

Understanding Longitudinal Effects of Recommender Systems

Dietmar Jannach, University of Klagenfurt, Austria

dietmar.jannach@aau.at

RecSys Workshop at TU Vienna, 2024

Goals and Purposes of Recommender Systems

- Consumer perspective, e.g.,
 - Find content that satisfy information needs
 - Discover new content
 - Entertainment
- Provider perspective, e.g.,
 - More revenue, cross-selling, commissions
 - Customer engagement and retention
- Other stakeholders (item providers)
 - Exposure, sales

Long-term perspective: Goals and Purposes

- Implicitly, we want to design algorithms to *achieve long-term goals*, e.g. **sustained**
 - sales increases
 - user engagement
 - customer retention
 - diverse consumption patterns
 - discovery support
- Which is different from
 - persuading consumers to make individual ad-hoc purchases
 - nudging them to inspect unexpected out-of-profile items

Long-term perspective: RS4Good

- Particularly, we focus on Recommender Systems for Good, e.g., **sustained**
 - environmentally friendly decisions
 - energy saving
 - travel choices
 - healthier eating and behavior
 - Increased physical activities
 - healthier food choices
 - balanced information consumption and news diet
 - life-long learning and self-actualization

But how do we measure?

- Predominant in the academic context:
 - Offline evaluations (estimated 80-90% of published research)
 - User studies

But how do we measure?

- The common **offline evaluation approach**
 - Use random or time-based train-test split of historical data
 - Make recommendations for each (test) user **once**
 - Evaluate quality of recommendations once
 - Accuracy, diversity, novelty, serendipity, coverage, popularity ...
- Characteristics
 - One-time measurement
 - Impact on stakeholders not in the focus, instead
 - “post-diction” of known preferences
 - characteristics of recommendation

But how do we measure?

- The common **user study approach**
 - Create study environment
 - Involving control and treatment groups
 - Invite participants to interact with the environment(s) **once**
 - Measure and analyze
 - Observed behavior
 - Explicit statements by participants
- Characteristics
 - Assessment of user perceptions
 - May help to estimate effects on user behavior
 - One-time measurement only

Alternative Evaluation Approaches

- Reinforcement Learning Approaches
 - Evaluation usually involves a specific form of simulation
 - Including assumptions about user behavior
 - Goal often is to maximize the long-term reward
 - Adds some level of exploration behavior
 - “Offline reinforcement learning” not too different from supervised learning anymore, with similar one-time evaluation setups
 - Limited to one specific machine learning approach

Alternative Evaluation Approaches

- A/B tests
 - Testing different system versions in production
 - Seen as a gold standard of evaluation
 - May reveal unexpected longitudinal aspects, e.g., decreased sales diversity over time
 - Duration often limited in published papers
 - A few days to a few weeks (with no information about ramp-up times)
 - Potential trade-off of an organization's short-term and long-term goals
 - E.g., higher short-term CTR vs. long-term retention

Possible Directions

- Offline:
 - Simulation-based Approaches
 - Retrospective data-based studies
- Online:
 - Long-running A/B tests
 - Longitudinal user studies

Simulation-based Approaches



SimuRec

Workshop on Simulation Methods for Recommender Systems

[Home](#) [CFP](#) [Organizers](#) [Schedule](#)

Simulation for Recommender Systems

The Workshop on Synthetic Data and Simulation Methods for Recommender Systems

Research (SimuRec) brings together researchers and practitioners to understand the current state of the art on best practices for using simulation and data synthesis for research, development, and education in recommender systems.

This workshop is a great opportunity to help shape the use of simulations and synthetic data in the next several years of recommender systems research. We look forward to your contributions and the ensuing discussion.

Simulation-based Approaches

Understanding Longitudinal Dynamics of Recommender Systems with Agent-Based Modeling and Simulation*

Gediminas Adomavicius[†]
University of Minnesota
gedas@umn.edu

Stephan Leitner
University of Klagenfurt
stephan.leitner@aau.at

Dietmar Jannach
University of Klagenfurt
dietmar.jannach@aau.at

Jingjing Zhang
Indiana University
jjzhang@indiana.edu

Simulation: ABM

- Agent-based Modeling and Simulation (ABM)
 - Applied in various scientific fields, including Social Sciences, Economics, and Information Systems
 - Simulating the actions of heterogeneous and autonomous agents
 - With the goal of understanding the effects of these actions and the interactions on the entire system *in a longitudinal perspective*
 - A Generative Science approach:
 - Modeling process at the micro-level (agent behavior)
 - Goal is to obtain insights on the macro-level (e.g., emergent behavior)

Simulation: ABM Elements in RS

- Agents
 - Central: The **consumers** are modeled as agents
 - Receiving recommendations and reacting on them
 - Communicating with others
 - Other agents
 - Stakeholders such as recommendation providers or item suppliers
 - Sometimes, also the items that enter and leave the catalog
- Environment
 - A physical space, technical environment, social space
- Interactions
 - Among agents and environment

ABM Simulation Examples

- Insights from ABM-based simulations in the literature
 - The recommendation “paradox”
 - The value of recommendations diminishes over time the more users rely on the provided recommendations
 - Hybrid method helps to counteract the effect
 - Effects of preference biases
 - Preferences provided by consumers might be biased (polluted), e.g., by public average ratings for items or by predicted ratings
 - Such polluted ratings then propagate in the system, leading to more noise in the data

ABM Simulation Examples

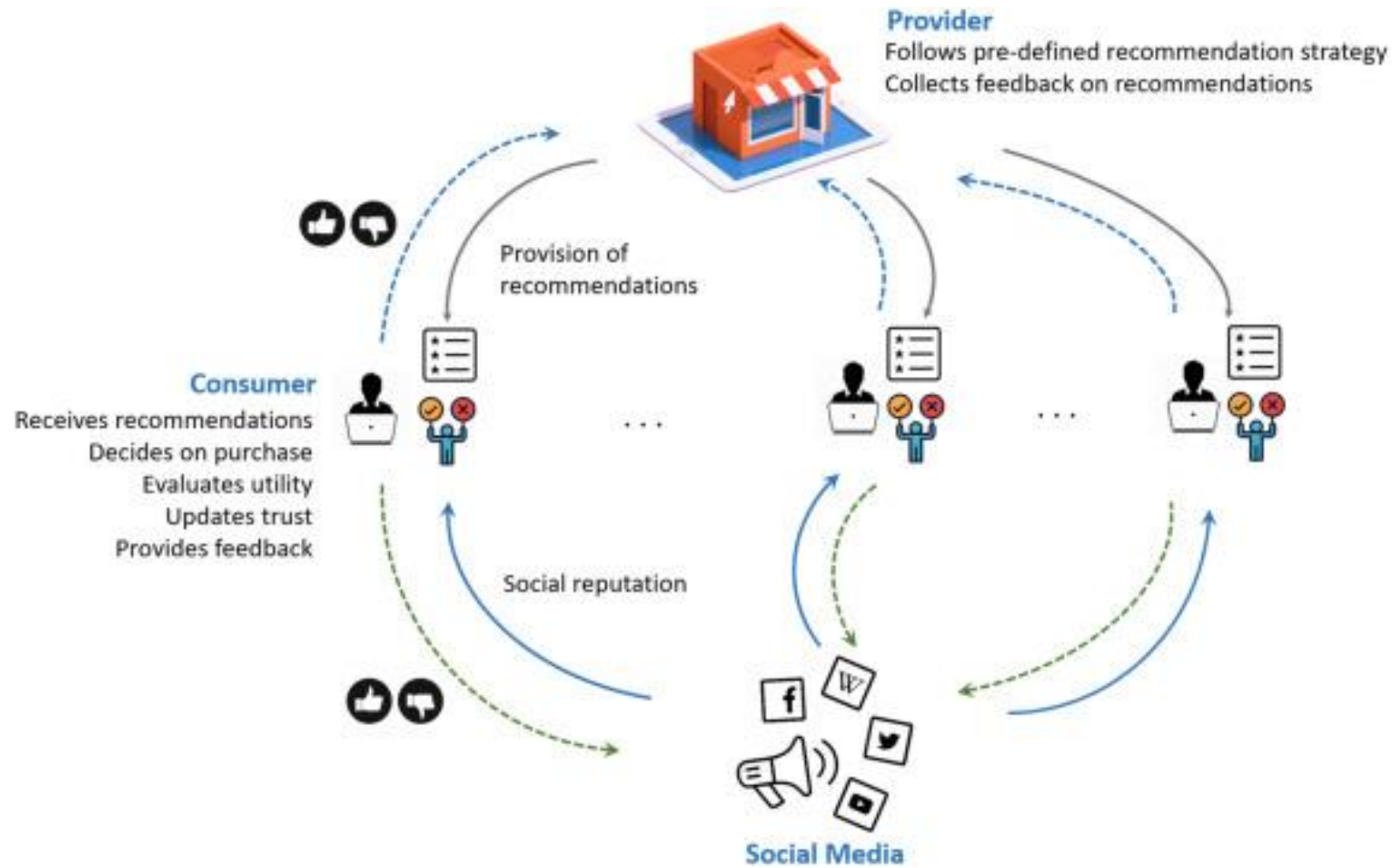
- Insights from ABM-based simulations in the literature
 - Addressing the profit-relevance trade-off



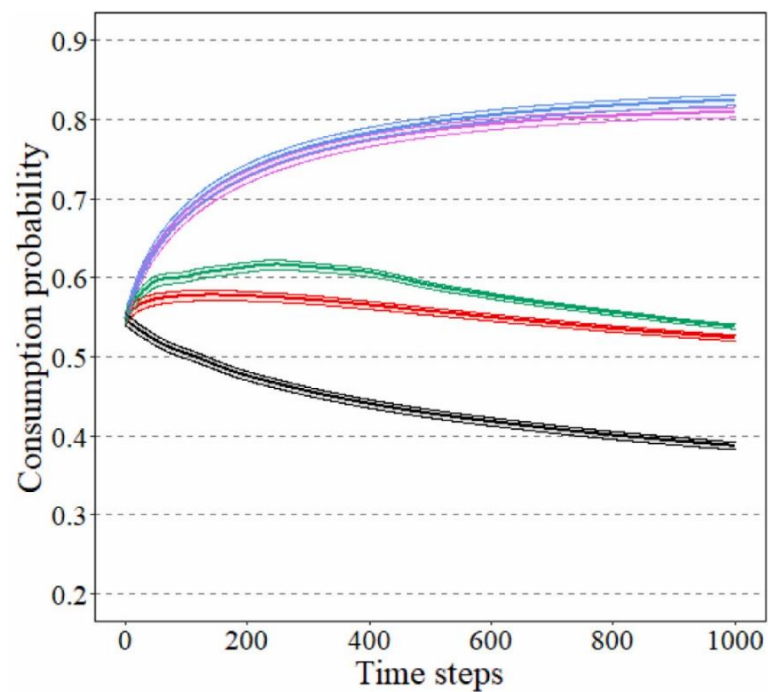
Profit-Relevance Trade-off

- Analyzing five provider strategies over time
 - Maximize **profit**, i.e., focus on items with highest profit margin
 - Maximize **consumer utility**, i.e., recommend the most relevant items for users
 - **Balanced**, considering both profit and relevance (2 settings)
 - **Popularity-based**, recommending the most popular items to everyone
- Modeling users as agents
 - Preference models
 - Item choice probabilities
 - Provider trust

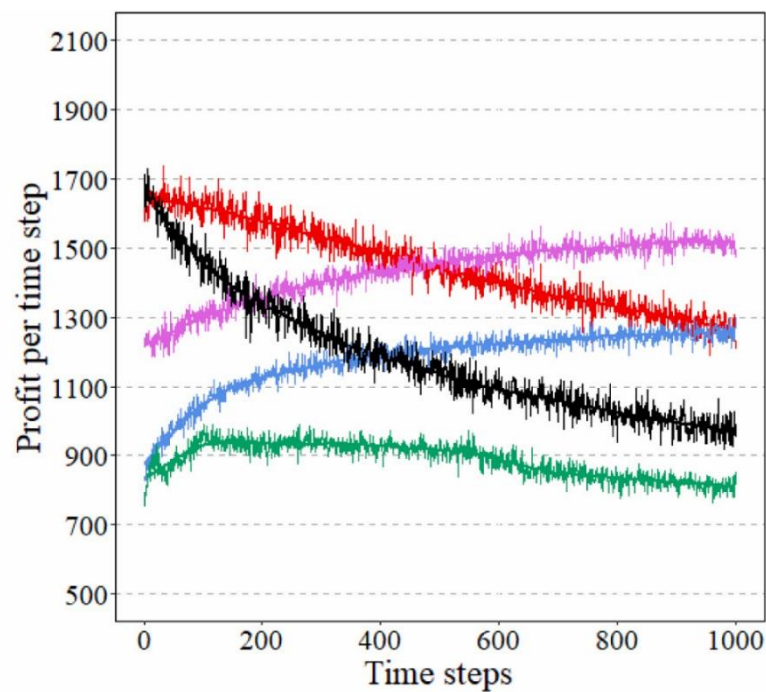
Profit-Relevance Trade-off: Environment



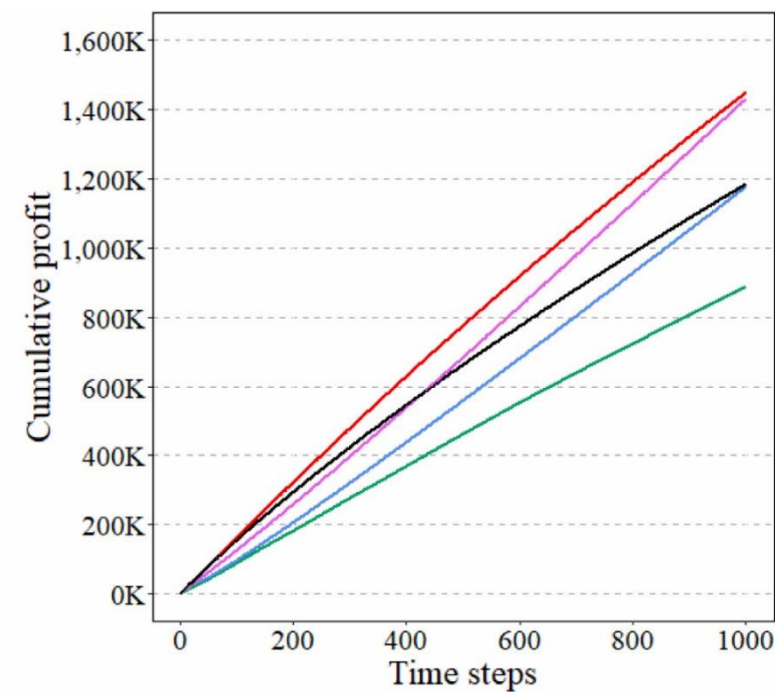
Outcomes



(a) Consumption probability



(b) Profit per time step



(c) Cumulative profit

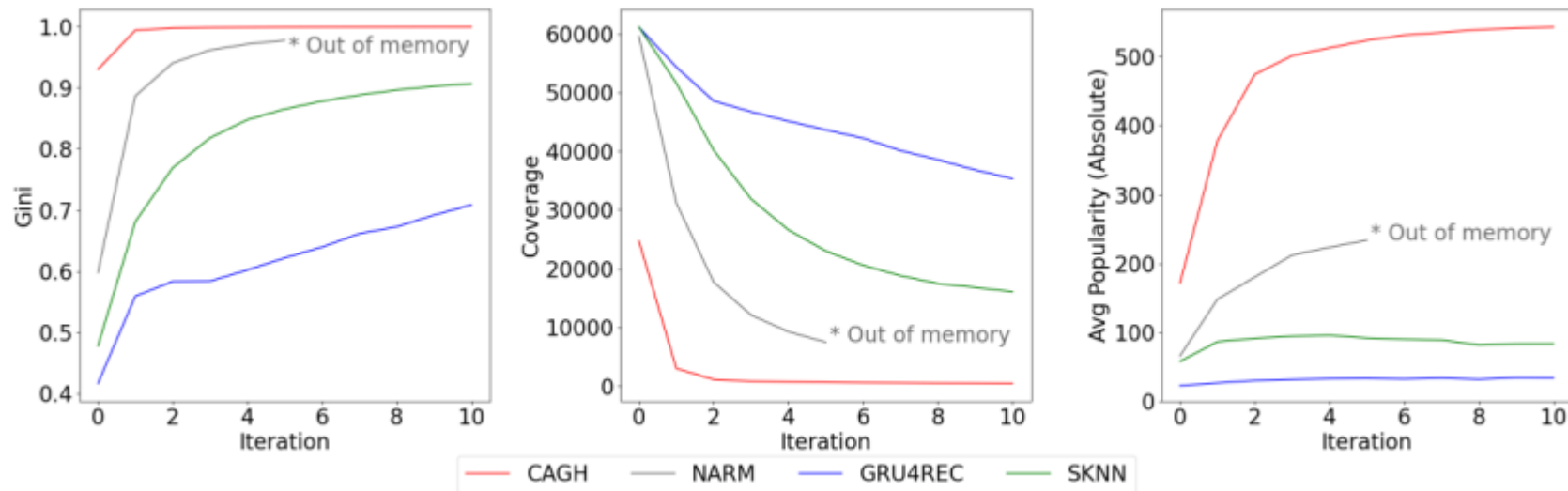
Strategy ■ Consumer-centric ■ Balanced ■ Consumer-biased ■ Profit-centric ■ Popularity-based

Other simulation-based research: Examples

- Goals of **non-ABM** simulation studies, effects of recommendations on
 - Sales diversity and concentration
 - Robustness against manipulations
 - Effects on news consumptions
 - Reinforcement of recommendation biases

Other simulation-based research: Examples

- Longitudinal effects of session-based recommendations



Simulation-based Models: Discussion

- Discussion
 - Simulation as a potential tool for analyzing complex system dynamics
 - Need to make (many) assumptions and simplifications
 - Often based on synthetic data
 - Potential gap between simulation and reality (Sim2Real)

Possible Directions

- Offline:

- Simulation-based Approaches



- Retrospective data-based studies

- Online:

- Long-running A/B tests

- Longitudinal user studies

Retrospective Data-based Studies

- Try to understand the effects of recommendations by analyzing collected logs
 - Requires access to production data
- Example:

Algorithmic Effects on the Diversity of Consumption on Spotify

Ashton Anderson
University of Toronto & Spotify
ashton@cs.toronto.edu

Lucas Maystre
Spotify
lucasm@spotify.com

Rishabh Mehrotra
Spotify
rishabhm@spotify.com

Ian Anderson
Spotify
iananderson@spotify.com

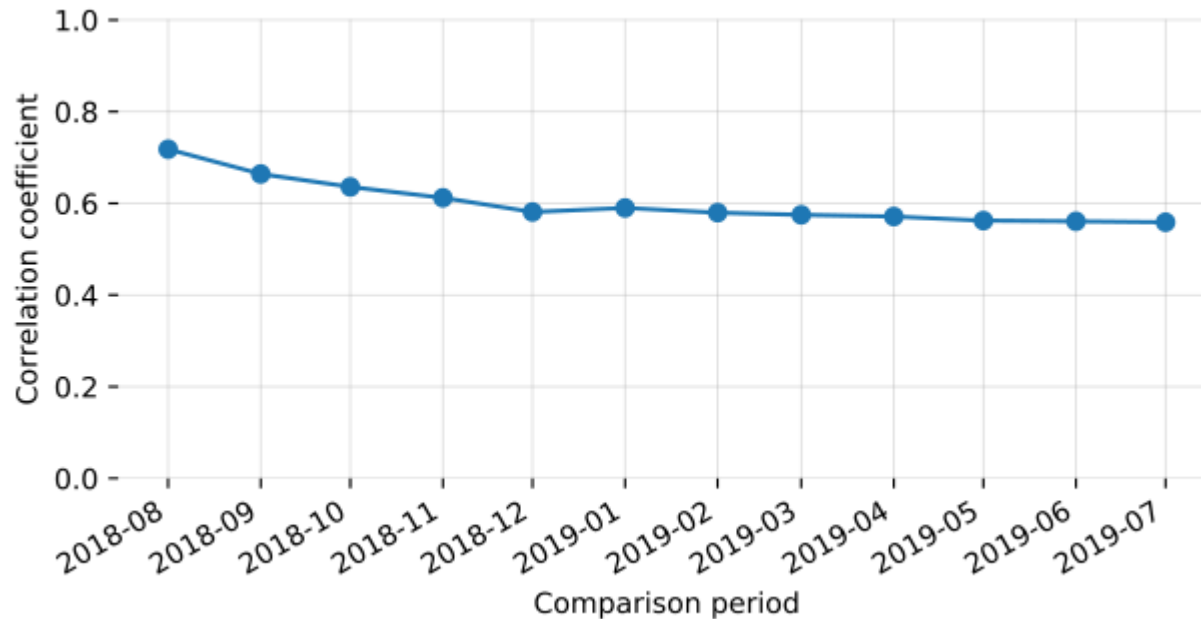
Mounia Lalmas
Spotify
mounia@acm.org

Retrospective Study by Spotify

- Different analyses performed
- Example: Correlate listening diversity with conversion and retention
 - **Insight**: High-consumption diversity is highly correlated with these business metrics
- Example: Effects of algorithm-driven listening
 - **Insight**: Leads to decreased consumption diversity
- Generally: Trade-off between short-term and long-term goals

Retrospective Study by Spotify

- Study covered millions of users
- Differences computed over a period of one year



Possible Directions

- Offline:
 - Simulation-based Approaches
 - Retrospective data-based studies
- • Online:
 - Long-running A/B tests
 - Longitudinal user studies

Longer-term A/B tests

- Very rare in the literature
- E.g., a study by Meta AI

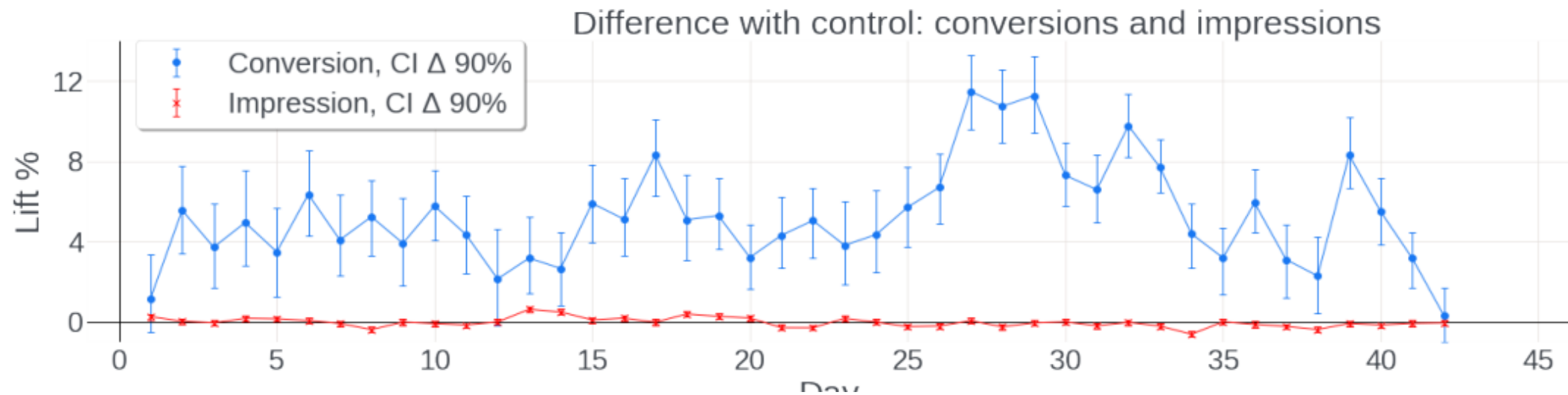
Optimizing Long-term Value for Auction-Based Recommender Systems via On-Policy Reinforcement Learning

Ruiyang Xu*, Jalaj Bhandari*, Dmytro Korenkevych, Fan Liu, Yuchen He, Alex Nikulkov,
Zheqing Zhu

Meta AI

Longer-Term A/B tests: Example by Meta AI

- A reinforcement learning approach for auction-based ad recommendation
 - Modified goal of policy: Target at long-term conversion
- A/B test outcome during six weeks
 - 4-10% improvement on conversion



Longer-term User Studies

- Also very rare, e.g., in the music domain by Porcaro et al.

Assessing the Impact of Music Recommendation Diversity on Listeners: A Longitudinal Study

LORENZO PORCARO, Music Technology Group, Universitat Pompeu Fabra, Spain and Joint Research Centre, European Commission, Italy

EMILIA GÓMEZ, Music Technology Group, Universitat Pompeu Fabra, Spain and Joint Research Centre, European Commission, Spain

CARLOS CASTILLO, Web Science and Social Computing Group, Universitat Pompeu Fabra, Spain and ICREA, Spain

Longer-term User Studies: Music Domain

- Porcaro et al.'s study:
 - Study lasted for 12 weeks
 - Involved 110 participants
 - Focus on effects of the recommendation of diverse and unfamiliar Electronic Music (EM), “Algorithmic Impact Assessment”
 - During the study ...
 - Participants did multiple listening sessions
 - Participants returned a EM feedback questionnaire multiple times
 - Multiple analyses over time
 - In particular the effects of diversification

Longer-term User Studies: Music Domain

- Porcaro et al.'s study, over 20

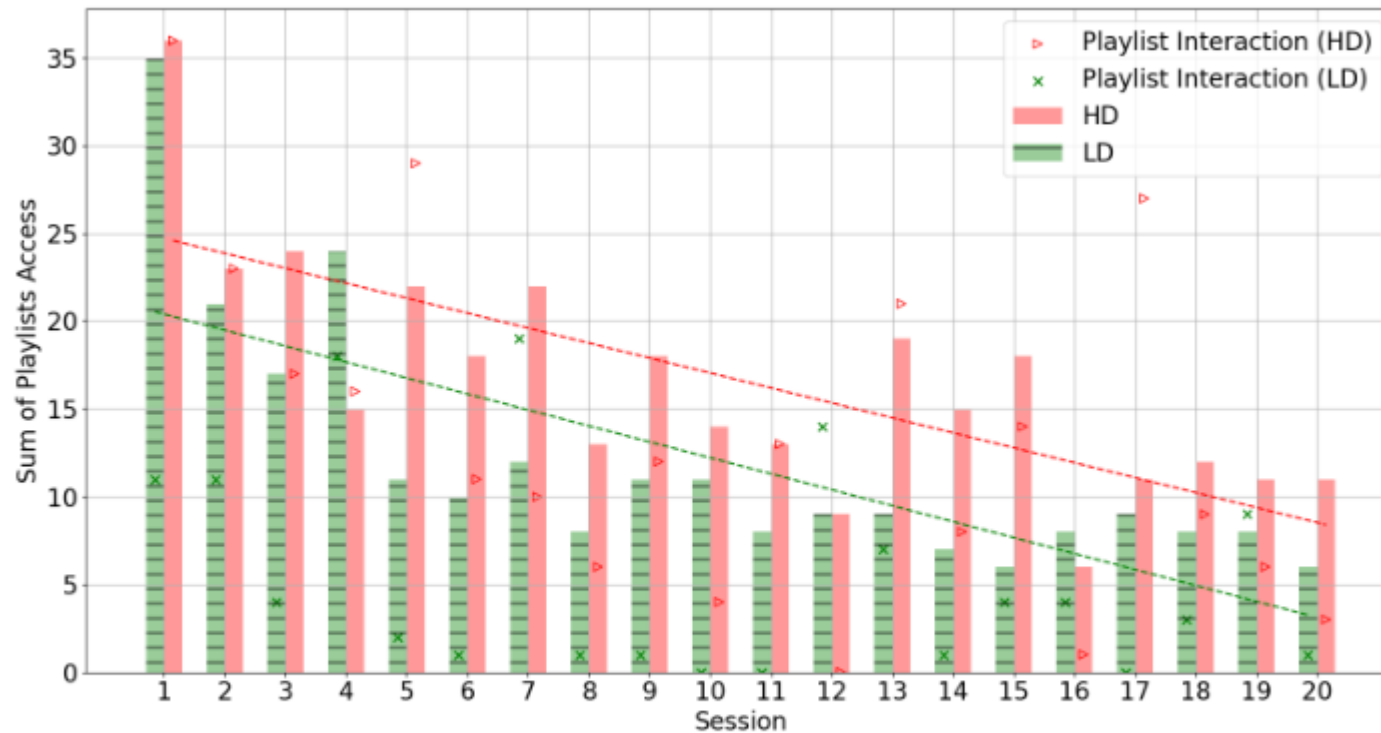


Fig. 4. Distribution of the playlists' accesses and interactions.

Summary / Discussion Points

- Understanding longitudinal aspects can be highly important in many application settings
- Only a limited number of works exist
- What are we not working on this?
 - It is more difficult to research and probably difficult to publish
 - Real-world data is often not available

Short-term and Long-term Call to Action

- What should we do in the long run?
 - Be courageous and attack difficult problems
 - Collaborate with industry and other disciplines
- Short-term activities
 - Raise awareness
 - Identify important application use cases
 - Organize workshops, e.g., a Dagstuhl seminar, propose journal special issues

-
- Thank you for your attention
 - Contact:
 - dietmar.jannach@aau.at

