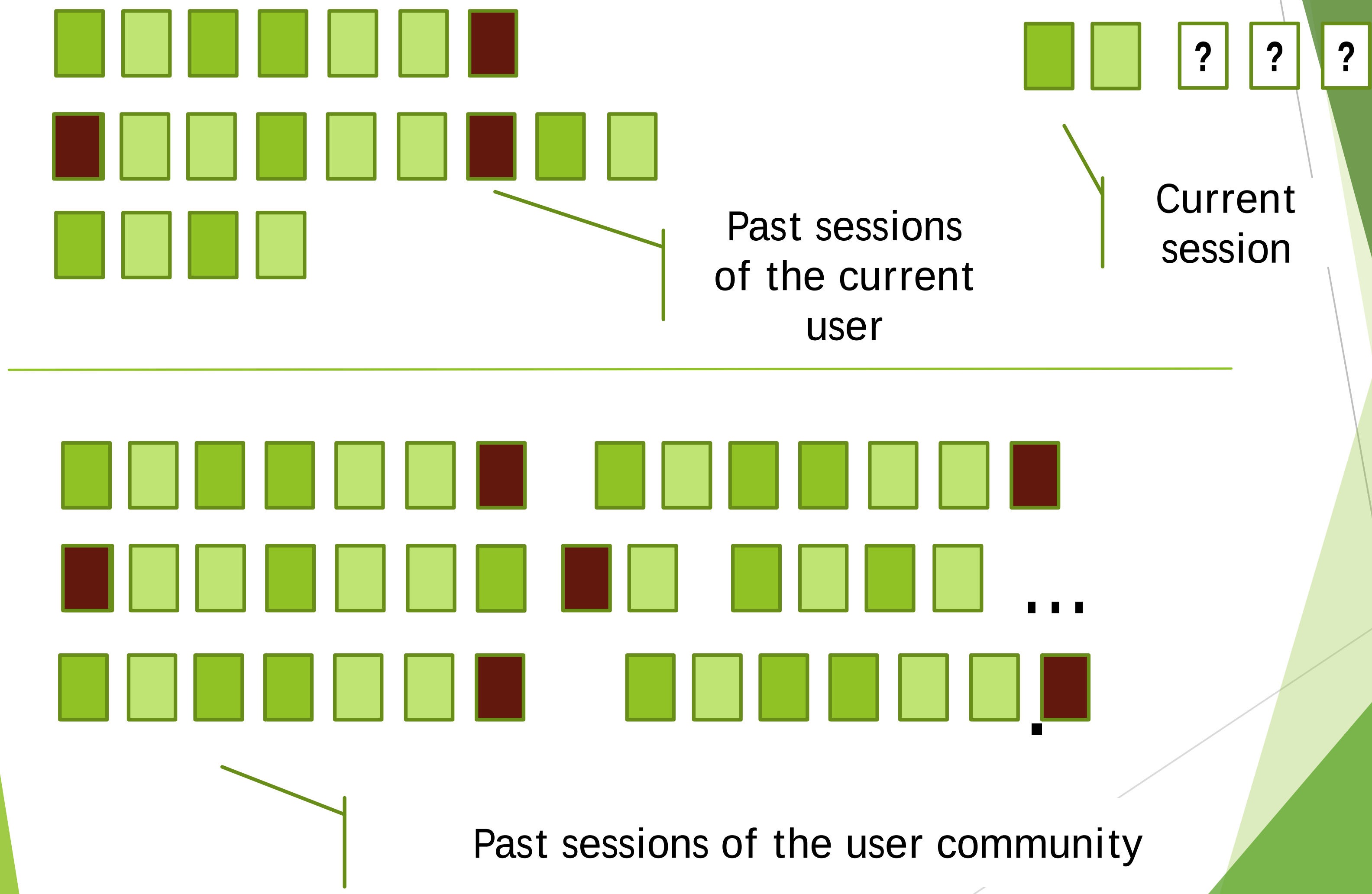
Only in e-commerce?

- No, consider in particular media recommendation
  - Video recommendation on YouTube
  - Music recommendation on Spotify
  - Next-POI recommendation, next-app recommendation for smartphones
  - Next-page recommendation on web sites
- Take YouTube, as an example
  - Seem to change their strategy often-times
  - Past “similar videos” recommendations often very messy, containing out-of-context recommendations
  - Today
    - Main page recommendations cover many topics
    - Similar videos are in fact all similar

Session-aware Recommendation

- Requires a different problem abstraction
  - Has to consider the user's most recent actions
  - But may also utilize past preferences
- Is based on different types of information
  - A sequentially ordered set of past user actions
  - Actions can have different types
- Recommendation task
  - Combine long-term and short-term preference signals predict the next user action
    - Item view, purchase, add-to-cart, watch, listen
  - Sometimes, the order of the actions can be important

General problem setup



Practical challenges

- Generally
  - How to automatically assess the user's current interests?
  - How to combine them with the long-term preference profile?
  - How can we do this in real-time?
- Additional opportunities, as we know more than just item ratings
  - Should we recommend things that the user has inspected last week, but not purchased?
  - Can we utilize individual and general interest drifts?
  - Are there sequentiality constraints to consider
    - Music transitions, recommendation of accessories

Categorization

- Introduction of the family of "sequence-aware" recommender systems
  - Are based on time-ordered log data
  - Different supporting computational tasks, e.g.,
    - Context adaptation
    - Trend detection
    - Repeated recommendation
    - Consideration of ordering constraints
- Context adaptation subcategories
  - Last(-n) item based recommendation
  - Session-based recommendation (short-term only)
  - Session-aware recommendation (long-term, short-term, our focus here)

Technical approaches

- Sequence-learning techniques
  - Frequent pattern mining
    - Frequent item sets, frequent sequential patterns
  - Sequence modeling
    - Markov Models, Recurrent Neural Networks
  - Distributed item representations
    - Distributional and Latent Markov embeddings
- Sequence-aware matrix factorization
- Hybrids
  - Factorized Markov Chains, others

What about Deep Learning?

- Recurrent Neural Networks (RNN) are a “natural” method to deal with ordered data (session data)
- Recent proposal(s) by Hidasi and colleagues
  - Usage of custom RNN Gated Recurrent Units for the problem
  - Evaluation on e-commerce data set
    - Millions of anonymous user sessions
    - Data provided in the ACM RecSys 2015 challenge dataset
  - Significant accuracy improvements over different baseline methods reported

What about Deep Learning?

- However:
  - True value of the method not fully clear
    - Choice of baselines
    - Choice of evaluation protocol
- Recent own experiments
  - Benchmark their method with a session-based kNN method
  - Observations
    - RNNs do not outperform the kNN method, and n most tested configurations they are worse
    - RNNs can be computationally complex, preventing systematic hyperparameter tuning
    - kNN with neighborhood sampling is light-weight and fast
    - Nonetheless: RNNs capture signals that are not covered by the kNN method – hybridization works best

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, forthcoming

Outline

- Why a common academic problem abstraction can be insufficient
- Defining Sequence-Aware Recommender Systems
- Case Studies
  - Session-aware Recommendation in E-Commerce
    - Considering long- and short-term user models in e-commerce
    - The role of reminders
  - Session-aware Next-Track Music Recommendation
- Outlook

Short-term and long-term profiles

- Our general approach:
  - Use two-stage methods
    - Stage 1: Pre-rank recommendable items based on long-term profile
    - Stage 2: Filter or re-rank items based on assumed short-term situation or intents
  - Stage 1 can be offline, stage 2 must be “real-time”
  - Furthermore:
    - Consider various types of data, if available
    - Consider domain-specific quality factors, when relevant

E-commerce case study

- Mainly based on e-commerce log dataset
  - By Zalando, contains sample of user activity logs
    - 1 million purchases; 20 million view events; 170,000 sessions; 800,000 users; 150,000 different items
- Goal of the study
  - Assess the relative importance of short-term and long-term user models
  - Long-term models
    - Used selection of complex and simple methods
    - Bayesian Personalized Ranking, Factorization Machines, Item-item-Nearest-Neighbors, Popularity-based and Random Baselines

Jannach, D., Lerche, L. and Jugovac, M.: "Adaptation and Evaluation of Recommendations for Short-term Shopping Goals". In: RecSys 2015, pp. 211-218

Contextualization Strategies

- Strategies
  - CoOccur
    - “Customers who bought … also bought”
  - CoOccur-Filter
    - Variant, where the ordering is done slightly different
  - Feature Matching (FM)
    - Rank items up when they have features in common with those from the current session (e.g., same brand)
  - Recently Viewed (RV)
    - Recommend recently viewed items in reverse chronological order
- Characteristics
  - All can be applied in real-time
  - Extend short lists with baseline recommendations

Empirical results

- Evaluation method
  - Use parameterizable evaluation protocol (see later)
  - Hit rate (recall) and MRR as evaluation measures
- Observations for dense dataset (example)
  - Recall of best baseline method (BPR): 40%
  - Other:
    - CoOccur : 49%
    - RecentlyViewed : 64%
    - FeatureMatching : 71%
    - Hybrid : 73%

Observations

- Combination of various short-term signals as the most effective strategy
- Choice of baseline is relevant
  - Better baseline in most cases leads to stronger overall results
- Importance of short-term adaptation
  - Contextualization-only methods often already better than the best long-term profile
  - Becomes more and more relevant, the more is known for the current session
  - Do the computational efforts of complex offline models truly pay off?
- Reminding is a very effective strategy

Outline

- Why a common academic problem abstraction can be insufficient
- Defining Sequence-Aware Recommender Systems
- Case Studies
  - Session-aware Recommendation in E-Commerce
    - Considering long- and short-term user models in e-commerce
    - The role of reminders
  - Session-aware Next-Track Music Recommendation
- Outlook

More on reminders

- Follow-up study
  - Deeper analysis of reminders
    - Using again the Zalando dataset
  - Development of more intelligent reminding strategies
  - Evaluation of reminding strategy in field test

Lerche, L., Jannach, D. and Ludewig, M.: *"On the Value of Reminders within E-Commerce Recommendations"*. In: UMAP 2016, 2016.

“Reco-minders” in practice

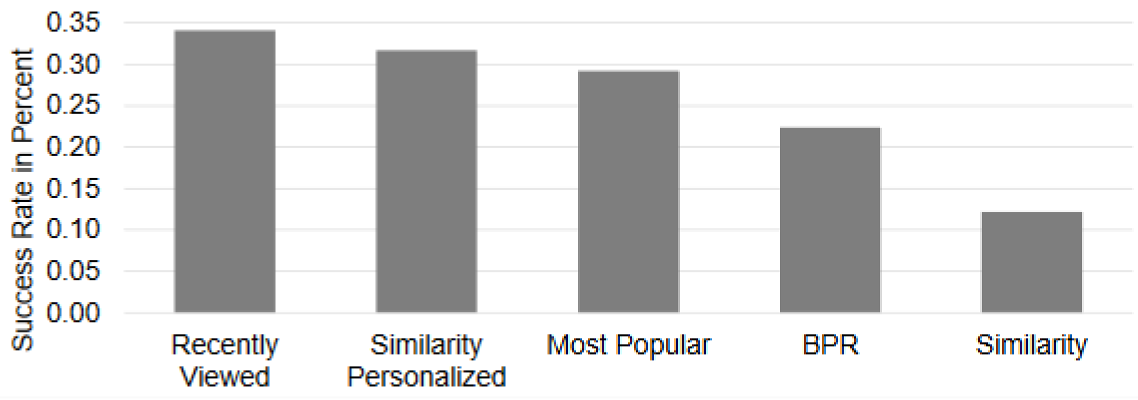
- Log data contains recommendation list for the view events
  - Every 10th recommendation was a reminder
  - More than 40% of the successful recommendations (recommendation clicks leading to purchases) were already known items
    - This also means that recommending unknown items is also very important, and helps users discover things
  - Users inspect an item multiple times before making a purchase
  - During one session, users inspect items of a small set of categories
    - Reminders as navigation shortcuts?

A field study

- A/B-tested different strategies on an e-commerce site for electronic gadgets
- Competing strategies
  - BPR as a learning-to-rank model
  - Similarity-based recommendation (using a reference item)
  - A personalized similarity-based approach
  - Popularity-based baseline
  - Present recently viewed items
    - In reverse chronological order

Field study outcomes

- "Success rate" as business measure
  - Click on recommendation and click on outgoing link to external retailer
  - Pure reminders led to best business value in this specific situation



Can we do better?

- Designed different “adaptive” reminding strategies
  - Recency-based baseline: Use reverse chronological order
  - Intensity-based ranking: Rank reminder items based on the number of past clicks
  - Item-similarity ranking: Select reminder items based on their fit for the current session
  - Session-similarity ranking: Select reminders based in their occurrence in similar past sessions
- General filtering strategy
  - Do not remind users of items in categories where recently a purchase was made

Empirical evaluation

- Done on three different datasets
- Baseline ranking method:
  - A session-based nearest neighbor technique
    - Configured to include reminders as well
  - More accurate than, e.g., BPR
- Parameterizable evaluation protocol
  - Configurable "obviousness gap"
- Results (hit rate, example, 2 evaluation variants)
    - v2 hides view event for target item.
  - Baseline: 0.156
  - Best result v1: 0.697
  - Best result v2: 0.363
- Adaptive reminders better than simple reminders

Outline

- Why a common academic problem abstraction can be insufficient
- Defining Sequence-Aware Recommender Systems
- Case Studies
  - Session-aware Recommendation in E-Commerce
    - Considering long- and short-term user models in e-commerce
    - The role of reminders
  - Session-aware Next-Track Music Recommendation
- Outlook

Music recommendation case study

- Problem setup: Next-track music recommendation
- Given:
  - List of recently listened tracks
  - Optional: history of past user listening sessions
- Recommend:
  - Tracks to listen to next
- Application scenarios
  - Automated radio stations
  - Playlist generation support

Jannach, D., Lerche, L. and Kamehkhosh, I.: "Beyond "Hitting the Hits" - Generating Coherent Music Playlist Continuations with the Right Tracks". In: RecSys 2015, pp. 187-194
Jannach, D., Kamehkhosh, I. and Lerche, L.: "Leveraging Multi-Dimensional User Models for Personalized Next-Track Music Recommendation". In: ACM SAC 2017

Basic recommendation strategies

- Not-so-bad baseline strategies

  (when using the hit rate as evaluation measure)

  (SAGH and CAGH lead to limited discovery)

  - SAGH: Recommend greatest hits of the artists appearing in the “history” of recently listened tracks
  - CAGH: In addition, recommend greatest hits of similar artists
- A simple but competitive strategy
  - kNN: Recommend tracks that appeared in similar listening sessions of other users
  - Outperforms also complex methods other like BPR
  - Existing, but smaller popularity bias

Phase 1: Multi-faceted scoring

- Idea:
  - 1) Determine basic score using kNN
  - 2) Consider variety of other signals
    - Compatibility of musical features for the given playlist, e.g., tempo or loudness
    - “Semantic” fit based on user-provided tags
    - Preferences of social friends
    - Fit according to the long-term preferences
      - Favorite artists
      - Content-based match with own past listening sessions
      - Neighbor sessions of past listening sessions
      - Statistic of previous listening events for same track (reminders)
  - 3) Determine final score as weighted combination

Phase 1: Results

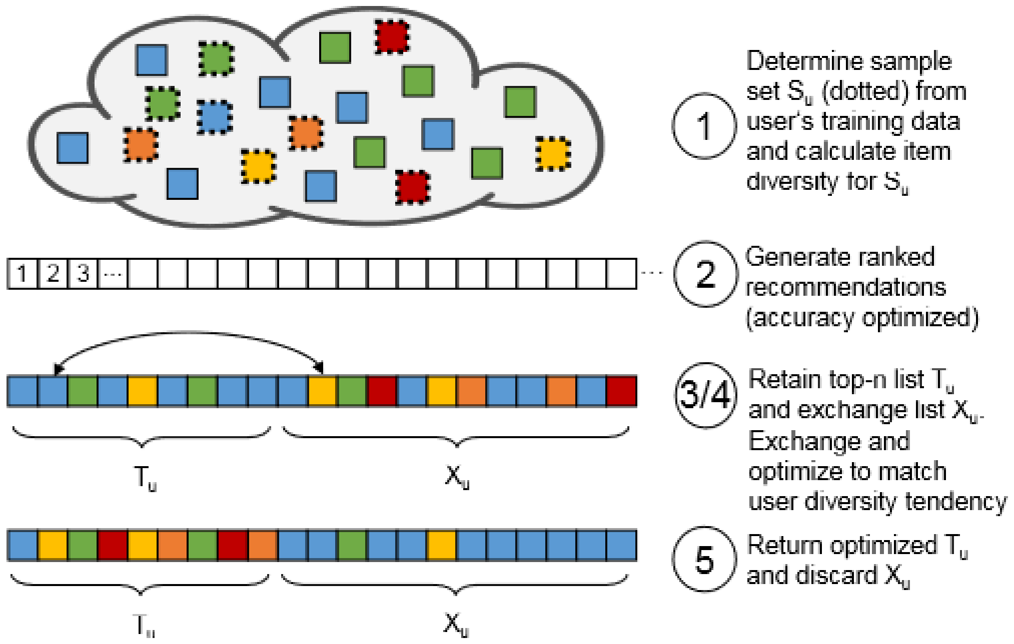
- Evaluation details
  - Different playlist datasets and listening logs
  - Weights determined in a manual process per dataset
  - Focus on information retrieval measures
- Observations
  - Repeated recommendation is advantageous in most cases, in particular when user is not in “exploration mode”
  - Playing favorite artists is good, but leads to lower artist diversity
  - Taking past playlists into account (both in terms of content and track occurrences) is helpful
  - Combining all scores leads to the best results

Phase 2: Greedy re-ranking

- Idea
  - 1) Create ranked list based on multi-faceted scoring technique
  - 2) Re-rank the first few tracks to optimize the user experience
    - Consider long-term quality preference patterns of the individual user (e.g., high diversity)
- Effects
  - Long-term preferences taken into account
  - General accuracy kept at high level
  - Multiple additional factors can be considered in parallel
    - Coherence or diversity of the immediate next few track
    - Coherence with the last few tracks
    - Smooth transitions, e.g., in terms of the tempo

Jugovac, M., Jannach, D. and Lerche, L.: *"Efficient Optimization of Multiple Recommendation Quality Factors According to Individual User Tendencies"*. Expert Systems With Applications. Forthcoming.

Phase 2: Visualization



Phase 2: Observations

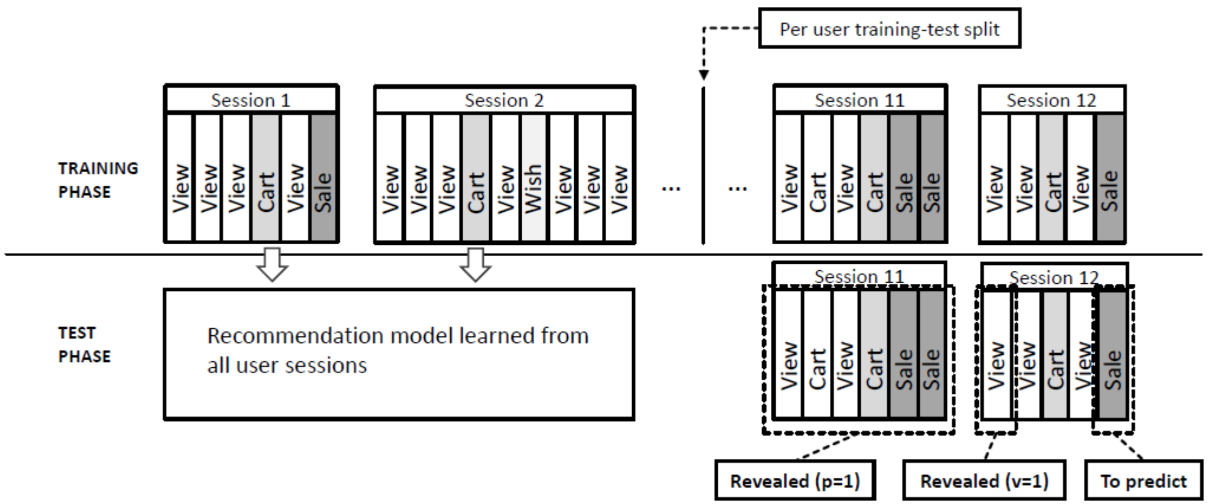
- Accuracy aspects
  - Accuracy depends on the number of tracks to re-rank. Accuracy compromises usually very low
  - In some cases, re-ranking even leads to higher accuracy (in terms of the hit rate)
- Optimization effects
  - Method proves to be effective in various dimensions
  - Multiple optimization goals related to the short-term situation can be considered in parallel
- Performance
  - Computational demands limited due to greedy approach

Outline

- Why a common academic problem abstraction can be insufficient
- Defining Sequence-Aware Recommender Systems
- Case Studies
  - Session-aware Recommendation in E-Commerce
    - Considering long- and short-term user models in e-commerce
    - The role of reminders
  - Session-aware Next-Track Music Recommendation
- Outlook

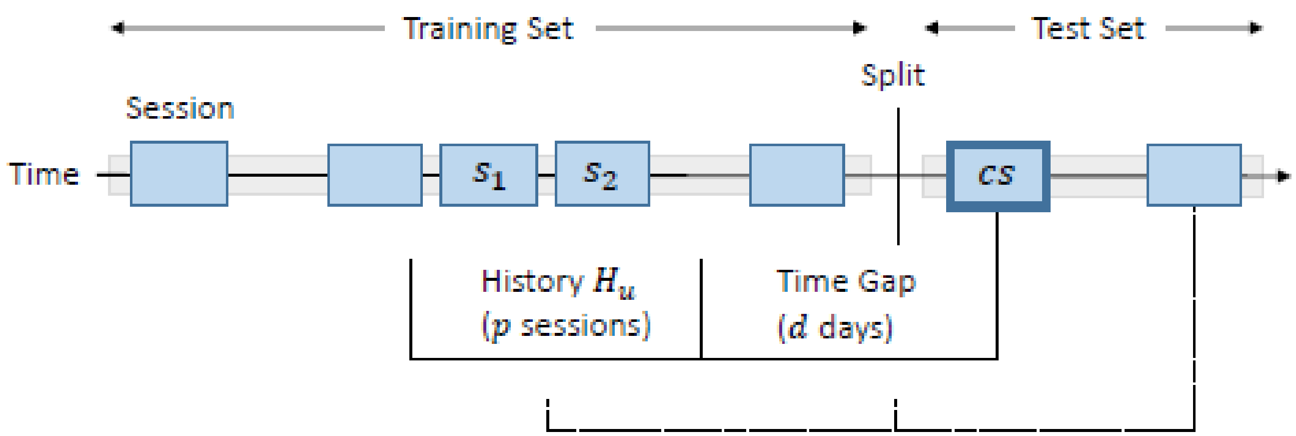
Methodology considerations

- Proposed or applied different procedures and measures for session-aware recommendations
- 1) Protocol for the offline evaluation of session-aware recommendation



Methodology considerations

- 2) Protocol variant for the investigation of reminders



Methodology considerations

- Validating outcomes of offline experiments through user studies
- General problem
  - Assume, we have a new next-track music recommendation method that leads to higher accuracy than previous methods (e.g., in terms of the hit rate)
  - Would such a list be perceived to be of better quality also by end users?
  - In the general field of recommender systems, a number of papers show that this is not necessarily the case

A user study in the music domain

- Selected outcomes of offline experiments in next-track music recommendation:
  - Recommending generally popular tracks of artists in the recent history is strong baseline
  - A session-based kNN method usually leads to very competitive results
  - Considering multiple aspects like the mood of a playlist or the genre within the kNN approach is even better and leads to more homogeneous recommendations (kNN+X)

Kamehkhosh, I. and Jannach, D.: *"User Perception of Next-Track Music Recommendations"*. In: The 2017 Conference on User Modeling Adaptation and Personalization (UMAP 2017).

A user study in the music domain

- Research questions (selection):
  - Does considering additional characteristics like the genre also translate into a higher perceived recommendation quality?
  - What is the quality perception of a method that focuses only on very popular tracks?
  - Can we observe any difference in quality perception when the recommended tracks are already known or new to the study participants?

A user study in the music domain

- Study setup (excerpt)
  - Performed a web-based online experiment
  - Main tasks:
    - Participants had to listen to a playlist beginning (4 tracks), with no track information revealed
    - Then they rated the suitability of continuations that were generated by different algorithms and indicated if they knew the track or artists
  - 277 students participated, leading to 300 trials
    - Participants who did not listen to the tracks long enough were removed from the analysis

Main outcomes

- The kNN method that used additional signals was also consistently better than the pure kNN method
  - The insights from the offline experiment were valid also in terms of the user's quality perception
- Strong differences exist depending on whether the participants knew the tracks or not
  - When considering all trials, the popularity-based method was perceived to lead to better playlists
  - When considering only situations when novel tracks were recommended, the kNN+X method was best
- Side implication
  - Potential familiarity biases in such user studies exist

Jannach, D., Lerche, L. and Jugovac, M.: *"Item Familiarity as a Possible Confounding Factor in User-Centric Recommender Systems Evaluation"*. i-com Journal of Interactive Media, Vol. 14(1). 2015, pp. 29-3

Summary

- Session-aware recommendation is a common problem in real-world application scenarios of recommenders
- A number of algorithmic approaches exist
- Examples of recent works discussed in the e-commerce and music domain
  - Considering both long-term preference models and short-term user preferences can be key to the success of recommenders
  - Reminding users of known items can be useful
- Open issues
  - Evaluation protocol – no standards yet exist
  - Validity of results from offline studies. Not fully clear – presented results of a recent study

- Thank you for your attention
- Contact:
  - dietmar.jannach@tu-dortmund.de

References

- Jannach, D. and Ludewig, M.: *"When Recurrent Neural Networks meet the Neighborhood for Session-Based Recommendation"*. In: Proceedings RecSys 2017. Como, Italy, forthcoming
- Jannach, D., Lerche, L. and Jugovac, M.: "Adaptation and Evaluation of Recommendations for Short-term Shopping Goals". In: RecSys 2015, pp. 211-218
- Lerche, L., Jannach, D. and Ludewig, M.: *"On the Value of Reminders within E-Commerce Recommendations"*. In: UMAP 2016, 2016.
- Jannach, D., Lerche, L. and Kamehkhosh, I.: "Beyond "Hitting the Hits" - Generating Coherent Music Playlist Continuations with the Right Tracks". In: RecSys 2015, pp. 187-194
- Jannach, D., Kamehkhosh, I. and Lerche, L.: "Leveraging Multi-Dimensional User Models for Personalized Next-Track Music Recommendation". In: ACM SAC 2017
- Jugovac, M., Jannach, D. and Lerche, L.: *"Efficient Optimization of Multiple Recommendation Quality Factors According to Individual User Tendencies"*. Expert Systems With Applications. Vol 81, 2017, 321-331.
- Kamehkhosh, I. and Jannach, D.: *"User Perception of Next-Track Music Recommendations"*. In Proceedings UMAP 2017, Bratislava, 2017.
- Jannach, D., Lerche, L. and Jugovac, M.: *"Item Familiarity as a Possible Confounding Factor in User-Centric Recommender Systems Evaluation"*. i-com Journal of Interactive Media, Vol. 14(1). 2015, pp. 29-3