UNIVERSITÄT KLAGENFURT

Multi-Objective Recommendation: Overview and Challenges

Dietmar Jannach, University of Klagenfurt, Austria

dietmar.jannach@aau.at

Keynote at the EvalRS Workshop at CIKM '22

Atlanta/Online, 2022

Materials

- Slides: https://tinyurl.com/evalrs2022-slides
- Survey/Summary: https://tinyurl.com/mors2022

Multi-Objective Recommender Systems: Survey and Challenges*

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

ABSTRACT

Recommender systems can be characterized as software solutions that provide users convenient access to relevant content. Traditionally, recommender systems research predominantly focuses on developing machine learning algorithms that aim to predict which content is relevant for individual users. In real-world applications, however, optimizing the accuracy of such relevance predictions as a single objective in many cases is not sufficient. Instead, multiple and often competing objectives have to be considered, leading to a need for more research in multi-objective recommender systems. We can differentiate between several types of such competing goals, including *(i)* competing recommendation quality objectives at the individual and aggregate level, *(ii)* competing objectives of different involved stakeholders, *(iii)* long-term vs. short-term objectives, *(iv)* objectives at the user interface level, and *(v)* system level objectives. In this paper we review these types of multi-objective recommendation settings and outline open challenges in this area.

considered a central task of a recommender and the corresponding objective was to minimize the mean absolute error (MAE), see [68] for work using MAE in 1996. Nowadays, item ranking is mostly considered to be more important than rating prediction, and a variety of corresponding ranking accuracy measures are used today.

While the metrics changed over time, the research community has been working on optimizing relevance predictions in increasingly sophisticated ways for almost 30 years now. The main objective of such research is to minimize the relevance prediction error or to maximize the accuracy of the recommendations. The underlying assumption of these research approaches is that better relevance predictions lead to systems that are more valuable for their users. This seems intuitive for many practical applications, because a better algorithm should surface more relevant items in the top-N lists shown to users.

Such an assumption might however not always be true, and it was pointed out many years ago that "being accurate is not enough" [53]. A recommender system might for example present

What are recommender systems good for?

- One possible generic characterization:

> *"Recommender systems are software solutions that provide users with **convenient access to relevant content.**"*

- The convenience of access may lie in that the system …
  - guides users to relevant items in situations of information overload,
  - helps them discover new things,
  - reminds them of what they were interested in earlier,
  - automatically takes action and plays suitable multimedia content
  - …

What’s predominant in the academic literature?

- The most common (implicit) objective
  - *“Guide users to relevant items in situations of information overload”*
  - Implemented through **personalized item ranking or filtering**

- The most common research operationalization
  - Recommendation as a supervised machine learning problem
  - Learn parameters of a prediction function from noisy data
  - Prediction is about the absolute or relative **relevance** of individual items for individual users

Prediction accuracy as a primary objective

- The operationalized objective
  - Create length-restricted (“top-n”) recommendation lists that have the most relevant items at the top

- How to quantify if the objective is achieved:
  - Through various accuracy measures in offline (data-based) experiments
  - Through monitoring user behavior or through questioning users in laboratory studies and field tests

Est. 1994 (and maybe even earlier)

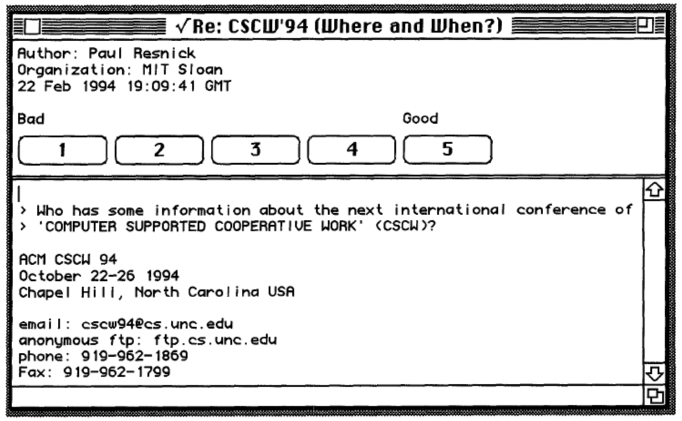
- GroupLens
  - Collaborative filtering based on a user item matrix of ratings
  - Predict empty places in the rating matrix
  - Based on nearest neighbors

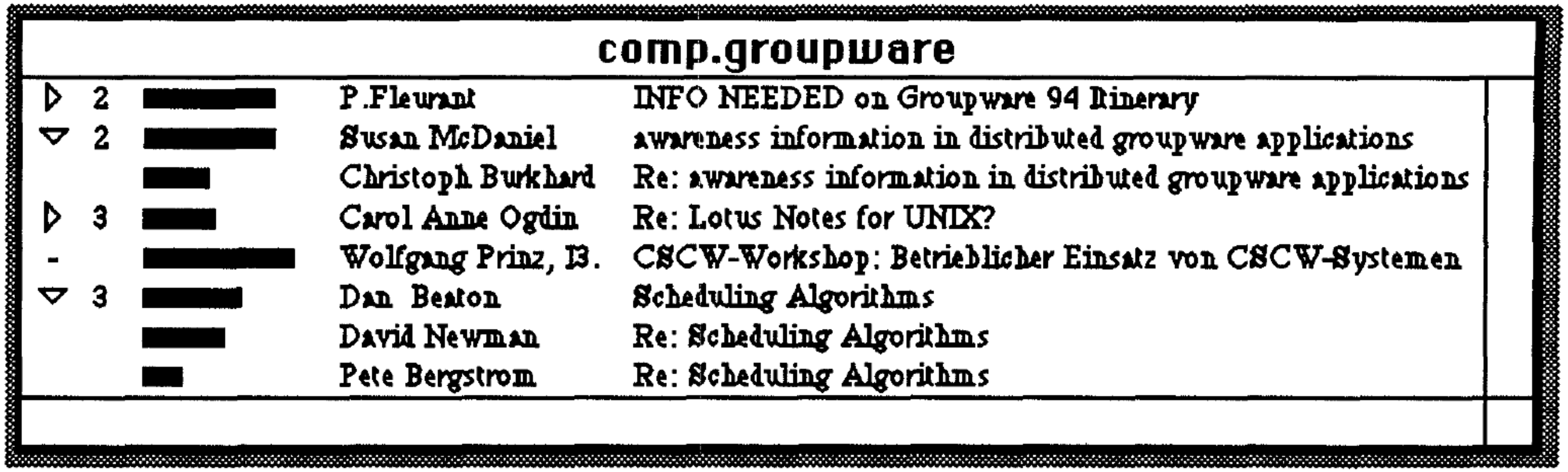
message #	Ken	Lee	Meg	Nan
1	1	4	2	2
2	5	2	4	4
3			3	
4	2	5		5
5	4	1		1
6	?	2	5	?

Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. 1994. *GroupLens: an open architecture for collaborative filtering of netnews.* CSCW '94. 175-186.

GroupLens

- Ratings and Recommendations





Measurements

- Rating predictions, from 1995
  - "The mean absolute [prediction] error must be minimized"
- Commonly used until today
  - RMSE, MRR, NDCG, Precision, Recall, F1, AUC, MAP …

7 : BOOM! One of my FAVORITE few! Can't live without it.
6 : Solid. They are up there.
5 : Good Stuff.
4 : Doesn't turn me on, doesn't bother me.
3 : Eh. Not really my thing.
2 : Barely tolerable.
1 : Pass the earplugs.

Figure 1: Ringo's scale for rating music.

Upendra Shardanand and Pattie Maes. 1995. Social information filtering: algorithms for automating "word of mouth". In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '95).

Optimizing …

- We work on prediction/ranking error minimization since 1994
  - With hundreds or even thousands of reported significant improvements over the state-of-the-art per year
  - Apparently, the problem is still not solved

- An important assumption:
  - We assume that improving on the algorithmic optimization objective (e.g., MRR, RMSE) leads to better systems

Our definition again, but what is relevance?

*"Recommender systems are software solutions that provide users with **convenient access to relevant content.**"*

- Let us assume this is what the user liked in the past

Our definition again, but what is relevance?

- And here's our recommendation

- In some interpretation relevant, and MAE might be even zero
- But no value for consumer and provider
  - **Consumer:** Boring, no discovery, knows items already
  - **Provider:** Could have used the space for promoting other content

Accuracy and beyond, since 2001 (at least)

Improving Recommendation Diversity

Keith Bradley [1] and Barry Smyth [1,2]

Being Accurate is Not Enough: How Accuracy Metrics have hurt Recommender Systems

Bradley K, Smyth B (2001) Improving recommendation diversity. In: Twelfth Irish Conference on Artificial Intelligence and Cognitive Science, pp 85–94

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).

Accuracy optimization in 2022

Neural net guesses memes

- Image prediction:
  people in conference room

- Confidence:
  100%

The Quest for the State-of-the-Art

- Evaluating on MovieLens-1M (2022)

Algorithm	Top@10					
	nDCG	**MAP**	**MRR**	**Pre**	**Rec**	**F1**
EASE^R	**0.336**	0.335	**0.583**	0.274	**0.194**	0.190
SLIM	0.335	**0.337**	0.580	**0.275**	0.189	0.188
MF2020	0.329	0.327	0.563	0.272	0.190	**0.192**
UserKNN	0.315	0.314	0.554	0.256	0.183	0.179
RP$^3\beta$	0.315	0.313	0.556	0.256	0.184	0.179
iALS	0.306	0.304	0.542	0.252	0.179	0.176
MultiVAE	0.294	0.284	0.514	0.243	0.183	0.175
ItemKNN	0.292	0.293	0.518	0.242	0.163	0.163
NeuMF	0.277	0.275	0.494	0.232	0.157	0.158
BPRMF	0.275	0.271	0.502	0.226	0.166	0.161
MostPop	0.159	0.159	0.317	0.137	0.084	0.086
Random	0.008	0.007	0.020	0.007	0.004	0.004

Anelli, V. W., Bellogin, A., Noia, T. D., Jannach, D. and Pomo, C.: *"Top-N Recommendation Algorithms: A Quest for the State-of-the-Art"*. In: 30th ACM Conference on User Modeling, Adaptation and Personalization (UMAP 2022). 2022

From a single objective to multiple objectives

- Being able to predict the relevance of individual items remains to be important
- However, optimizing for one type of objectives (and metrics) may be too simplistic
  - Recommendations can serve a variety of purposes
    - including, e.g., information filtering, discovery, or entertainment
  - And they can create value both for consumers and providers
    - e.g., in terms of revenue and customer engagement

Jannach, D. and Adomavicius, G.: "*Recommendations with a Purpose*". In: Proceedings of the 10th ACM Conference on Recommender Systems (RecSys 2016). Boston, Massachusetts, USA, 2016, pp. 7-10

From a single objective to multiple objectives

- For more impactful research, we must improve the way we design and evaluate recommender systems
  - And escape the "McNamara Fallacy", where we rely too much on a single set of measures to assess progress

- Considering multiple objectives is one key step in this direction

Jannach, D. and Bauer, C.: "Escaping the McNamara Fallacy: Towards more Impactful Recommender Systems Research". AI Magazine, Vol. 41(4). 2020, pp. 79-95

Competing objectives

- Sometimes objectives are aligned / synergistic
  - e.g., when higher diversity leads to more engagement
- Sometimes objectives are competing
  - e.g., when higher diversity leads to lower accuracy in top-n lists
- Algorithmically balancing the objectives often is the central problem in the literature
  - What the right balance should be is often assumed to be given
  - A balanced “macro”-score in the EvalRS@CIKM ‘22 Challenge

J. Tagliabue, F. Bianchi, T. Schnabel, G. Attanasio, C. Greco, G. de Souza P. Moreira, P. John Chia: “EvalRS: a Rounded Evaluation of Recommender Systems”, https://arxiv.org/abs/2207.05772

Types of multi-objective problems

- Competing quality objectives
  - e.g., increase diversity without sacrificing accuracy
- Competing stakeholder objectives
  - e.g., increase profitability while maintaining consumer satisfaction
- Competing temporal objectvies (short-term vs long-term)
  - e.g., high short-term click-through-rate (CTR) vs long-term consumer retention

Types of multi-objective problems

- Competing quality objectives
  - e.g., increase diversity without sacrificing accuracy
- Competing stakeholder objectives
  - e.g., increase profitability while maintaining consumer satisfaction
- Competing temporal objectvies (short-term vs long-term)
  - e.g., high short-term click-through-rate (CTR) vs long-term consumer retention

Competing quality objectives – Individual level

- Consumers may have multiple decision criteria or needs
  - e.g., distance to a certain restaurant, types of dishes offered, price category
- The recommender system must take these into account and balance them
  - Support a multi-criteria decision problem, balance consumer objectives/needs which might be competing
  - Might be helpful to make recommendation reasoning explicit, using explanations
  - Maybe help users find solutions when no "optimal" solution exists

Competing quality objectives – Individual level

- Calibration – Idea:
  - Match recommendation list with individual past user tendencies
  - No global target, e.g., for diversity, needed or desirable
- Technical approach, e.g.,
  - Minimize difference between distributions
    - Past profile and recommendations
    - Related to next-track music recommendation problem
  - Often based on re-ranking

Oh, J. , Park, S. , Yu, H. , Song, M. , & Park, S. (**2011**). Novel recommendation based on personal popularity tendency. In Proceedings ICDM ’11

Jugovac, M., Jannach, D. and Lerche, L. (**2017**) "Efficient Optimization of Multiple Recommendation Quality Factors According to Individual User Tendencies". Expert Systems With Applications, Vol. 81, pp. 321-33

Harald Steck. (**2018**). **Calibrated recommendations.** In Proceedings of the 12th ACM Conference on Recommender Systems (RecSys '18)

H. Abdollahpouri, M. Mansoury, R. Burke, B. Mobasher, and E. Malthouse. (**2021**). User-centered Evaluation of Popularity Bias in Recommender Systems. In Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '21).

Competing quality objectives – System level

- Often: “Accuracy vs the rest”, e.g.,
  - Diversity
  - Novelty
  - Serendipity
  - Popularity bias
  - Catalog coverage
  - Fairness

- Objective
  - “Increase other objective(s) while retaining accuracy”

Accuracy vs the rest: Not always a contrast

- Sometimes, objectives are (almost) synergistic

- Next-track music recommendation:
  - Objective is not only to find the "right tracks" (accuracy), but also a smooth continuation, e.g., in terms of the tempo
  - Study results indicate that smoothness can be improved while accuracy is even slightly increased

Dietmar Jannach, Lukas Lerche, and Iman Kamehkhosh. 2015. Beyond "Hitting the Hits": Generating Coherent Music Playlist Continuations with the Right Tracks. In Proceedings of the 9th ACM Conference on Recommender Systems (RecSys '15).

Accuracy vs the rest: Variations

- Number of objectives
  - Accuracy vs one other quality dimension
  - Multiple optimization goals (including accuracy)

- Nature of the optimization target, e.g.,
  - Reach given global threshold or target distribution
  - Match previous individual user tendencies (calibration)
  - Be better than other algorithm with similar accuracy
  - Limit the loss in accuracy

Jambor, T. , & Wang, J. (2010). Optimizing multiple objectives in collaborative filter- ing. In Proceedings of the 4th ACM Conference on Recommender Systems (RecSys '10)

Zhang, Y. C. , Séaghdha, D. O. , Quercia, D. , & Jambor, T. (2012). Auralist: Introducing serendipity into music recommendation. In Proceedings of the 5th International Conference on Web Search and Web Data Mining (WSDM '12) (pp. 13–22).

Accuracy vs the rest: Explore-Exploit

- A common dilemma: explore vs exploit
  - Continue to make more “safe” recommendations, or help users discover from time to time with more risky ones?
  - Risky or uncertain recommendations may lead to increased engagement, e.g., in the music domain

- Technically, often addressed with Contextual Bandit/ Reinforcement Learning approaches

J. McInerney, B. Lacker, S. Hansen, K. Higley, H. Bouchard, A. Gruson, and R. Mehrotra. 2018. Explore, exploit, and explain: personalizing explainable recommendations with bandits. In Proceedings ACM RecSys '18

Accuracy vs the rest: Main methods

- Re-ranking an accuracy-optimized list
  - Can be used with any baseline ranker
  - Simple, fast, and effective
  - Greedy methods may not find the optimum
- Modeling (e.g., a multi-element loss function)
  - Joint consideration of multiple quality factors
  - May be tied to specific algorithm, no global guarantees
- Mathematical multi-objective optimization
  - Pareto-optimal solutions possible
  - Scalability issues

Accuracy vs the rest: Measuring offline

- Offline evaluation
  - Based on computational proxies for diversity etc., e.g., Intra-List-Diversity
  - Mostly focus on not compromising accuracy too much while increasing other factors
- Common challenges
  - How do we know that metrics are representative of user perception?
    - For Intra-List-Diversity: Would users notice? Would the position of items matter?
  - What is the right level of, e.g., diversity?
    - In case of calibration, how do we know if users will prefer the calibrated recommendations?

M Ge, F Gedikli, D Jannach, Placing high-diversity items in top-n recommendation lists, Proceedings ITWP Workshop 2011

Accuracy vs the rest: Measuring offline

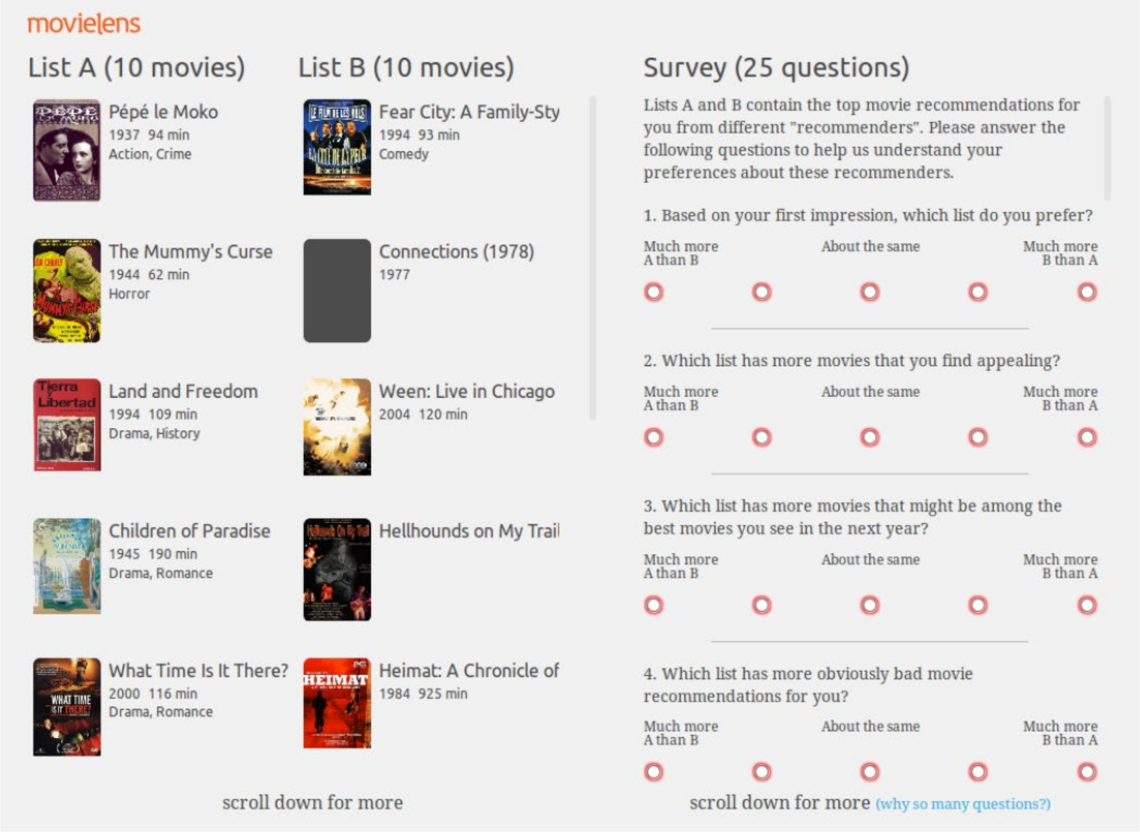
- Fairness and biases as recent topics
  - Fairness often measured in terms of popularity distributions
- Substantial limitations of offline evaluations
  - No validation for underlying assumptions and metrics
    - Less popular items might just be of poor quality
    - In the case of Netflix, even a "healthy dose" of popularity is added
  - Absolute "target" fairness is assumed to be given
  - Achieving fairness algorithmically is actually mostly trivial
- Severe danger of an "abstraction trap", as in XAI in general

Carlos A. Gomez-Uribe and Neil Hunt. 2016. The Netflix Recommender System: Algorithms, Business Value, and Innovation. ACM Trans. Manage. Inf. Syst. 6, 4

Y. Deldjoo, D. Jannach, A. Bellogin, A. Difonzo, D. Zanzonelli: A Survey of Research on Fair Recommender Systems, https://arxiv.org/abs/2205.11127

Accuracy vs the rest: Measuring with users

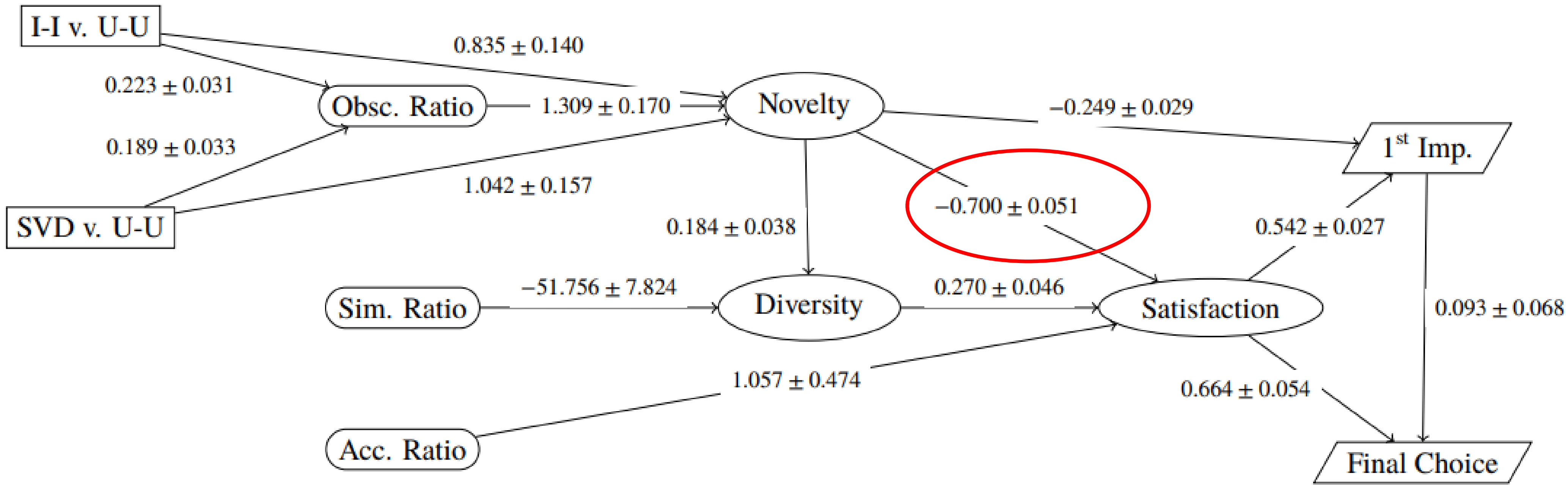
- Evaluation with users
  - Assessing subjective quality perceptions
  - E.g., by comparing the output of different algorithms



Michael D. Ekstrand, F. Maxwell Harper, Martijn C. Willemsen, and Joseph A. Konstan. 2014. User perception of differences in recommender algorithms. In Proceedings RecSys '14.

Accuracy vs the rest: Measuring with users

- Assessment based on established evaluation frameworks or specific models
  - One study: too much of one thing (e.g., novelty) may be detrimental to overall satisfaction



Accuracy vs the rest: System-level aspects

- A general question:
  - Will slightly more accurate models translate into better systems in practice, e.g., in terms of consumer or provider value?
- Even if we expect this, what is the price for the better models?
  - The most accurate models might be computationally complex and may not scale well
  - The most accurate models might be difficult to implement and maintain
  - The most accurate models might be difficult to scrutinize
- Thus, trade-offs may occur also at this level

Types of multi-objective problems

- Competing quality objectives
  - e.g., increase diversity without sacrificing accuracy
- Competing stakeholder objectives
  - e.g., increase profitability while maintaining consumer satisfaction
- Competing temporal objectvies (short-term vs long-term)
  - e.g., high short-term click-through-rate (CTR) vs long-term consumer retention

Types of multi-objective problems

- Competing quality objectives
  - e.g., increase diversity without sacrificing accuracy

- Competing stakeholder objectives
  - e.g., increase profitability while maintaining consumer satisfaction

- Competing temporal objectvies (short-term vs long-term)
  - e.g., high short-term click-through-rate (CTR) vs long-term consumer retention

Mulitstakeholder Recommendation

- Historically, focus on consumer value only
  - Assuming better recommendations (and consumer satisfaction) would indirectly be beneficial for the provider
- In practice, however
  - Providers use recommender systems for a certain purpose and/or value expectation
    - e.g., to increase sales, customer engagement, change consumer behavior
- May lead to multiple (competing) objectives

Abdollahpouri, H., Adomavicius, G., Burke, R., Guy, I., Jannach, D., Kamishima, T., Krasnodebski, J. and Pizzato, L.: "Multistakeholder Recommendation: Survey and Research Directions". User Modeling and User-Adapted Interaction, Vol. 30. 2020, pp. 127–158

Jannach, D. and Zanker, M.: "Impact and Value of Recommender Systems". In: Recommender Systems Handbook. Ricci, F., Shapira, B. and Rokach, L. (Eds.), Springer US, 2021

Mulitstakeholder Recommendation: Examples

- Two stakeholders: consumer and provider
  - Some research exists on “price and profit aware” recommender systems
    - E.g., when two items have the same assumed utility for a consumer, it might be tempting to recommend the one with higher profit
    - E.g., an existing discount may drastically increase the chances a recommendation is accepted by a consumer
- Two stakeholders: special cases
  - Social recommendation, job recommendation, dating recommendations
  - E.g., “reciprocal” aspects, mutual compatibility required

Jannach, D. and Adomavicius, G.: "Price and Profit Awareness in Recommender Systems". In: Proceedings of the ACM RecSys 2017 Workshop on Value-Aware and Multi-Stakeholder Recommendation. Como, Italy, 2017

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).

Mulitstakeholder Recommendation: Examples

- Group recommendation, a long-studied problem
  - Goal is to make recommendations that satisfy a group as a whole, considering the preferences of all group members (stakeholders)
  - Technical solutions often based on studying various preference aggregation strategies
- Unique aspect:
  - All group members receive the same recommendations at the end
- Research directions
  - Studying group decision making processes using observational and simulation studies

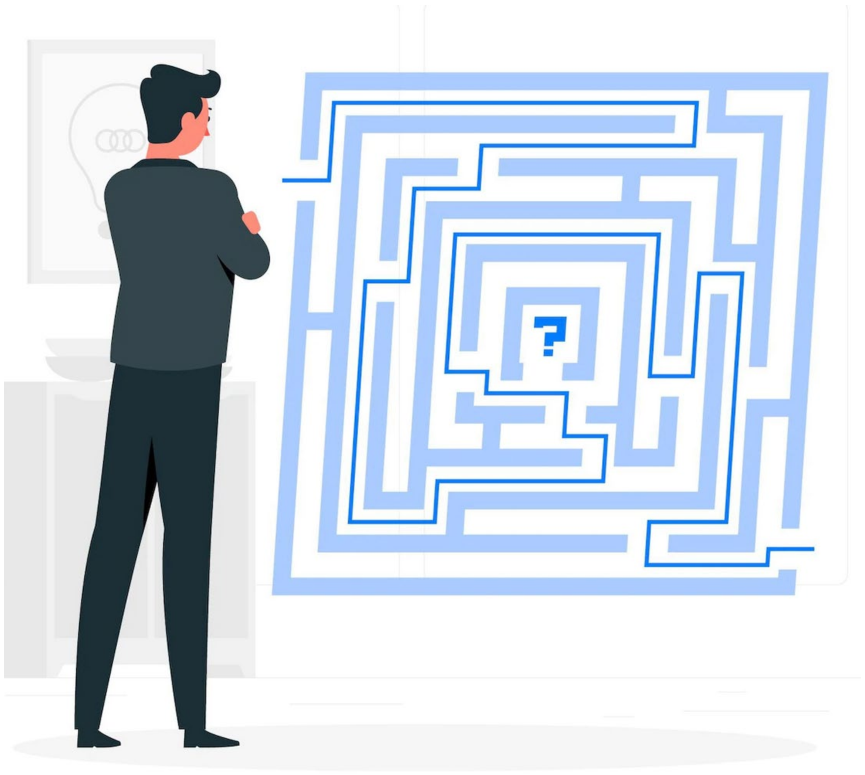
Masthoff, J., Delić, A. (2022). Group Recommender Systems: Beyond Preference Aggregation. In: Recommender Systems Handbook. Springer

Nguyen, T.N., Ricci, F., Delic, A. et al. Conflict resolution in group decision making: insights from a simulation study. User Model User-Adap Inter 29, 895–941 (2019)

Delic, A., Neidhardt, J., Nguyen, T.N. et al. An observational user study for group recommender systems in the tourism domain. Inf Technol Tourism 19, 87–116 (2018)

Mulitstakeholder Recommendation: Challenges

- A challenging research environment
  - Lack of real-world data
    - Profitability data is business critical and usually confidential
  - Assumptions must be made regarding the “optimal” balance for a given application setting
  - Longitudinal aspects may be important
    - An unbalanced approach may produce an effect only in the long run



Types of multi-objective problems

- Competing quality objectives
  - e.g., increase diversity without sacrificing accuracy
- Competing stakeholder objectives
  - e.g., increase profitability while maintaining consumer satisfaction
- Competing temporal objectvies (short-term vs long-term)
  - e.g., high short-term click-through-rate (CTR) vs long-term consumer retention

Types of multi-objective problems

- Competing quality objectives
  - e.g., increase diversity without sacrificing accuracy

- Competing stakeholder objectives
  - e.g., increase profitability while maintaining consumer satisfaction

- Competing temporal objectvies (short-term vs long-term)
  - e.g., high short-term click-through-rate (CTR) vs long-term consumer retention

Short-term vs Long-Term Optimization

- In some domains, it is easy to increase certain Key Performance Indicators, e.g.,
  - **Hotels:** Recommend relatively cheap hotels to maximize revenue
  - **News:** Recommend recently trending items and latest celebrity gossip to increase click rates; use clickbait headlines
  - **Music:** Recommend Ed Sheeran to everyone, to be on the safe side
- In the long run however,
  - **Hotels:** More important goal may be long-term profitability
  - **News:** Readers may be disappointed by article behind clickbait headline
  - **Music:** Subscribers may lose interest when there’s nothing new

Longitudinal effects of recommender systems

- Very limited research so far in academic environment
  - Reports from industry on long-term effects missing as well
- Some uptake of the topic in recent years
  - Mostly based on various forms of **simulation** (reinforcement learning based, agent-based)
  - Often considering also multistakeholder settings
  - In individual cases, considerations of the **customer lifetime value** are taken into account

D.-R. Liu and Y.-Y. Shih. Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences. J Systems and Software, 2005.

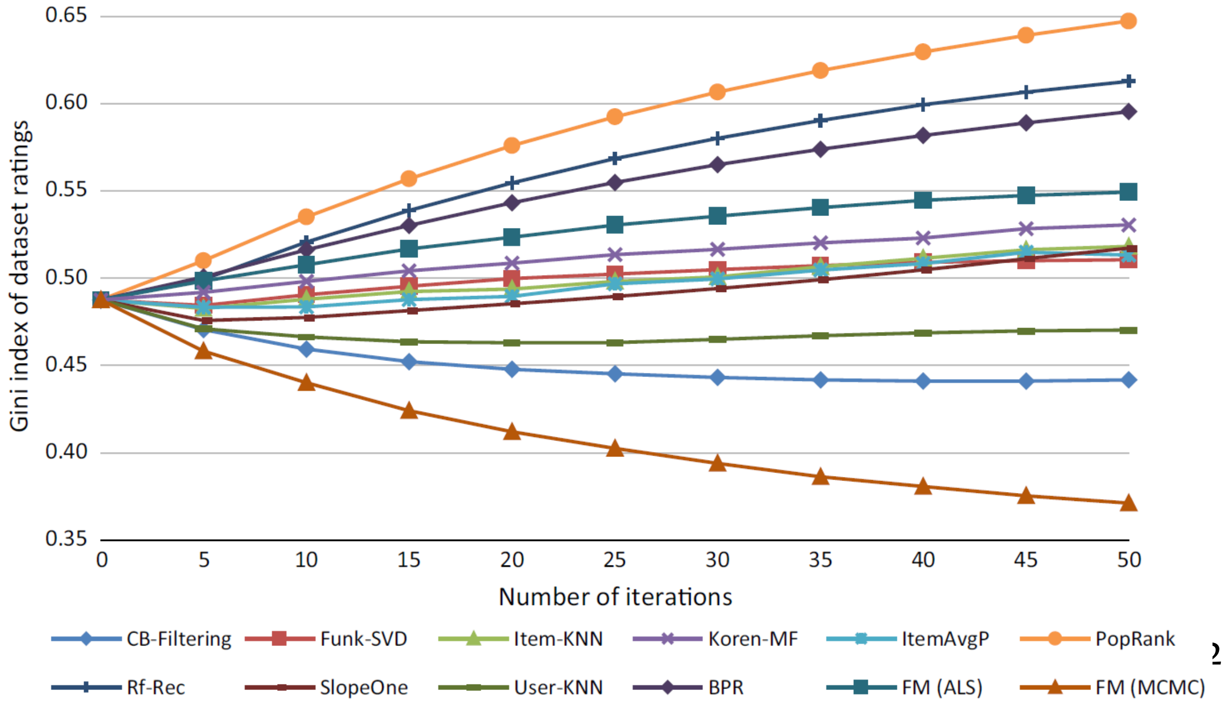
D. Fleder and K. Hosanagar. 2009. Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity. Mngmt Sci 2009

Krasnodebski and John Dines. 2016. Considering Supplier Relations and Monetization in Designing Recommendation Systems. RecSys '16.

U. Panniello, S. Hill, M. Gorgoglione, The impact of profit incentives on the relevance of online recommendations, Electron. Commer. Rec. Appl. 20(2016) 87–104

Studying longitudinal effects: Examples

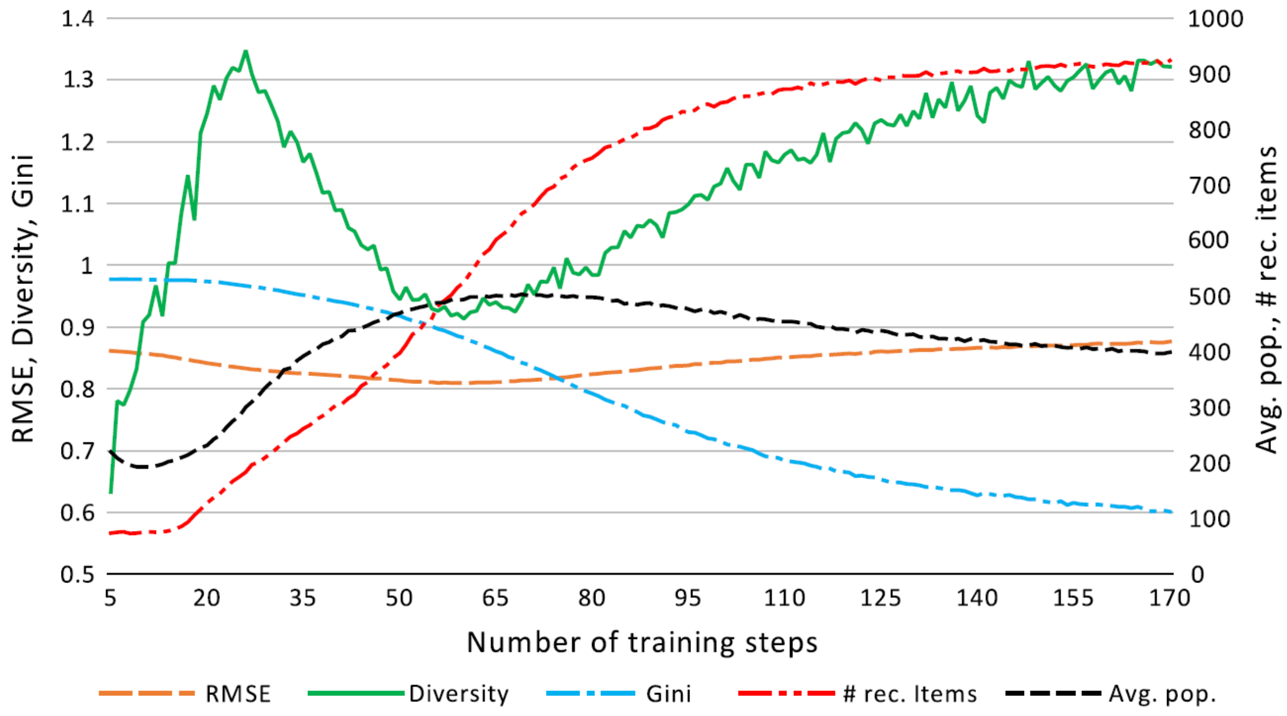
- Studying reinforcement of concentration effects
  - Some algorithms tend to focus on a small set of items
  - Assuming that these few items will be often accepted and then rated by users, the concentration may be intensified
  - Study shows that algorithms may differ significantly in terms of this effect
  - Effects may even depend on specifics of algorithm configuration



Jannach, D., Lerche, L., Kamehkhosh, I. and Jugovac, M.: "What recommenders recommend: an analysis of recommendation biases and possible countermeasures". User Modeling and User-Adapted Interaction, Vol. 25(5). Springer Nature, 2015, pp. 427-491

Studying longitudinal effects: Examples

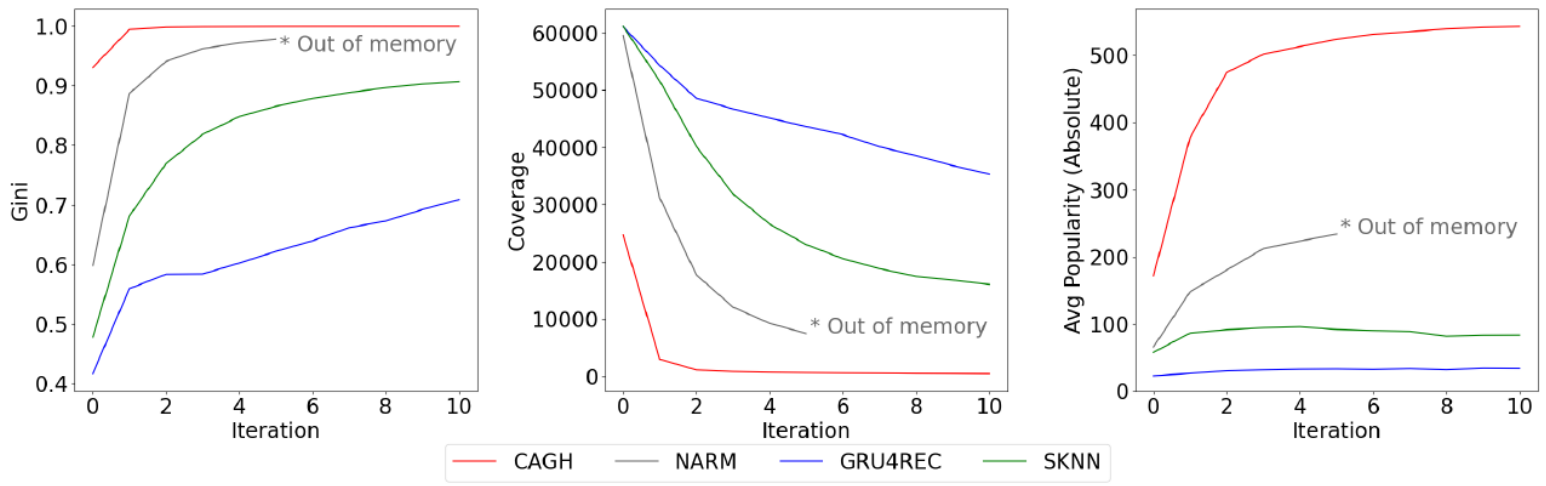
- Studying effects of hyper-parameter settings (statically)
  - RMSE values do not vary too much
  - Diversity, Gini, etc., may in contrast can be very different, depending on the training steps



Jannach, D., Lerche, L., Kamehkhosh, I. and Jugovac, M.: "What recommenders recommend: an analysis of recommendation biases and possible countermeasures". User Modeling and User-Adapted Interaction, Vol. 25(5). Springer Nature, 2015, pp. 427-491

Studying longitudinal effects: Examples

- Studying concentration, coverage, and popularity effects in session-based recommendation



- Significant differences between algorithms
  - Two neural methods show quite opposite effects

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

Studying longitudinal effects: Examples

- Studying recommendation performance in the long run
- Study reveals a "performance paradox"
  - "[…] users' reliance on the system's recommendations to make item choices generally tends to make the recommender system less useful in the long run"
  - E.g., if users rely too much on the recommendations, their average consumption diversity will decrease over time
- Study uses an Agent-based Modeling (ABM) approach

Zhang, J., Adomavicius, G., Gupta, A., Ketter, W., Consumption and Performance: Understanding Longitudinal Dynamics of Recommender Systems via an Agent-Based Simulation Framework, Inf. Syst. Res. 31(1): 76-101 (2020)

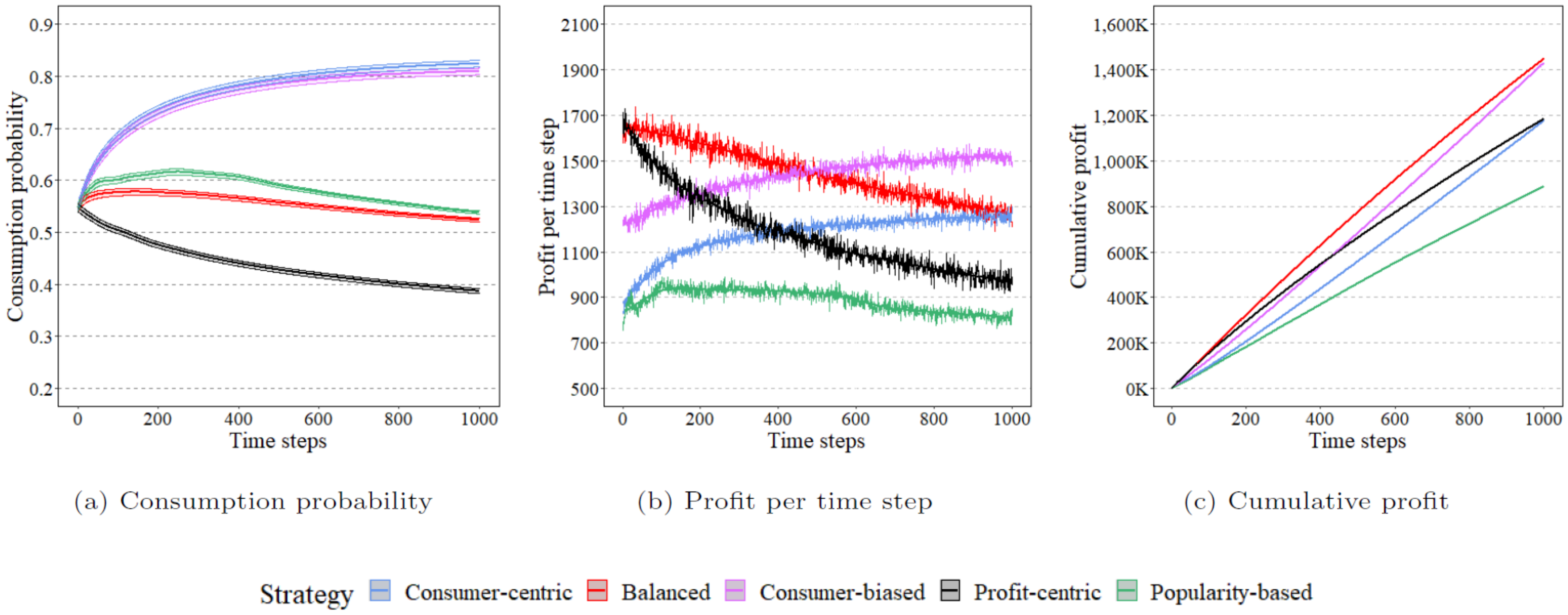
Studying longitudinal effects: Examples

- ABM approach to study consumer and provider value over time
- Model assumptions:
  – Each recommendable item has a profit margin
  – Provider can focus either more on profit or more on consumer relevance
    - Different strategies implemented
  – Agents receive recommendations and make choices based on time-varying consumption probability
    - Influenced by own past experiences and social reputation of the provider
  – Agents may post their experiences on social media

N. Ghanem, S. Leitner, D. Jannach, Balancing Consumer and Business Value of Recommender Systems: A Simulation-based Analysis, E-Commerce Research and Applications, forthcoming, https://arxiv.org/abs/2203.05952

Studying longitudinal effects: Examples

- Effects of different strategies clearly visible
  - Various configurations can be simulated and analyzed



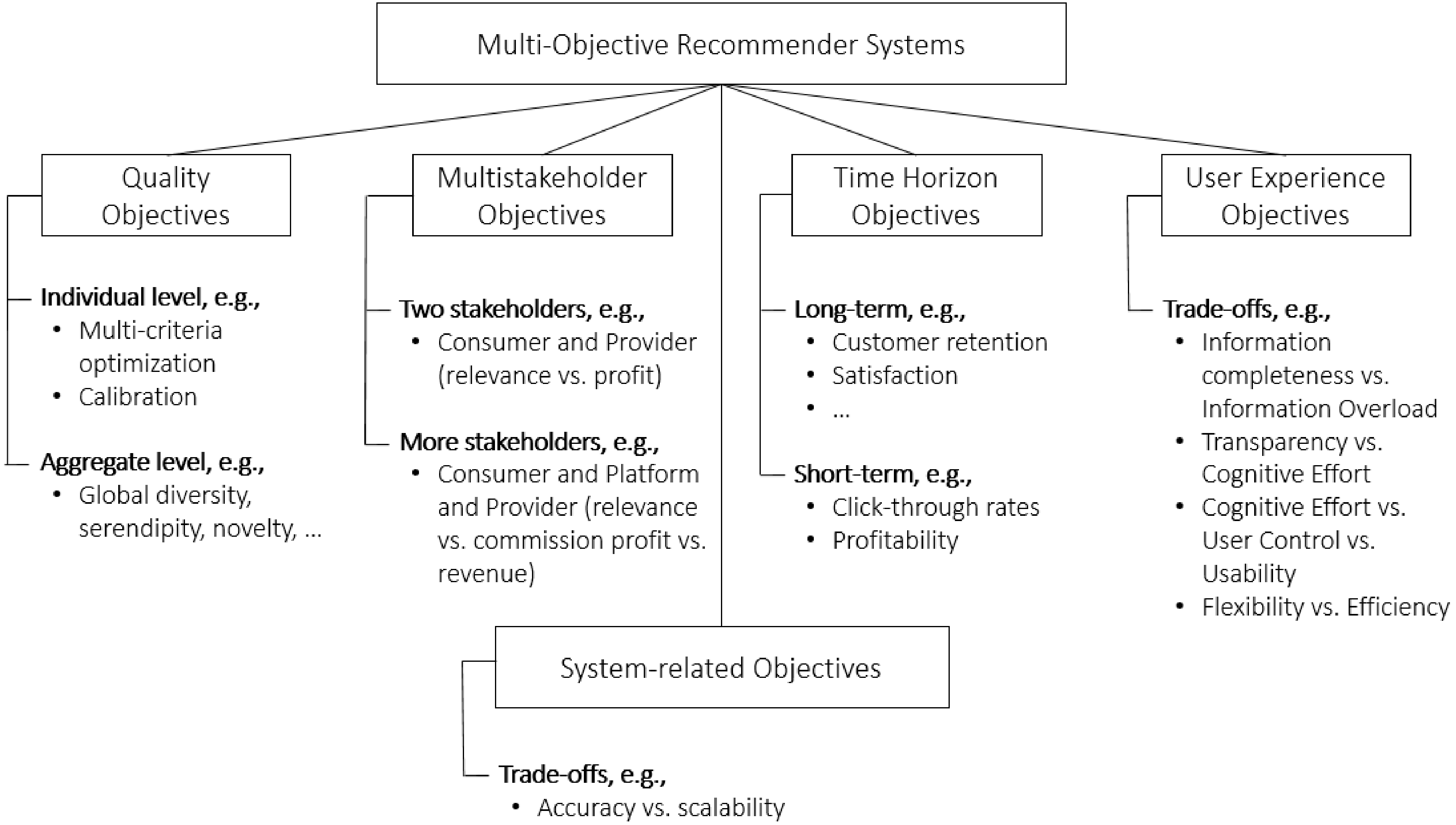
(a) Consumption probability (b) Profit per time step (c) Cumulative profit

User Experience Objectives

- Multiple objectives in the **User Interface**
  - Information completeness vs. Information Overload
    - How many items to show?
  - Transparency vs. Cognitive Effort
    - Should we show (simple or complex) explanations?
  - Cognitive Effort vs. User Control vs. Usability
    - Should we allow users to fine-tune the recommendations or strategy?
  - Flexibility vs. Efficiency
    - Should we use natural language input or web forms (buttons) in a conversational recommender chat bot?

Andrea Iovine, Fedelucio Narducci, Giovanni Semeraro, Conversational Recommender Systems and natural language: A study through the ConveRSE framework, Decision Support Systems, Volume 131, 2020, 113250

An expanded taxonomy: UI and system-related



Summary & Outlook

- In practice, recommendation may often not be a single-objective problem
  - Accuracy vs the rest, multiple stakeholders, long-term vs short-term
- Considering multiple objectives can be key to more impactful recommender systems research
  - Lack of real-world datasets currently hampers progress, cooperation with industry required
  - Simulation-based approaches may help, both agent-based ones as well as ones on reinforcement learning and counterfactual reasoning
  - Let’s be aware of abstraction traps

Thank you!

- Time for questions, contact: dietmar.jannach@aau.at
- Survey/Summary: https://tinyurl.com/mors2022
- Slides: https://tinyurl.com/mors2022-slides

Multi-Objective Recommender Systems: Survey and Challenges*

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

ABSTRACT

Recommender systems can be characterized as software solutions that provide users convenient access to relevant content. Traditionally, recommender systems research predominantly focuses on developing machine learning algorithms that aim to predict which content is relevant for individual users. In real-world applications, however, optimizing the accuracy of such relevance predictions as a single objective in many cases is not sufficient. Instead, multiple and often competing objectives have to be considered, leading to a need for more research in multi-objective recommender systems. We can differentiate between several types of such competing goals, including *(i)* competing recommendation quality objectives at the individual and aggregate level, *(ii)* competing objectives of different involved stakeholders, *(iii)* long-term vs. short-term objectives, *(iv)* objectives at the user interface level, and *(v)* system level objectives. In this paper we review these types of multi-objective recommendation settings and outline open challenges in this area.

considered a central task of a recommender and the corresponding objective was to minimize the mean absolute error (MAE), see [68] for work using MAE in 1996. Nowadays, item ranking is mostly considered to be more important than rating prediction, and a variety of corresponding ranking accuracy measures are used today.

While the metrics changed over time, the research community has been working on optimizing relevance predictions in increasingly sophisticated ways for almost 30 years now. The main objective of such research is to minimize the relevance prediction error or to maximize the accuracy of the recommendations. The underlying assumption of these research approaches is that better relevance predictions lead to systems that are more valuable for their users. This seems intuitive for many practical applications, because a better algorithm should surface more relevant items in the top-N lists shown to users.

Such an assumption might however not always be true, and it was pointed out many years ago that "being accurate is not enough" [53]. A recommender system might for example present