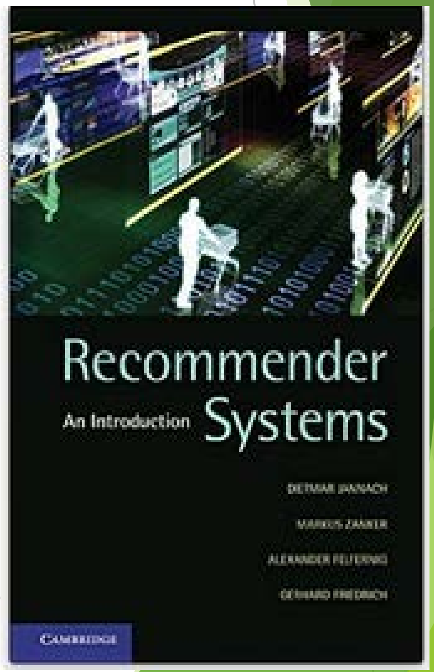
Challenges of Session-aware Recommendation in E-Commerce

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

Talk at the ComplexRec Workshop at ACM RecSys 2017, Como, Italy

About me

- Professor of Computer Science at TU Dortmund, Germany
- Co-founder of a tech company selling interactive selling solutions (2003-2008)
- Research interests
  - Recommender Systems
    - E-Commerce applications, business value of recommenders
    - Interactive advisory systems
  - Artificial Intelligence
    - Model-based Diagnosis, Constraints
  - Software Engineering
    - Debugging of Spreadsheets



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…

Outline

- Why a common academic problem abstraction can be insufficient in practice
- Sequence-Aware Recommender Systems
- An e-commerce case study:
  - On short-term intents, reminders, trends and discounts
- Summary and outlook

Matrix Completion

- A familiar common problem abstraction in academia
- 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

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

Also familiar, but somehow less researched

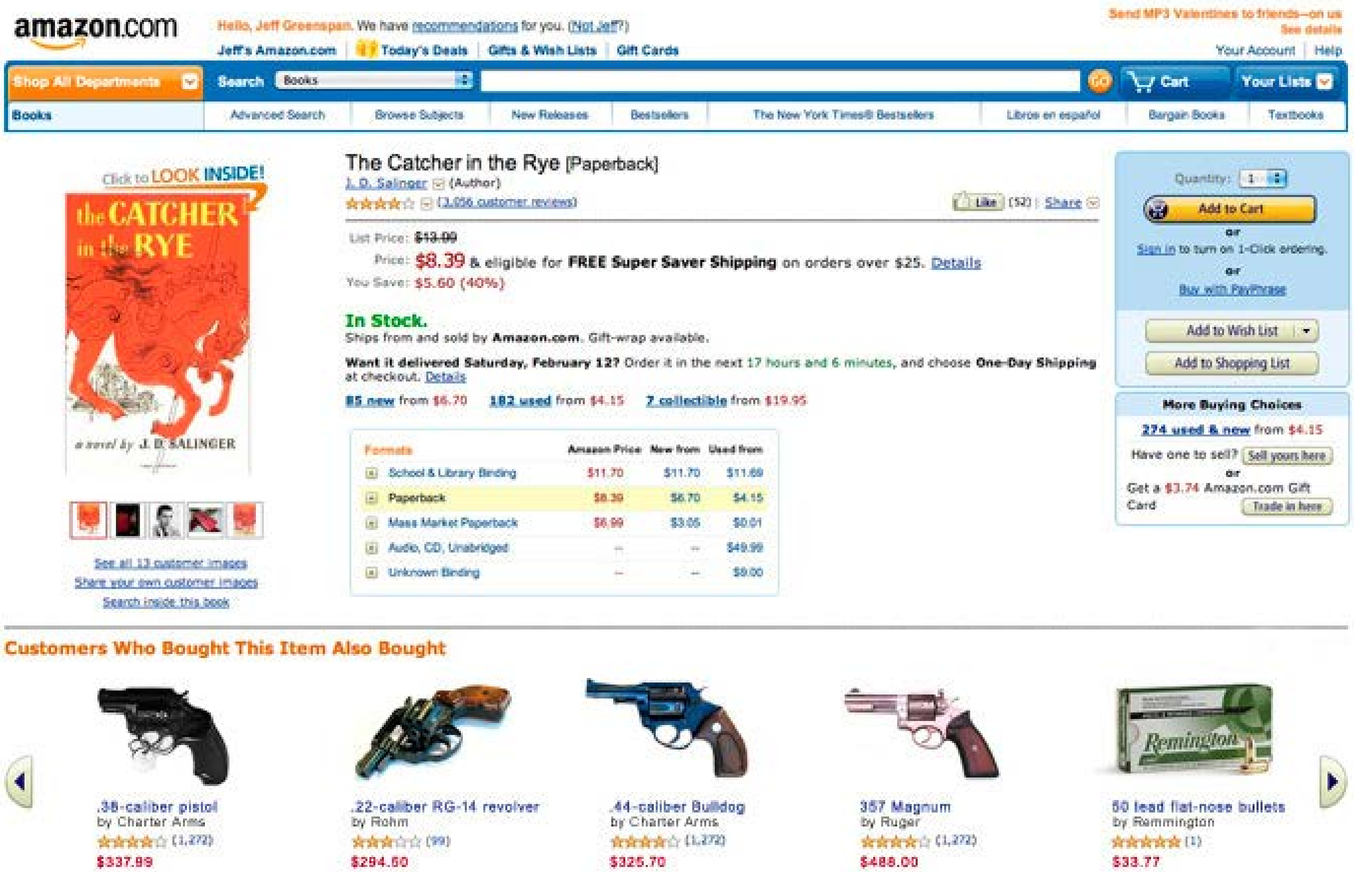
Session-aware recommendation

- Consider the user's most recent interests, their specific goals in their ongoing session, e.g.,
  - the single item that is currently looked at on an e-commerce shop
  - the last five tracks picked by the user on a music listening platform
  - the videos watched last time on a video streaming platform
- Optional:
  - Consider what is generally trending and popular
  - Recommend on-sale items (in e-commerce)

- Not covered by the matrix completion abstraction
- Highly relevant in practice

Jannach, D., Resnick, P., Tuzhilin, A. and Zanker, M.: *"Recommender Systems - Beyond Matrix Completion"*. Communications of the ACM, Vol. 59(11). Association for Computing Machinery (ACM), 2016, pp. 94-102

It can be a complex problem



Customers who bought ...

Roll over image to zoom in

Minnow Sports

Minnow Sports Aluminum Baseball Bat For Baseball & Teeball

8 customer reviews

Price: ~~$29.99~~

Sale: $19.99

You Save: $10.00 (33%)

In Stock.

This item does not ship to **Germany**. Please check other sellers who may ship internationally. Learn more

Sold by BBro Store and Fulfilled by Amazon. Gift-wrap available.

Item Display Length:

32.0 inches

- Made from lightweight high grade Aluminum alloy for faster swing speed
- Ultra-thin 32" handle with All Sports grip for increased stability and accuracy
- Stylish design featuring full rolled-over end for ultimate performance
- Ideal for all levels of baseball players from practice to matches
- 32 inches in length & 24 ounces

Long- and short-term interests

- 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

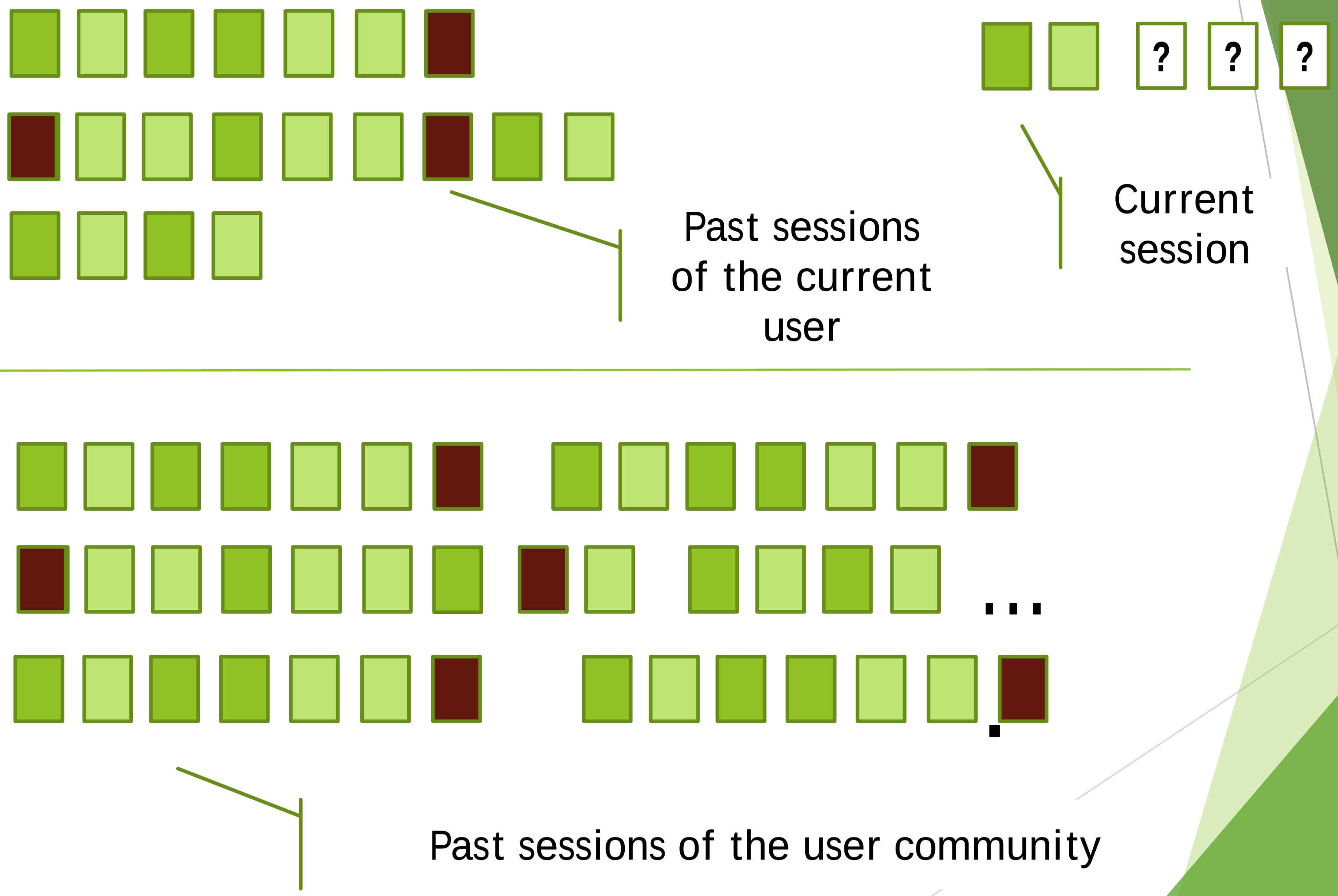
Outline

- Why a common academic problem abstraction can be insufficient in practice
- Defining Sequence-Aware Recommender Systems
- An e-commerce case study:
  - On short-term intents, reminders, trends and discounts
- Summary and outlook

Sequence-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 situation



Categorization

- Introducing the family of “sequence-aware” recommender systems
  - Are based on time-ordered log data
  - Different main and 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)

Existing technical approaches

- Mostly designed for context adaptation

- 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

Outline

- Why a common academic problem abstraction can be insufficient in practice
- Defining Sequence-Aware Recommender Systems
- An e-commerce case study:
  - On short-term intents, reminders, trends and discounts
- Summary and outlook

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. Springer, (forthcoming)

On short-term intents

- Research question:
  - What is the relative importance of adapting recommendations to users' short-term intents (shopping goals) when they visit the site?
- Research approach:
  - "Hide-and-predict" simulation experiments on log data from a large online shop (Zalando)
  - Inputs
    - Behavior of larger user community (including current user)
    - Behavior of user in current session (first *n* actions)
  - Compare capability of session-aware and session-agnostic algorithms of predicting the purchased items in a given session

Technical approach

- Used a two-stage approach to compare session-aware and session agnostic algorithms
  - Stage 1: Learn a long-term user profile
  - Stage 2: Filter or re-rank items based on the assumed short-term situation or intents
- Stage 1 can be done offline
  - Various algorithms were tested
    - Bayesian Personalized Ranking, Factorization Machines, Item-item-Nearest-Neighbors, Popularity-based and Random Baselines
- Stage 2 must be “real-time”
  - Can still be based on complex models, e.g., Recurrent Neural Networks

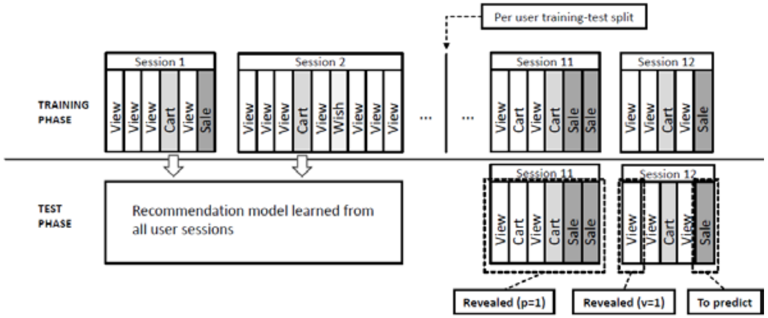
Jannach, D., Lerche, L. and Jugovac, M.: "Adaptation and Evaluation of Recommendations for Short-term Shopping Goals". In: RecSys 2015, pp. 211-218

Contextualization Strategies

- Various comparably simple strategies tested, e.g.,
  - CoOccur
    - “Customers who bought … also bought”
  - 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, method

- No off-the-shelf standard protocol exists
  - e.g., in terms of forming training/test splits
  - Defined a parameterizable evaluation protocol
    - Define how much to reveal of the current and of previous sessions
    - Helps to understand how quickly different strategies are successful to adapt to the short-term goals
  - Hit rate (recall) and MRR as evaluation measures



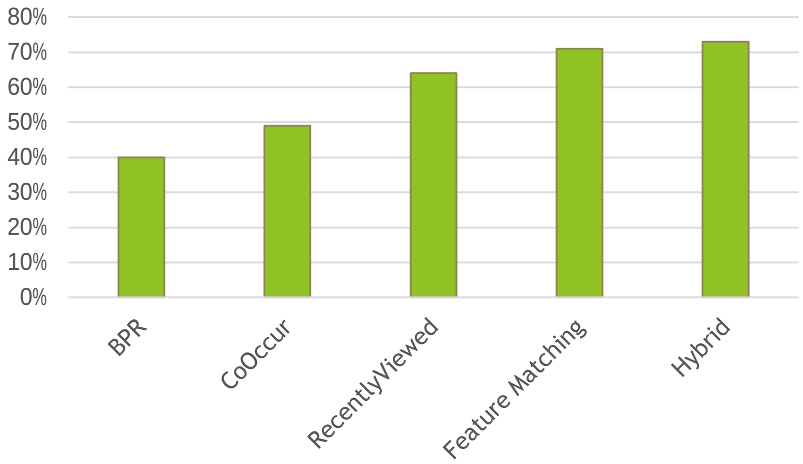
Jannach, D., Lerche, L. and Jugovac, M.: "Adaptation and Evaluation of Recommendations for Short-term Shopping Goals". In: RecSys 2015, pp. 211-218

Dataset

- Evaluations mostly based on an e-commerce log dataset
  - By Zalando, a major European online fashion retailer
  - Dataset contains sample of user activity logs
    - 1 million purchases
    - 20 million view events
    - 170,000 sessions
    - 800,000 users
    - 150,000 different items
- Dataset is very sparse
  - Many users without any purchase

Empirical results

- Observations for dense dataset (example)
  - Recall of best baseline method (BPR): 40%
  - Other:
    - Customers who bought … : 49%
    - Just show me what I have seen : 64%
    - Show me similar things : 71%
    - Combining long- and short-term : 73%



Alternative session-based baselines

- C-KNN:
  - Take the current session and find *n* past sessions that are similar to the current one
  - Do sampling to ensure scalability
  - Performance is competitive with recent neural network approach
- Sequential Rules:
  - “Learn” rules of size two, i.e., count how often two items appeared together in a given sequence in past sessions
  - Apply discount for distance between events
  - Often even better than C-KNN

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

Kamehkhosh, I., Jannach, D. and Ludewig, M.: *"A Comparison of Frequent Pattern Techniques and a Deep Learning Method for Session-Based Recommendation"*. In: Proceedings of the ACM RecSys 2017 Workshop on Temporal Reasoning in User Modeling. Como, Italy,

Observations

- Combination of various short-term and long-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 in practice
- Defining Sequence-Aware Recommender Systems
- An e-commerce case study:
  - On short-term intents, reminders, trends and discounts
- Summary and 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

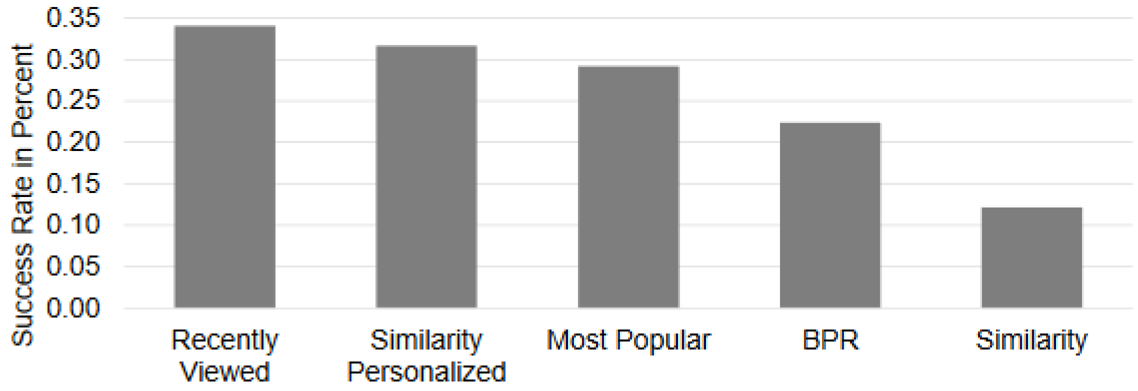
- Log data contains recommendation list for the view events
  - Every 10th recommendation was a reminder
  - More than 40% of the successful recommendations (recommendation clicks leading to purchases) were already known items
    - This also means that recommending unknown items is also very important, 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 on the business value

- 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

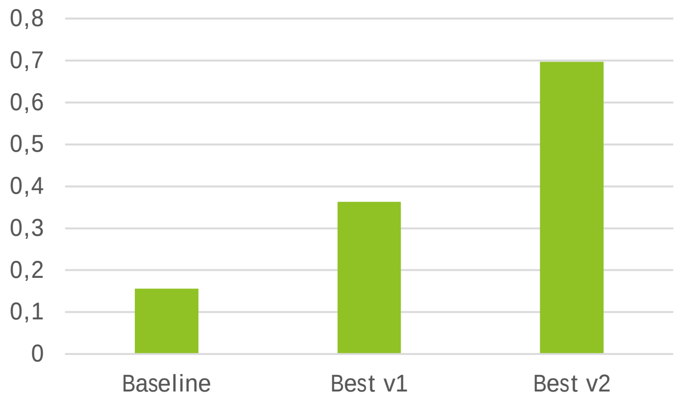


Can we do better?

- 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

- Baseline ranking method:
  - A session-based nearest neighbor technique
    - Configured to include reminders as well, more accurate than, e.g., BPR
- Results (hit rate, example, 2 evaluation variants)
    - v1 hides view event for target item, v2 reveals them



- Adaptive reminders better than simple reminders

Outline

- Why a common academic problem abstraction can be insufficient in practice
- Defining Sequence-Aware Recommender Systems
- An e-commerce case study:
  - On short-term intents, reminders, trends and discounts
- Summary and outlook

Recommendation success factors

- Our dataset includes additional information:
  - For each view event, the three recommendations displayed on the item detail page
  - Click events on the recommended items
  - (Purchase information)
- Our research goal:
  - Analyze which recommendations are successful, i.e., lead to a purchase event later on
  - Operationalize these insights in new recommendation algorithms

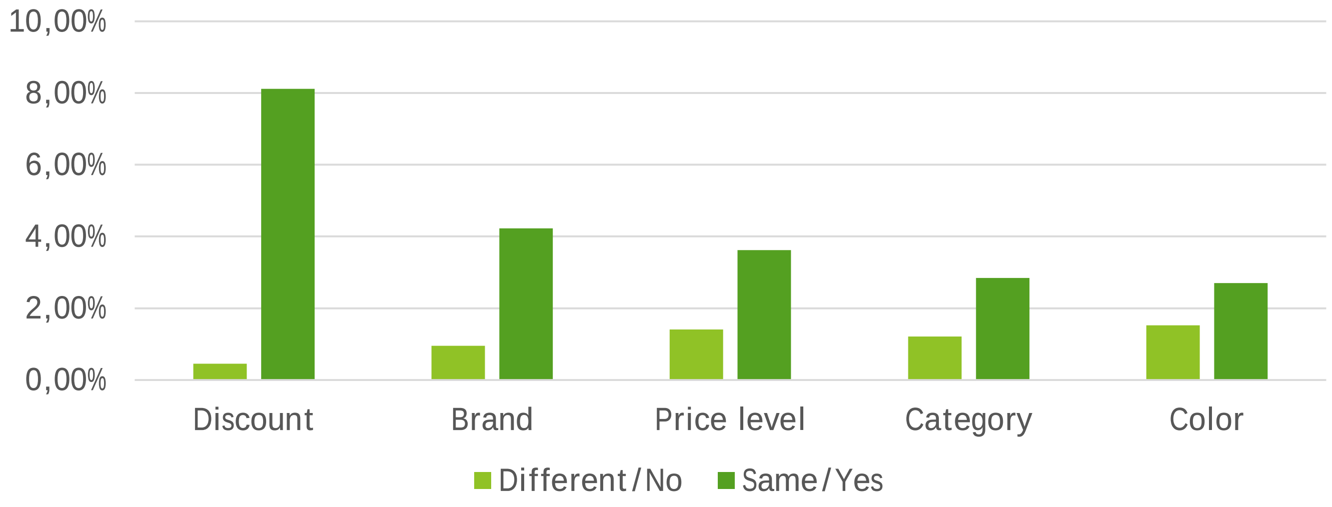
Jannach, D. and Ludewig, M.: *"Determining Characteristics of Successful Recommendations from Log Data – A Case Study"*. In: ACM Symposium on Applied Computing (ACM SAC 2017). Marrakesh, Morocco, 2017

Analysis results

- Reminders:
  - Only 10% of the recommendations seen before
  - But 44% of the successful ones were already known
- Short-term intents
  - Recommendations are more likely to be successful when from the same brand, category etc.
- Trends
  - Success rate of four times higher when the recommended item is trending on that day
- Discounts
  - Recommending on-sale items boosts the success rate

Analysis results

- Impact of item features on success of recommendations visualized
- About 1% success rate in general



Systematic feature analysis

- A more systematic approach
  - Engineered about 90 features to predict the success of a recommendation
  - Framed a classification problem
  - Determined the feature importance

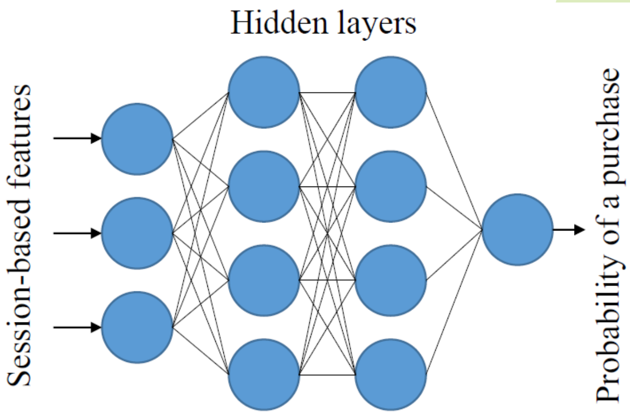
Feature	Gain Ratio	Chi Squared
Discount level	0.439	0.556
Current popularity (day)	0.371	1.000
Discount flag	0.325	0.556
Viewed before	0.286	0.435
Current popularity (week)	0.242	0.785
Distance to first item view (in days)	0.232	0.428
Distance to last item view (in days)	0.217	0.441
Distance to first item view (in sessions)	0.214	0.428
Distance to last item view (in sessions)	0.210	0.443
Current popularity (month)	0.201	0.563

Operationalization in algorithms

- Approach 1: A weighted two-phase approach
  - Create a candidate list of, e.g., 200 items using a baseline technique, in our case nearest-neighbors
  - Re-Rank items based on a weighted scoring scheme
  - Scoring functions
    - Content-wise similarity
    - Appearance in recent history (reminders)
    - Recent popularity on site
    - Level of discount
  - Importance weights have to be fine-tuned

Operationalization in algorithms

- Approach 2: A classification based approach
  - Framed scoring problem as a classification problem
  - Engineered 32 features in a similar way as was done for the feature importance analysis
    - Slightly different problem formulation
    - We are not interested analyzing success factors in general, but to predict the purchase probability for the given session
  - Used a deep neural-network approach for classification (H2O.ai library)
    - Outperformed Random Forests and manually tuned weights



Top features

Popularity	Normalized popularity of the item in the same day, week, or month
Viewed before	True, if the item was viewed before by the user
Views count	Number of previous views of the item by the user
Distance to first view	Distance to the first item view by the user in days or sessions
Distance to last view	Distance to the last view of the item in days or sessions
Brand ratio	Fraction of items of the same brand in the last 1, 2, and 3 sessions
Brand popularity	Overall popularity of the brand for the same day, week, or month
Color ratio	Fraction of items with the same color in the last 1, 2, and 3 sessions
Color popularity	Overall popularity of the color for the same day, week, or month
Category ratio	Fraction of items of the same category in the last 1, 2, and 3 sessions
Category popularity	Overall popularity of the category for the same day, week, or month
Price level ratio	Fraction of items of the same price level in the last 1, 2, and 3 sessions
Discount granted?	True, if the item is discounted
Discount level	Level of discount from 0 (no discount) to 3 (high discount)

Results

- Deep Learning based method led to best results
  - Independent of the chosen baseline ranking technique
- Random Forests were not better than the manually tuned weighted hybrid

Baseline	C-KNN		C-CoOcc		BPR	
Metric@10	HR	MRR	HR	MRR	HR	MRR
No post-processing	0.268	0.091	0.123	0.046	0.062	0.021
FM	0.281	0.093	0.145	0.052	0.119	0.046
IRec-FM	0.306	0.097	0.266	0.096	0.262	0.111
DR-FM	0.316	0.177	0.242	0.120	0.168	0.094
RPOP-FM	0.361	0.187	0.233	0.103	0.216	0.096
RFPredict	0.381	0.248	0.274	0.150	0.241	0.119
WR(RPOP,DR,0.5)-FM	0.382	0.220	0.262	0.121	0.225	0.100
DeepPredict	**0.405**	**0.284**	0.322	0.205	0.301	0.188

General insights

- First approach in academia to “reconstruct” success factors of recommendations from log data
- Could successfully operationalize the insights in a new prediction method
- Feature engineering is important
- Domain-dependent aspects should be considered
  - Reminding or not
  - Recommending discounted items or not
  - Recommending trending items or not

Outline

- Why a common academic problem abstraction can be insufficient in practice
- Defining Sequence-Aware Recommender Systems
- An e-commerce case study:
  - On short-term intents, reminders, trends and discounts
- Summary and outlook

Summary & Outlook

- Session-based recommendations relevant in many domains
- They require a different problem abstraction and evaluation methodology
- Recently, more awareness in the research community
  - e.g., using deep neural networks for next-item prediction
- Particularities in the e-commerce domain
  - E.g., considering discounts
- More general aspects (e.g., also for music recommendation)
  - E.g., considering trends, repeated recommendations

Outlook

- Perform additional experiments in different domains
  - Only partially validated so far
- Investigating the user perception of next-item recommendations
  - Recently made a study in the music domain
- Practical aspects
  - Assessing the business value of different algorithms
  - Assess the utility of complex models
    - See short paper in main program
- Academic aspects
  - Need for standardized problem abstraction and protocols for reproducibility

- Thank you for your attention

- Contact:
  - dietmar.jannach@tu-dortmund.de

References

- Jannach, D., Lerche, L. and Jugovac, M.: "Adaptation and Evaluation of Recommendations for Short-term Shopping Goals". In: RecSys 2015, pp. 211-218
- Jannach, D. and Ludewig, M.: *"Determining Characteristics of Successful Recommendations from Log Data – A Case Study"*. In: ACM Symposium on Applied Computing (ACM SAC 2017). Marrakesh, Morocco, 2017
- Lerche, L., Jannach, D. and Ludewig, M.: *"On the Value of Reminders within E-Commerce Recommendations"*. In: UMAP 2016, 2016.
- Kamehkhosh, I. and Jannach, D.: *"User Perception of Next-Track Music Recommendations"*. In Proceedings UMAP 2017, Bratislava, 2017.
- 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., 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. Springer, (forthcoming)