Third Workshop: Perspectives on the Evaluation of Recommender Systems (PERSPECTIVES 2023)

Alan Said
alansaid@acm.org
University of Gothenburg
Sweden

Eva Zangerle
eva.zangerle@uibk.ac.at
Universität Innsbruck
Austria

Christine Bauer
christine.bauer@plus.ac.at
Paris Lodron University Salzburg
Austria

ABSTRACT

Evaluation is important when developing and deploying recommender systems. The PERSPECTIVES workshop sheds light on the different, potentially diverging or contradictory perspectives on the evaluation of recommender systems. Building on the discussions and outcomes of the PERSPECTIVES workshops held at RecSys 2021 and 2022, the third edition of the PERSPECTIVES workshop held at RecSys 2023 brought together researchers and practitioners from academia and industry to reflect on the evaluation of recommender systems critically. The workshop featured a keynote and focused on the interactive part with discussions in small groups and the plenum. We discussed problems and lessons learned, encouraged the exchange of the many perspectives on evaluation, and aimed to move the discourse forward within the community.

CCS CONCEPTS

• **Information systems** → **Personalization**; **Recommender systems**; **Evaluation of retrieval results**; • **Human-centered computing** → **HCI design and evaluation methods**.

KEYWORDS

evaluation, personalized systems, recommender systems, methods, information retrieval

1 INTRODUCTION

Evaluation is central in every phase in a recommender system's lifecycle—in design and development and for continuous improvement while in operation. Thereby, the evaluation may assess the core performance of a system in its very sense or embrace the entire context in which the system is used [3, 5, 7, 9, 11]. Thereby, the evaluation of recommender systems may target a wide spectrum of different aspects that have to be considered and evaluated, which was the starting point for the PERSPECTIVES workshop series. While the initial two editions of the workshop (held at RecSys 2021

and 2022, respectively) were well-attended and raised interesting questions, the discourse on evaluation continues.

The goal of the PERSPECTIVES workshop series is to capture the current state of evaluation and discuss the different targets that recommender systems evaluation should strive for. Building on the previous PERSPECTIVES workshops, this third edition addressed the question: *Where should we go from here as a community?* and aims to come up with concrete steps for action. Our particular concern is to *give a voice to the various perspectives involved.* A critical interest of this workshop is integrating the perspectives from both academia and industry. A unique difference between PERSPECTIVES and other workshops is the focus on interaction, discussions, and group work. While the workshop accepts papers and published those in proceedings, the papers are not presented at the workshop so that we can focus on the interactive part. This way, we also specifically focus on guiding junior community members and further an (interdisciplinary) multi-method approach to evaluation.

From the discussions in the first PERSPECTIVES workshop held at RecSys 2021, it became clear that despite all the ongoing efforts in the field of evaluating recommender systems, there are still numerous open issues that we need to tackle in the recommender systems research and industry community [14, 15]. In the second edition of the PERSPECTIVES workshop held at RecSys 2023, we continued the discussions to bring consensus and consolidation into an ever-increasingly active field of recommender systems research. The discussions in the second PERSPECTIVES workshop substantiated that we need to take and embrace a wide(r) scope of perspectives, covering the full spectrum of factors relevant when assessing the quality of recommender systems. These include, amongst other factors: evaluation methods and experimental designs [4], impact and purposes [6, 8], levels of maturity of the system employed, the context of the recommender system (industry or academia), stakeholders [1, 2] or fairness considerations [10].

2 TOPICS OF INTEREST AND MATERIAL

The workshop solicited papers addressing the following topics:

- Case studies of difficult, hard-to-evaluate scenarios
- Evaluations with contradicting results
- Showcasing (structural) problems in recommender systems evaluation
- Integration of offline and online experiments
- Multi-stakeholder evaluation
- Divergence between evaluation goals and what is actually captured by the evaluation
- Nontrivial and unexpected experiences from practitioners

Before and during the workshop, we collected pressing issues related to the evaluation of recommender systems from participants that should be addressed during the workshop. Hence, the topics discussed during the workshop extended beyond the list of topics above. In addition, we deliberately solicited papers reporting problems and (negative) experiences and results regarding recommender systems evaluation, as we consider reflections on unsuccessful, inadequate, or insufficient evaluations as a fruitful source to provide yet another perspective on recommender systems evaluation that can spark discussion. This also includes papers reporting negative study results (which complements, rather than duplicates, the topics of the main conference track). Accordingly, submissions could also address the following themes: (a) "lessons learned" from the successful application of recommender systems evaluation or from "post mortem" analyses describing specific evaluation strategies that failed to uncover decisive elements, (b) "overview papers" analyzing patterns of challenges or obstacles to evaluation, and (c) "solution papers" presenting solutions for specific evaluation scenarios. Additionally, (d) "visionary papers" discussing the potential novel and future evaluation aspects have been considered.

The workshop materials are available on the workshop website at https://perspectives-ws.github.io/. Similar to the 2021 edition [12] and the 2022 edition [13] of the PERSPECTIVES workshop, accepted papers are published as open-access workshop proceedings via ceur-ws.org[1]. Supplemental material (e.g., presentation slides or videos) are available on the workshop website (on authors' approval). Furthermore, we also aim to submit a workshop report to SIGIR Forum (cf. [14] for the report on PERSPECTIVES 2021 and [15] on PERSPECTIVES 2022).

ACKNOWLEDGMENTS

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. https://doi.org/10.1007/s11257-019-09256-1

[2] Christine Bauer and Eva Zangerle. 2019. Leveraging Multi-Method Evaluation for Multi-Stakeholder Settings. In *Proceedings of the 1st Workshop on the Impact of Recommender Systems* (Copenhagen, Denmark, 19 September) *(ImpactRS '19)*. CEUR Workshop Proceedings, CEUR-WS.org, 3 pages. http://ceur-ws.org/Vol-2462/short3.pdf

[3] Joeran Beel, Stefan Langer, Marcel Genzmehr, Bela Gipp, Corinna Breitinger, and Andreas Nürnberger. 2013. Research Paper Recommender System Evaluation: A Quantitative Literature Survey. In *Proceedings of the International Workshop on Reproducibility and Replication in Recommender Systems Evaluation* (Hong Kong, China) *(RepSys '13)*. Association for Computing Machinery, New York, NY, USA, 15–22. https://doi.org/10.1145/2532508.2532512

[4] Asela Gunawardana, Guy Shani, and Sivan Yogev. 2022. Evaluating Recommender Systems. In *Recommender Systems Handbook* (3rd ed.), Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, New York, NY, USA, 547–601. https://doi.org/10.1007/978-1-0716-2197-4_15

[5] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. 2004. Evaluating Collaborative Filtering Recommender Systems. *ACM Transaction on Information Systems* 22, 1 (Jan. 2004), 5–53. https://doi.org/10.1145/963770.963772

[6] Dietmar Jannach and Gediminas Adomavicius. 2016. Recommendations with a Purpose. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, MA, USA) *(RecSys '16)*. Association for Computing Machinery, New York, NY, USA, 7–10. https://doi.org/10.1145/2959100.2959186

[7] Dietmar Jannach, Oren Sar Shalom, and Joseph A Konstan. 2019. Towards More Impactful Recommender Systems Research. In *Proceedings of the 1st Workshop on the Impact of Recommender Systems, co-located with 13th ACM Conference on Recommender Systems* (Copenhagen, Denmark) *(ImpactRS@RecSys 2019, Vol. 2462)*. CEUR Workshop Proceedings, CEUR-WS.org, 3 pages. http://ceur-ws.org/Vol-2462/short6.pdf

[8] Dietmar Jannach and Markus Zanker. 2022. Value and Impact of Recommender Systems. In *Recommender Systems Handbook* (3rd ed.), Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, New York, NY, USA, 519–546. https://doi.org/10.1007/978-1-0716-2197-4_14

[9] Alan Said, Domonkos Tikk, Klara Stumpf, Yue Shi, Martha Larson, and Paolo Cremonesi. 2012. Recommender Systems Evaluation: A 3D Benchmark. In *Proceedings of the Workshop on Recommendation Utility Evaluation: Beyond RMSE* (Dublin, Ireland) *(RUE '12, Vol. 910)*. CEUR Workshop Proceedings, CEUR-WS.org, 21–23. http://ceur-ws.org/Vol-910/

[10] Sirui Yao and Bert Huang. 2017. Beyond parity: Fairness Objectives for Collaborative Filtering. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Long Beach, CA, USA) *(NIPS '17)*. Curran Associates, Red Hook, NY, USA, 2925–2934. https://doi.org/10.5555/3294996.3295052

[11] Eva Zangerle and Christine Bauer. 2022. Evaluating recommender systems: survey and framework. *Comput. Surveys* 55, 8, Article 170 (2022), 38 pages. https://doi.org/10.1145/3556536

[12] Eva Zangerle, Christine Bauer, and Alan Said (Eds.). 2021. *Proceedings of the Perspectives on the Evaluation of Recommender Systems Workshop 2021, co-located with the 15th ACM Conference on Recommender Systems (RecSys 2021), Amsterdam, The Netherlands, September 25, 2021.* PERSPECTIVES '21, Vol. 2955. CEUR Workshop Proceedings, CEUR-WS.org. http://ceur-ws.org/Vol-2955

[13] Eva Zangerle, Christine Bauer, and Alan Said (Eds.). 2022. *Proceedings of the Perspectives on the Evaluation of Recommender Systems Workshop 2022, co-located with the 16th ACM Conference on Recommender Systems (RecSys 2022), Seattle, WA, USA, September 22, 2022.* PERSPECTIVES '22, Vol. 3228. CEUR Workshop Proceedings, CEUR-WS.org. http://ceur-ws.org/Vol-3228

[14] Eva Zangerle, Christine Bauer, and Alan Said. 2022. Report on the 1st Workshop on the Perspectives on the Evaluation of Recommender Systems (PERSPECTIVES 2021) at RecSys 2021. *SIGIR Forum* 55, 2, Article 16 (mar 2022), 5 pages. https://doi.org/10.1145/3527546.3527565

[15] Eva Zangerle, Christine Bauer, and Alan Said. 2022. Report on the 2nd Workshop on the Perspectives on the Evaluation of Recommender Systems (PERSPECTIVES 2022) at RecSys 2022. *ACM SIGIR Forum* 56, 2, Article 15 (2022), 4 pages. https://doi.org/10.1145/3582900.3582919

---

[1]https://ceur-ws.org