Multi-Objective Recommender Systems: Survey and Challenges*

DIETMAR JANNACH, University of Klagenfurt, Austria

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.

Additional Key Words and Phrases: Recommender systems, Multi-objective optimization, Beyond-accuracy evaluation, Multistakeholder recommendation, Short-term and long-term objectives

1 INTRODUCTION

Generically defined, recommender systems can be characterized as *software solutions that provide users convenient access to relevant content.* The types of conveniences that such systems provide can be manifold. Historically, recommender systems were mainly designed as information filtering tools, like the early GroupLens system [62] from 1994. Later on, various other ways were investigated how such systems can create value, e.g., by helping users *discover* relevant content, by providing easy access to related content (e.g., accessories), or by even taking automatic action like creating and starting a music playlist.

While a recommender systems can serve various purposes and create value in different ways [40], the predominant (implicit) objective of recommender systems in literature today can be described as "guide users to relevant items in situations of information overload", or simply "find good items" [26]. The most common way of operationalizing this information filtering problem is to frame the recommendation task as a supervised machine learning problem. The core of this problem is to learn a function from noisy data, which accurately predicts the *relevance* of a given item for individual users, sometimes also taking contextual factors into account.

Although the actual relevance of recommended items can be assessed in different ways [25], data-based offline experiments dominate the research landscape. In early years, rating prediction was 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.

Author's address: Dietmar Jannach, dietmar.jannach@aau.at, University of Klagenfurt, Austria.

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 users with obvious recommendations, e.g., recommending new Star Wars sequels to a Star Wars lover. The prediction error for such recommendations might be even close to zero. But so will be the value of the recommendations to users, who most probably know these movies already. Observations like this led to a multitude of research efforts on "beyond-accuracy" measures like diversity, novelty, or serendipity, see [12] for an early work from 2001. Which beyond-accuracy dimension is relevant for a given setting depends on the purpose the recommender is intended to serve [32]. Independent of the purpose, beyond-accuracy measures however often compete with accuracy measures, leading to the problem that multiple objectives have to be balanced when serving recommendations.

Historically, when considering the purpose of a recommender system, the focus of research was on the value of such a system for *consumers*. Only in recent years more attention was paid to the fact that recommender systems in practice factually serve some business or organizational objectives. Considering these provider-side aspects therefore requires that we see recommendation as a problem where the interests and objectives of multiple stakeholders must be considered [1, 2], often also taking different optimization time horizons into account.

Overall, while being able to predict the relevance of individual items for users remains to be a central and relevant problem, considering only one type of objectives, i.e., prediction accuracy, and the corresponding metrics may be too simplistic and ultimately limit the impact of academic research efforts in practice. One important way to escape the limitations of current research practice is to consider multiple quality dimensions and stakeholder objectives in parallel [34]. Next, in Section 2, we will discuss various forms of multi-objective recommender systems found in the literature.

2 TYPES OF MULTI-OBJECTIVE RECOMMENDATION SETTINGS

On a very general level, we can define that "*a multi-objective recommender system (MORS) is a system designed to jointly optimize or balance more than one optimization goal.*" Figure 1 provides an taxonomy of different and mostly orthogonal types of multi-objective recommendation settings. We note that the various objectives to be pursued with a recommender system are in many cases competing. Diversity goals, for instance, typically stand in contrast with accuracy. However, not all objectives necessarily represent such a contrast. Increasing the musical coherence of a playlist for example showed to be advantageous in terms of accuracy in some cases in [36]. Also, when considering short-term and long-term objectives, taking measures to increase interactivity and engagement with the system in the short term is usually considered beneficial for customer retention in the long run [24]. Note that while most discussed objectives are *algorithmic* ones, there may also be *non-algorithmic* objectives at the user experience level, i.e., objectives that are not tied to a specific underlying algorithm.

2.1 Recommendation Quality Objectives

Under this category, we subsume problem settings, where more than one quality objective of recommendations for users must be considered. We can differentiate between the system considering such objectives at the level of individual users or at an aggregate level, i.e., for the entire user base.

Individual level. At the individual level, consumers can have specific (short-term) preferences, e.g., regarding item features that should be considered in parallel. For instance, a user of a hotel booking platform might be interested in a relatively cheap hotel, which in addition is close to the city center. In such a situation, a recommender system might

```
┌─────────────────────────────────────────┐
│     Multi-Objective Recommender Systems  │
└─────────────────────────────────────────┘
```

Quality Objectives

Individual level, e.g.,
- Multi-criteria optimization
- Calibration

Aggregate level, e.g.,
- Global diversity, serendipity, novelty, ...

Multistakeholder Objectives

Two stakeholders, e.g.,
- Consumer and Provider (relevance vs. profit)

More stakeholders, e.g.,
- Consumer and Platform and Provider (relevance vs. commission profit vs. revenue)

Time Horizon Objectives

Long-term, e.g.,
- Customer retention
- Satisfaction
- ...

Short-term, e.g.,
- Click-through rates
- Profitability

User Experience Objectives

Trade-offs, e.g.,
- Information completeness vs. Information Overload
- Transparency vs. Cognitive Effort
- Cognitive Effort vs. User Control vs. Usability
- Flexibility vs. Efficiency

System-related Objectives

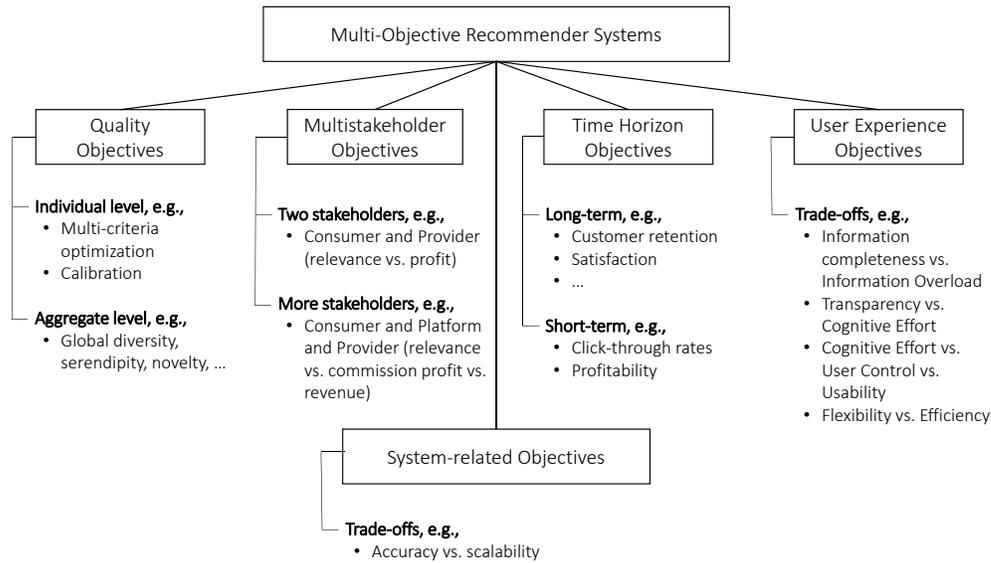**Trade-offs, e.g.,**
- Accuracy vs. scalability

Fig. 1. Taxonomy of different types of multi-objective recommendation settings.

therefore strive to find an offering that respects both preferences as good as possible. Such problem settings often occur in *interactive* or *conversational* recommendation scenarios. Technically, a variety of Multi-Criteria Decision-Making (MCDM) methods such as Analytic Hierarchical Process (AHP), the Weighted Sum Method (WSM) or Multi-Attribute Utility Theory (MAUT) can be applied, see [48, 74]. Such methods can also be part of interactive constraint-based recommendation approaches [31] to rank items. In addition, some of these constraint-based approaches support the automatic relaxation of individual user preferences in case none of the remaining items matches all consumer preferences [20].

A different way to consider multiple user preferences at the individual level is called *calibration*. In such an approach, the idea usually is to match certain beyond-accuracy aspects of the recommendation list with past preference profiles of individual users. In an early work, Oh et al. [57] tried to align the recommendations with the past popularity tendencies of a user. Later, Jugovac et al. [42] extended the approach for multiple optimization objectives. A more formal characterization of calibration was introduced in [69] by Steck, and [4] represents another recent work into that direction. In most cases the central idea of these approaches is to match two distributions, e.g., the popularity distribution of items in the user profile and the popularity distribution of recommendations. An alternative optimization goal was used in [36] for the music domain, where the objective was to find musically coherent playlist continuations while preserving prediction accuracy.

Aggregate level. The majority of published research on balancing different recommendation quality aspects targets the aggregate level. The objective of such works is to balance the recommendations for the entire user base, the corresponding metrics are therefore usually averages. The most common beyond-accuracy measures in the literature include diversity, novelty, serendipity, catalog coverage, popularity bias or fairness, see, e.g., [3, 6, 18, 43, 75]. Most commonly, the goal is to balance accuracy with exactly one of these measures, assuming that there is a trade-off between these quality factors. Increasing diversity is for example commonly assumed having a negative impact on accuracy

metrics. A few works exist which consider more than two factors. In an earlier work in this area [64], the authors describe an effort to build a *talent recommendation* system at LinkedIn, which not only considers the semantic match between a candidate profile and a job, but which also take side constraints into account, for instance, the presumed willingness of a candidate to change positions.

Technically, a variety of approaches to balance competing goals can be found in the literature. Re-ranking accuracy-optimized lists is probably the most common approach and was also used in early approaches for diversification [12]. Alternative techniques for creating a balanced lists or for merging different lists include constraint optimization techniques, graph-based and neighborhood-based approaches, methods that are based on the concept of Pareto efficiency, methods that integrate different objectives in their loss function, or bandit-based approaches that address the explore-exploit problem for improved discovery through novel recommendations [16, 29, 30, 52, 63, 66, 79–81].

2.2 Multistakeholder Objectives

The beyond-accuracy quality metrics discussed in the previous section were historically mostly introduced to improve recommendations for end users. Higher diversity, for example, should avoid monotonicity, and novelty should support discovery. The underlying assumption—also of pure accuracy-oriented works—is that improving different quality aspects for users would be at least indirectly beneficial for the providers. Only in recent years, more attention was paid in the literature to the fact that many recommendation scenarios in the real world are situated in environments, where the objectives of multiple stakeholders have to be considered. The common players in such *multistakeholder* recommendation problems include *end consumers*, recommendation *service providers*, item *suppliers*, and sometimes even parts of a broader *society* [1, 34]. In such settings, a recommender system may serve different purposes for different stakeholders [32], and the related objectives may stand in conflict.[1]

A typical problem setting in practice that involves *two stakeholders* is that of balancing consumer and provider objectives. In many cases, there may be a potential trade-off between (a) recommending the *most relevant* items for consumers and (b) recommending items that are also somewhat relevant but assumed to be favorable in terms of the provider's business objectives. Some of the discussed beyond-accuracy metrics can actually be seen as serving both stakeholders. Making more novel recommendations not only potentially leads to a better user experience, but also to more engagement with the service and longer term customer retention [8].

A number of research works however also consider business (or: organizational) objectives more directly, in particular in the form of recommender systems that are "'price and profit aware" [33]. In [14], for instance, two heuristic profit-aware strategies are proposed and the authors found that such methods can increase the profit from cross-selling without losing much recommendation accuracy. A number of other possible solutions to maximize provider profitability were proposed over the years, using a variety of approaches including mathematical optimization with side constraints, re-ranking accuracy-optimized lists based on profit considerations, random walks over graphs or analytical modeling [9, 9, 59, 76]

Besides such situations with potential consumer-provider trade-offs, there are settings that involve even more stakeholders to consider. A problem that has been studied for several years—even though not under the name multi-stakeholder recommendation—is that of *group recommendation* [50]. In such settings, the system's goal is to determine a set of recommendations that suit the preferences of a group of users, e.g., friends who want to watch a movie together. A variety of strategies to aggregate the individual user preferences were proposed over the years. Early works on the

---

[1]In some cases there may even be subgroups within the consumer stakeholder group, e.g., free vs. premium or new vs. existing customers, for which different objectives may exist.

topic can be found in [56] and [49]. In [49], for instance, Masthoff reports the outcomes of different user studies aimed to understand how humans make choices for a group and finds that humans indeed sometimes follow strategies inspired from Social Choice Theory. We note here that the group recommendation setting differs from other multistakeholder scenarios in that all stakeholders receive the same set of recommendations.

Reciprocal recommendation is another specific set of problem settings involving multiple stakeholders. Here, instead of recommending items to users, the problem is to recommend users to users, also known as people-to-people recommendation. Typical application scenarios are recommendations on dating and recruiting platforms. A particularity of such settings is that the success of a recommendation is not determined solely by the recipient of the recommendation, but there must be a mutual preference match or compatibility between the two people involved, see [58] for an in-depth discussion on the topic. The recommendation service provider therefore faces additional complexities in the matching process and in parallel has to observe its own business objectives and constraints. On a job recommendation platform, for example, the provider may have to additionally ensure that each paid job advertisement receives a minimum number of relevant impressions, i.e., exposure [5].

Similar considerations may generally apply when the recommendation platform serves as a marketplace with multiple suppliers of identical or comparable items. Let us consider again the example of a typical hotel booking platform, which serves personalized recommendations to its users [34]. Besides the consumer, who already might have competing objectives, there are the property owners, who have their offerings listed on the booking platform and pay a commission for each booking. The goal of the property owners is that their offerings are exposed to as many matching customers as possible in order to increase the chances of being booked. The booking platform, finally, is not only interested in recommending matching hotels to consumers, but might also seek to maximize their commission, e.g., by recommending slightly more expensive hotels. In addition, to balancing these objectives, the platform may furthermore have to ensure that *all* listed properties reach a sufficient level of exposure, i.e., chance of being booked. This may be required to ensure a long-term relationship with property owners, who might otherwise discontinue to list their offerings on the platform at some stage [44]. We discuss short-term and long-term objectives next.

2.3 Time Horizon Objectives

In some application domains, it might be quite simple to increase short-term Key Performance Indicators (KPIs). In the hotel booking scenario which we have just discussed, boosting short-term revenue might be achieved by recommending hotels with currently discounted rates, which maximizes the probability of a transaction [38]. In the news domain, recommending articles on trending topics, articles with click-bait headlines or generally popular content such as celebrity gossip may lead to high click-through-rates (CTR). In the music domain, recommending tracks of trending or popular artists, which the user already knows, may be a safe strategy when the target metric is to avoid "skip" events.

Such strategies that are successful in the short term may however be non-optimal or even detrimental in the long run. The recommendation of discounted hotel rooms may be bad for profit. News readers may be disappointed when actually reading articles with a clickbait headline and may not trust these recommendations in the future. Music listeners finally may have difficulties to discover new artists over time and may quit using the service after some time.

Most academic research is based on one-shot evaluations, typically focusing on the prediction accuracy given a static dataset and a certain point in time. Longitudinal effects of different recommendation strategies are much less explored and there is also limited literature on the long-term effects of recommender systems in industry. A/B tests in industry may last from a few weeks to several months. In [24], the case of Netflix is discussed, where one main KPI is customer retention, which is oriented towards the long-term perspective. In their case, attributing changes in the

recommender system to such long-term effects is reported to be challenging, e.g., because of already high retention rates and the need for large user samples. Other reports from real-world deployments a recommender systems can be found in [59] or [45]. In [45], the authors for example found that using a recommender system led to decreased sales diversity compared to a situation without a recommender. A similar effect was reported in [8], where the recommender system on a music streaming site led to a reduced consumption diversity. A survey of other reports on real-world applications of recommender systems can be found in [35].

Given the limitations of one-shot evaluations, we have observed an increased interest in longitudinal studies in recent years. One prominent line of research lies in the area of reinforcement learning (RL) approaches in particular in the form of contextual bandits, see e.g. [46] for an earlier work in the news domain. In such approaches, the system sequentially selects items to recommend to users and then incorporates the users' feedback for subsequent recommendations. Different recommendation algorithms can be evaluated offline with the help of simulators, e.g., [51, 65]. A common challenge in this context is to ensure that such evaluations are unbiased [27, 47]. We note that the consideration of temporal aspects such as different time horizons or delayed feedback have been explored in the RL literature for the related problem of computational advertising for several years [13, 72].

Reinforcement learning approaches typically aim at finding a strategy to maximize the expected *reward*. During the last few years, a number of studies that use other forms of simulations were published that focus on other important long-term phenomena of recommender systems. These studies for example focus on longitudinal effects of recommender systems on sales diversity [22], potential reinforcement effects in terms of popularity bias and other aspects for traditional and session-based recommendations [21, 37], longitudinal performance effects of recommender systems and the "performance paradox" [78], differences in terms of long-term effects of consumer-oriented and profit-oriented recommendation strategies [23].

2.4   User Experience Objectives

Going beyond specifics of individual algorithms, there can be also various objectives to be pursued at the user interaction level of a recommender system. The design space for the user interface of recommender systems is actually large, see [41], and there thus may be a number of competing objectives at the user interface (UI) level.

Here, we only list a few examples of potential trade-offs that may be common for many recommender system applications.

- *Information Completeness vs. Information Overload*: This, for instance, refers to the question of how many items should be shown to users and if we should completely filter out certain items from the result list. Showing too few options may give users the feeling that the system holds back some information. If there is too much information users will find themselves again in a situation of information overload [7, 11]. Besides the question of how many options to show, a related question is how much detail and additional information to show for each recommendation.
- *Transparency and User Control vs. Cognitive Effort*: Transparency and explanations are commonly considered to be trust-establishing factors in recommender systems [61]. A variety of different ways of explaining recommendations were proposed in the literature [55, 73]. Many of these academic proposals are quite complex and may easily cognitively overload average end users. Similar considerations apply for approaches that implement mechanisms for *user control* in recommender systems [19, 39].

- *Flexibility vs. Efficiency*: This question arises in the context of modern conversational recommender systems that are implemented in the form of chatbots. Chatbots typically support two forms of interactions: a) natural language input and b) form-based input (i.e., using buttons). While natural language inputs may allow for more flexible interactions, the study in [28], for instance, indicated that a combination of interaction modalities was most effective.

Several other more general design trade-offs may exist depending on the specific application, e.g., regarding acceptable levels automating adaptivity of the user interface, which may hamper usability [60].

2.5 System Objectives

In this final category we discuss technical aspects and their potential trade-offs. We call them "system objectives", as they refer to more general system properties.

One such trade-off in practice may lie in the complexity of the underlying algorithms and the gains that one may obtain in terms of business-related KPIs. Already in the context of the Netflix Prize [10] we could observe that the winning solutions were finally not put into production, partly due to their complexity. Similar considerations can be made for today's sometimes computationally demanding methods based on deep learning. In some cases there might be a diminishing return of deploying the most sophisticated models in production, only because they lead to slightly better accuracy values in offline testing. In some research works, it even turns out that "embarrassingly shallow" models can be highly competitive in offline evaluations [70].

With highly complex models, not only scalability issues may arise and monetary costs for computing resources may increase, the complexity of the architectures might also make such systems more difficult to maintain, debug, and explain. On the other hand, solutions built upon modern deep learning frameworks are sometimes reported to be advantageous over conceptually simpler, but specialized solutions, because these frameworks and deep learning architectures make it very easy to integrate various types of information into the models [71].

3 SUMMARY AND CHALLENGES

Our review outlines that providing automated recommendations is a problem that may require the consideration of more than one objective in many real-world use cases. Such multi-objective settings may include competing objectives of consumers, possible tensions between goals of different stakeholders, conflicts when optimizing for different time horizons, competing design choices at the UI level, as well as system-level and engineering-related considerations. In this work, we reviewed the literature in this area and provided a taxonomy to organize the various dimensions of multi-objective recommendation. We note here that the categories of the taxonomy are not mutually exclusive. For instance, a multi-objective recommendation approach may address both aspects regarding different time horizons as well as the possibly competing goals of the involved stakeholders.

In practice, one main challenge may usually lie in deciding on the right balance between the competing goals from an organizational perspective. Various stakeholders from different organizational units may have to agree on such decisions, and corresponding KPIs need to be defined and monitored. Given these KPIs, suitable optimization goals and possibly proxy measures have to be implemented and validated at the technical level.

In academic settings, researchers typically abstract from the specifics of a given application context, aimed at developing generalizable algorithmic solutions to deal with multi-objective problem settings. This abstraction process commonly involves the use of *offline evaluation approaches*, the establishment of certain assumptions, and the introduction of

computational metrics which should be optimized. After such an abstraction, one main challenge, however, lies in the evaluation process and, in particular, in making sure that improvements that are observed in terms of abstract evaluation measures would translate to better systems in practice [15].

Unfortunately, in many of today's research works, we observe phenomena similar to the "abstraction traps" described by Selbst et al. [67] in the context of research on algorithmic works in *Fair Machine Learning*. In the case of competing individual-level quality goals, for example, how can we be sure that a particular diversity metric, which we optimize such as an intra-list similarity, matches human perceptions and what would be the right balance for a given application setting or an individual user? How do we know if calibrated recommendations are liked more by users, and what would be the effects of calibration on organizational goals? Answering such questions requires corresponding user studies to, e.g., validate that the computational metrics are good proxies for human perceptions.

The problem however becomes even more challenging when not even the target concepts are entirely clear. In recent years, a widely investigated multi-objective problem setting is the provision of *fair* recommendations [18]. Unfortunately, optimizing for fairness turns out to be challenging, as fairness is a societal construct. Researchers therefore came up with various types of ways of operationalizing fairness constraints. However, in many of such works, little or no evidence or argumentation is provided why the chosen fairness metrics are meaningful in practice in general or in a particular application setting, see [17] for a survey on the recent literature. In several cases, making fair recommendations is simply equated with reducing the popularity bias of recommendations, e.g., by matching it with a target distribution, which is assumed to be given. In reality, however, it is not clear if it would be fair to frequently recommend items that are not popular—these items might simply be of poor quality—or if users would even *perceive* these recommendations to be fair.[2] Similar considerations apply for many other types of multi-objective recommender systems.

Overall, these observations call for more studies involving humans in the evaluation loop and industry partners in the research process. However, only few works exist in that direction so far. An example of a user study can be found in [9], outcomes of field studies are described in [59]. An offline evaluation with real-world data from industry is done in [54], but even in this case it is not clear if the computational metrics truly correspond to the real-world goals, e.g., if more listening events on the music platform lead to higher user satisfaction as claimed. In a very recent work, Wang and colleagues from Google [77] investigated the relationship between observed user behavior—using data from a large platform—and desired long-term outcomes in terms of user experience. With their work, they aim at providing guidance for researchers and practitioners when selecting surrogate measures to address the difficult problem of optimizing for long-term objectives.

Ultimately, despite such recent progress, multi-objective recommender systems remains a highly important research area with a number of challenging research questions. Addressing such questions will however help to pave the way towards more impactful recommender systems research in the future.

REFERENCES

[1] Himan Abdollahpouri, Gediminas Adomavicius, Robin Burke, Ido Guy, Dietmar Jannach, Toshihiro Kamishima, Jan Krasnodebski, and Luiz Pizzato. 2020. Multistakeholder Recommendation: Survey and Research Directions. *User Modeling and User-Adapted Interaction* 30 (2020), 127–158.

[2] Himan Abdollahpouri and Robin Burke. 2022. Multistakeholder Recommender Systems. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, New York, NY, 647–677.

---

[2]Avoiding popularity biases in general, i.e., beyond fairness considerations, certainly is important for different reasons in practice, e.g., to ensure a certain level of exposure for novel (cold-start) items.

[3] Himan Abdollahpouri, Robin Burke, and Bamshad Mobasher. 2017. Controlling Popularity Bias in Learning-to-Rank Recommendation. In *Proceedings of the Eleventh ACM Conference on Recommender Systems (RecSys '17)*. 42–46.

[4] Himan Abdollahpouri, Masoud Mansoury, Robin Burke, Bamshad Mobasher, and Edward C. 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 2021*. 119–129.

[5] Fabian Abel, Yashar Deldjoo, Mehdi Elahi, and Daniel Kohlsdorf. 2017. RecSys Challenge 2017: Offline and Online Evaluation. In *Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys 2017*. 372–373.

[6] Gediminas Adomavicius and YoungOk Kwon. 2012. Improving Aggregate Recommendation Diversity Using Ranking-Based Techniques. *IEEE TKDE* 24, 5 (2012), 896–911.

[7] Muhammad Aljukhadar, Sylvain Senecal, and Charles-Etienne Daoust. 2012. Using recommendation agents to cope with information overload. *International Journal of Electronic Commerce* 17, 2 (2012), 41–70.

[8] Ashton Anderson, Lucas Maystre, Ian Anderson, Rishabh Mehrotra, and Mounia Lalmas. 2020. Algorithmic Effects on the Diversity of Consumption on Spotify. In *Proceedings of The Web Conference 2020 (WWW '20)*. 2155–2165.

[9] Amos Azaria, Avinatan Hassidim, Sarit Kraus, Adi Eshkol, Ofer Weintraub, and Irit Netanely. 2013. Movie Recommender System for Profit Maximization. In *Proceedings of the 7th ACM Conference on Recommender Systems* (Hong Kong, China) *(RecSys '13)*. 121–128.

[10] James Bennett, Stan Lanning, et al. 2007. The Netflix Prize. In *Proceedings of KDD cup and workshop*, Vol. 2007. 35.

[11] Dirk G. F. M. Bollen, Bart P. Knijnenburg, Martijn C. Willemsen, and Mark P. Graus. 2010. Understanding choice overload in recommender systems. In *Proceedings of the 4th Conference on Recommender Systems (RecSys '10)*. 63–70.

[12] K. Bradley and B. Smyth. 2001. Improving Recommendation Diversity. In *Proceedings of the 12th National Conference in Artificial Intelligence and Cognitive Science*. 75–84.

[13] Olivier Chapelle. 2014. Modeling Delayed Feedback in Display Advertising. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14)*. 1097–1105.

[14] Long-Sheng Chen, Fei-Hao Hsu, Mu-Chen Chen, and Yuan-Chia Hsu. 2008. Developing recommender systems with the consideration of product profitability for sellers. *Information Sciences* 178, 4 (2008), 1032–1048.

[15] Paolo Cremonesi and Dietmar Jannach. 2021. Progress in recommender systems research: Crisis? What crisis? *AI Magazine* 42, 3 (2021), 43–54.

[16] Gabriel de Souza Pereira Moreira, Dietmar Jannach, and Adilson Marques da Cunha. 2019. Contextual Hybrid Session-based News Recommendation with Recurrent Neural Networks. *IEEE Access* 7 (2019), 169185–169203.

[17] Yashar Deldjoo, Dietmar Jannach, Alejandro Bellogin, Alessandro Difonzo, and Dario Zanzonelli. 2022. A Survey of Research on Fair Recommender Systems. https://arxiv.org/abs/2205.11127

[18] Michael D. Ekstrand, Anubrata Das, Robin Burke, and Fernando Diaz. 2022. Fairness in Information Access Systems. *Foundations and Trends® in Information Retrieval* 16, 1-2 (2022), 1–177.

[19] Michael D. Ekstrand, Daniel Kluver, F. Maxwell Harper, and Joseph A. Konstan. 2015. Letting Users Choose Recommender Algorithms: An Experimental Study. In *Proceedings of the 9th ACM Conference on Recommender Systems (RecSys '15)*. 11–18.

[20] Alexander Felfernig, Gerhard Friedrich, Dietmar Jannach, and Markus Zanker. 2006. An Integrated Environment for the Development of Knowledge-Based Recommender Applications. *International Journal of Electronic Commerce* 11, 2 (2006), 11–34.

[21] Andres Ferraro, Dietmar Jannach, and Xavier Serra. 2020. Exploring longitudinal effects of session-based recommendations. In *Proceedings of the 2020 ACM Conference on Recommender Systems*. 474–479.

[22] Daniel Fleder and Kartik Hosanagar. 2009. Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity. *Management Science* 55, 5 (2009), 697–712.

[23] Nada Ghanem, Stephan Leitner, and Dietmar Jannach. 2022. Balancing Consumer and Business Value of Recommender Systems: A Simulation-based Analysis. *E-Commerce Research and Applications* forthcoming (2022).

[24] Carlos A. Gomez-Uribe and Neil Hunt. 2015. The Netflix Recommender System: Algorithms, Business Value, and Innovation. *Transactions on Management Information Systems* 6, 4 (2015), 13:1–13:19.

[25] Asela Gunawardana and Guy Shani. 2015. Evaluating Recommender Systems. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer, 265–308.

[26] Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. 2000. Explaining Collaborative Filtering Recommendations. In *CSCW '00*. 241–250.

[27] Jin Huang, Harrie Oosterhuis, Maarten de Rijke, and Herke van Hoof. 2020. Keeping Dataset Biases out of the Simulation: A Debiased Simulator for Reinforcement Learning Based Recommender Systems. In *Fourteenth ACM Conference on Recommender Systems (RecSys '20)*. 190–199.

[28] Andrea Iovine, Fedelucio Narducci, and Giovanni Semeraro. 2020. Conversational Recommender Systems and Natural Language: A study through the ConveRSE Framework. *Decision Support Systems* 131 (2020), 113250–113260.

[29] Elvin Isufi, Matteo Pocchiari, and Alan Hanjalic. 2021. Accuracy-diversity trade-off in recommender systems via graph convolutions. *Information Processing & Management* 58, 2 (2021), 102459.

[30] Tamas Jambor and Jun Wang. 2010. Optimizing Multiple Objectives in Collaborative Filtering. In *Proceedings of the Fourth ACM Conference on Recommender Systems (RecSys '10)*. 55–62.

[31] Dietmar Jannach. 2004. ADVISOR SUITE – A Knowledge-based Sales Advisory System. In *ECAI '04*. 720–724.

[32] Dietmar Jannach and Gediminas Adomavicius. 2016. Recommendations with a Purpose. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, Massachusetts, USA) *(RecSys '16)*. 7–10.

[33]   Dietmar Jannach and Gediminas Adomavicius. 2017. Price and Profit Awareness in Recommender Systems. *Proceedings VAMS Workshop at ACM RecSys 2017* CoRR abs/1707.08029 (2017).

[34]   Dietmar Jannach and Christine Bauer. 2020. Escaping the McNamara Fallacy: Towards more Impactful Recommender Systems Research. *AI Magazine* 41, 4 (2020), 79–95.

[35]   Dietmar Jannach and Michael Jugovac. 2019. Measuring the business value of recommender systems. *ACM TMIS* 10, 4 (2019), 1–23.

[36]   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 2015)*. Vienna, Austria, 187–194.

[37]   Dietmar Jannach, Lukas Lerche, Iman Kamehkhosh, and Michael Jugovac. 2015. What recommenders recommend: an analysis of recommendation biases and possible countermeasures. *User Modeling and User-Adapted Interaction* 25, 5 (2015), 427–491.

[38]   Dietmar Jannach, Malte Ludewig, and Lukas Lerche. 2017. Session-based item recommendation in e-commerce: on short-term intents, reminders, trends and discounts. *User Modeling and User-Adapted Interaction* 27, 3 (2017), 351–392.

[39]   Dietmar Jannach, Sidra Naveed, and Michael Jugovac. 2016. User Control in Recommender Systems: Overview and Interaction Challenges. In *17th International Conference on Electronic Commerce and Web Technologies (EC-Web 2016)*.

[40]   Dietmar Jannach and Markus Zanker. 2021. Impact and Value of Recommender Systems. In *Recommender Systems Handbook*, Francesco Ricci, Bracha Shapira, and Lior Rokach (Eds.). Springer US.

[41]   Michael Jugovac and Dietmar Jannach. 2017. Interacting with Recommenders - Overview and Research Directions. *ACM Transactions on Intelligent Interactive Systems (ACM TiiS)* 7, 3 (2017).

[42]   Michael Jugovac, Dietmar Jannach, and Lukas Lerche. 2017. Efficient Optimization of Multiple Recommendation Quality Factors According to Individual User Tendencies. *Expert Systems With Applications* 81 (2017), 321–331.

[43]   Marius Kaminskas and Derek Bridge. 2016. Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems. *ACM Transactions on Interactive Intelligent Systems* 7, 1 (2016), 2:1–2:42.

[44]   Jan Krasnodebski and John Dines. 2016. Considering Supplier Relations and Monetization in Designing Recommendation Systems. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. 381–382.

[45]   Dokyun Lee and Kartik Hosanagar. 2019. How Do Recommender Systems Affect Sales Diversity? A Cross-Category Investigation via Randomized Field Experiment. *Information Systems Research* 30, 1 (2019), 239–259.

[46]   Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. 2010. A Contextual-Bandit Approach to Personalized News Article Recommendation. In *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*. 661–670.

[47]   Lihong Li, Wei Chu, John Langford, and Xuanhui Wang. 2011. Unbiased Offline Evaluation of Contextual-Bandit-Based News Article Recommendation Algorithms. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining* (Hong Kong, China) *(WSDM '11)*. 297–306.

[48]   Nikos Manouselis and Constantina I. Costopoulou. 2007. Analysis and Classification of Multi-Criteria Recommender Systems. *World Wide Web* 10, 4 (2007), 415–441.

[49]   Judith Masthoff. 2004. Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers. In *Personalized Digital Television: Targeting Programs to individual Viewers*. Springer Netherlands, 93–141.

[50]   Judith Masthoff. 2015. Group Recommender Systems: Aggregation, Satisfaction and Group Attributes. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer.

[51]   James McInerney, Ehtsham Elahi, Justin Basilico, Yves Raimond, and Tony Jebara. 2021. Accordion: A Trainable Simulator for Long-Term Interactive Systems. In *Fifteenth ACM Conference on Recommender Systems (RecSys '21)*. 102–113.

[52]   James McInerney, Benjamin Lacker, Samantha Hansen, Karl Higley, Hugues Bouchard, Alois Gruson, and Rishabh Mehrotra. 2018. Explore, Exploit, and Explain: Personalizing Explainable Recommendations with Bandits. In *Proceedings of the 12th ACM Conference on Recommender Systems* (Vancouver, British Columbia, Canada) *(RecSys '18)*. 31–39.

[53]   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)*. 1097–1101.

[54]   Rishabh Mehrotra, James McInerney, Hugues Bouchard, Mounia Lalmas, and Fernando Diaz. 2018. Towards a Fair Marketplace: Counterfactual Evaluation of the Trade-off between Relevance, Fairness & Satisfaction in Recommendation Systems. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM '18)*. 2243–2251.

[55]   Ingrid Nunes and Dietmar Jannach. 2017. A Systematic Review and Taxonomy of Explanations in Decision Support and Recommender Systems. *User Modeling and User-Adapted Interaction* 27, 3–5 (Dec. 2017), 393–444.

[56]   Mark O'Connor, Dan Cosley, Joseph A. Konstan, and John Riedl. 2001. *PolyLens: A Recommender System for Groups of Users*. Springer Netherlands, 199–218.

[57]   Jinoh Oh, Sun Park, Hwanjo Yu, Min Song, and Seung-Taek Park. 2011. Novel Recommendation Based on Personal Popularity Tendency. In *ICDM '11*. 507–516.

[58]   Iván Palomares, Carlos Porcel, Luiz Pizzato, Ido Guy, and Enrique Herrera-Viedma. 2021. Reciprocal Recommender Systems: Analysis of state-of-art literature, challenges and opportunities towards social recommendation. *Information Fusion* 69 (2021), 103–127.

[59]   Umberto Panniello, Shawndra Hill, and Michele Gorgoglione. 2016. The Impact of Profit Incentives on the Relevance of Online Recommendations. *Electron. Commer. Rec. Appl.* 20, C (2016), 87–104.

[60] Tim F. Paymans, Jasper Lindenberg, and Mark Neerincx. 2004. Usability Trade-Offs for Adaptive User Interfaces: Ease of Use and Learnability. In *Proceedings of the 9th International Conference on Intelligent User Interfaces (IUI '04)*. 301–303.

[61] Pearl Pu, Li Chen, and Rong Hu. 2011. A User-centric Evaluation Framework for Recommender Systems. In *RecSys '11*. 157–164.

[62] 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.

[63] Marco Tulio Ribeiro, Nivio Ziviani, Edleno Silva De Moura, Itamar Hata, Anisio Lacerda, and Adriano Veloso. 2015. Multiobjective Pareto-Efficient Approaches for Recommender Systems. *ACM Trans. Intell. Syst. Technol.* 5, 4 (2015).

[64] Mario Rodriguez, Christian Posse, and Ethan Zhang. 2012. Multiple Objective Optimization in Recommender Systems. In *Proceedings of the Sixth ACM Conference on Recommender Systems (RecSys '12)*. 11–18.

[65] David Rohde, Stephen Bonner, Travis Dunlop, Flavian Vasile, and Alexandros Karatzoglou. 2018. RecoGym: A Reinforcement Learning Environment for the problem of Product Recommendation in Online Advertising. *arXiv preprint arXiv:1808.00720* (2018).

[66] Alan Said, Ben Fields, Brijnesh J. Jain, and Sahin Albayrak. 2013. User-Centric Evaluation of a K-Furthest Neighbor Collaborative Filtering Recommender Algorithm. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. 1399–1408.

[67] Andrew D. Selbst, Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and Abstraction in Sociotechnical Systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*. 59–68.

[68] 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)*. 210–217.

[69] Harald Steck. 2018. Calibrated Recommendations. In *ACM RecSys '18*. 154–162.

[70] Harald Steck. 2019. Embarrassingly Shallow Autoencoders for Sparse Data. In *The World Wide Web Conference (WWW '19)*. 3251–3257.

[71] Harald Steck, Linas Baltrunas, Ehtsham Elahi, Dawen Liang, Yves Raimond, and Justin Basilico. 2021. Deep Learning for Recommender Systems: A Netflix Case Study. *AI Mag.* 42, 3 (2021), 7–18.

[72] Georgios Theocharous, Philip S. Thomas, and Mohammad Ghavamzadeh. 2015. Personalized Ad Recommendation Systems for Life-Time Value Optimization with Guarantees. In *Proceedings of the 24th International Conference on Artificial Intelligence (IJCAI'15)*. 1806–1812.

[73] Nava Tintarev and Judith Masthoff. 2011. Designing and evaluating explanations for recommender systems. In *Recommender systems handbook*. Springer, 479–510.

[74] Evangelos Triantaphyllou. 2000. Multi-Criteria Decision Making Methods. In *Multi-criteria Decision Making Methods: A Comparative Study*. 5–21.

[75] Saúl Vargas and Pablo Castells. 2011. Rank and Relevance in Novelty and Diversity Metrics for Recommender Systems. In *RecSys '11*. 109–116.

[76] Hsiao-Fan Wang and Cheng-Ting Wu. 2009. A mathematical model for product selection strategies in a recommender system. *Expert Systems with Applications* 36, 3, Part 2 (2009), 7299–7308.

[77] Yuyan Wang, Mohit Sharma, Can Xu, Sriraj Badam, Qian Sun, Lee Richardson, Lisa Chung, Ed H. Chi, and Minmin Chen. 2022. Surrogate for Long-Term User Experience in Recommender Systems. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22)*. 4100–4109.

[78] Jingjing Zhang, Gediminas Adomavicius, Alok Gupta, and Wolfgang Ketter. 2019. Consumption and Performance: Understanding Longitudinal Dynamics of Recommender Systems via an Agent-Based Simulation Framework. *Information Systems Research* forthcoming (2019).

[79] Mi Zhang and Neil Hurley. 2008. Avoiding Monotony: Improving the Diversity of Recommendation Lists. In *Proceedings of the 2008 ACM Conference on Recommender Systems (RecSys '08)*. 123–130.

[80] Yuan Cao Zhang, Diarmuid Ó Séaghdha, Daniele Quercia, and Tamas Jambor. 2012. Auralist: Introducing Serendipity into Music Recommendation. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining (WSDM '12)*. 13–22.

[81] Tao Zhou, Zoltán Kuscsik, Jian-Guo Liu, Matúš Medo, Joseph Rushton Wakeling, and Yi-Cheng Zhang. 2010. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences* 107, 10 (2010), 4511–4515.