Report on the 3rd Workshop on the Perspectives on the Evaluation of Recommender Systems (PERSPECTIVES 2023) at RecSys 2023

Alan Said
University of Gothenburg
Sweden
alansaid@acm.org

Eva Zangerle
University of Innsbruck
Austria
eva.zangerle@uibk.ac.at

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

Abstract

Evaluation is a central step when developing, optimizing, and deploying recommender systems. The PERSPECTIVES 2023 workshop, held as part of the 17th ACM Conference on Recommender Systems (RecSys 2023), served as a forum where researchers from both academia and industry critically reflected on the evaluation of recommender systems. The goal of the PERSPECTIVES workshop series is to capture the current state of evaluation from different perspectives and discuss the different targets that recommender systems evaluation should strive for. In the third edition of the workshop, we discussed problems and lessons learned, and aimed to move the discourse forward within the community.

Date: 19 September 2023.

Website: https://perspectives-ws.github.io/2023/.

1 Introduction

The third workshop on Perspectives on the Evaluation of Recommender Systems (PERSPECTIVES 2023) [Said et al., 2023a] took place at the 17th ACM Conference on Recommender Systems (RecSys 2023) in Singapore, which was organized in a hybrid manner. In this report, we summarize the insights gained from the discussion.

The primary goal of the workshop was to capture the current state of evaluation from different perspectives. In this workshop, we built on the discussions from the previous two editions of the workshop [Zangerle et al., 2021, 2022b]. In the 2023 edition, we addressed the question: *Where should we go from here as a community?* and aimed to come up with concrete steps for action. We brought together academia and industry to reflect on the evaluation of recommender systems critically, while also representing the perspectives of young academic talents and established researchers in the field.

2   Organization and Structure

The workshop was organized with high interactivity in mind. Before the event, we solicited submissions for the workshop. Accepted papers were published in the workshop proceedings [Said et al., 2023b]. Furthermore, authors of accepted papers (see Section 3) presented their work in the form of 3-minute videos. These videos are available on the workshop's website[1] and formed the basis for the discussion at the workshop.

The half-day workshop was held in two sessions. The workshop started with a keynote by Noam Koenigstein, Lead of Deep Learning Technologies and Applications (DELTA) at the Department of Industrial Engineering at Tel Aviv University, Israel, with the title "Teaching Algorithms to Explain Recommender Systems: A Counterfactual Evaluation Approach?". Koenigstein introduced the Learning to eXplain Recommendations (LXR) framework[2], which employs a counterfactual-loss-function approach which helps to assess explanations in recommender systems.

After the keynote, the author of the accepted abstract submission ("A Common Misassumption in Online Experiments with Machine Learning Models" [Jeunen, 2023]) pitched his idea. Further, Tobias Vente pitched his idea on reproducibility issues. The workshop participants then chose to join a discussion for either of these two topics; overall, we had two on-site and two online groups for each topic. The discussions were subsequently wrapped up in the plenum. We further asked each of the groups to briefly summarize their discussions in a shared document.

After the plenum discussion, the organizers gave a brief overview of the previous PERSPECTIVES activities, which include three editions of the workshop, the Special Issue on "Perspectives on Recommender Systems Evaluation"[3] in ACM Transactions on Recommender Systems, and the Dagstuhl Seminar 24211 "Evaluation Perspectives of Recommender Systems: Driving Research and Education"[4]. The workshop closed with a brief discussion of ideas and reflections for future PERSPECTIVES activities.

3   Accepted Contributions

The workshop received 8 paper submissions; 5 of these were accepted to be published in the proceedings [Said et al., 2023b]. Further, the workshop received 3 abstract submissions; 1 of these was accepted to be presented at the workshop to generate ideas for the group discussions . While the submissions were heterogeneous in terms of perspectives, all focused on evaluation topics.

All accepted papers were published in the CEUR-WS volume *Proceedings of the 3rd Workshop Perspectives on the Evaluation of Recommender Systems 2023* [Said et al., 2023b]:

Paper Annotation Practices in Societally Impactful Machine Learning Applications: What are Popular Recommender Systems Models Actually Trained On?
[Sav et al., 2023]

Paper Exploring Effect-Size-Based Meta-Analysis for Multi-Dataset Evaluation
[Sertkan et al., 2023]

---

[1] https://perspectives-ws.github.io/2023/
[2] LXR's code is publicly available at https://github.com/ExplainingRecommendations/LXR.
[3] see the Call for Papers at https://tinyurl.com/SI-PERSPECTIVES
[4] https://www.dagstuhl.de/24211

Paper	Multiobjective Hyperparameter Optimization of Recommender Systems [Moscati et al., 2023]
Paper	The Effect of Random Seeds for Data Splitting on Recommendation Accuracy [Wegmeth et al., 2023]
Paper	Unveiling Challenging Cases in Text-based Recommender Systems [Haratinezhad Torbati et al., 2023]
Abstract	A Common Misassumption in Online Experiments with Machine Learning Models [Jeunen, 2023]

The corresponding videos are available on the workshop website[5].

4 Summary and Discussion

This was the third edition of the Perspectives on the Evaluation of Recommender Systems workshop (see Zangerle et al. [2022a] for the 2021 report and Zangerle et al. [2023] for the 2022 report). The discussions during the workshop substantiated that there are still many open issues and that we need to continue building a community around the topic of evaluation and advance this important aspect of recommender systems research. The discussion groups at the workshop identified reproducibility, open access to data, and online evaluation as key open challenges in the recommender systems community. All these open issues have direct impact on the evaluation of recommender systems, while not being directly related to the algorithmic performance of recommendation models.

Acknowledgments

References

Ghazaleh Haratinezhad Torbati, Anna Tigunova, and Gerhard Weikum. Unveiling challenging cases in text-based recommender systems. In Said et al. [2023b]. URL https://ceur-ws.org/Vol-3476/paper5.pdf.

Olivier Jeunen. A common misassumption in online experiments with machine learning models. In Said et al. [2023b]. URL https://ceur-ws.org/Vol-3476/abstract1.pdf.

---

[5]https://perspectives-ws.github.io/2023/

Marta Moscati, Yashar Deldjoo, Giulio Davide Carparelli, and Markus Schedl. Multiobjective hyperparameter optimization of recommender systems. In Said et al. [2023b]. URL `https://ceur-ws.org/Vol-3476/paper3.pdf`.

Alan Said, Eva Zangerle, and Christine Bauer. Third workshop: perspectives on the evaluation of recommender systems (PERSPECTIVES 2023). In *17th ACM Conference on Recommender Systems*, RecSys '23, pages 1221–1222, New York, NY, USA, 2023a. ACM. doi: 10.1145/3604915.3608748.

Alan Said, Eva Zangerle, and Christine Bauer, editors. *Proceedings of the Perspectives on the Evaluation of Recommender Systems Workshop 2023*, volume 3476 of *CEUR Workshop Proceedings*, Aachen, Germany, 2023b. CEUR-WS.org. URL `https://ceur-ws.org/Vol-3476/`.

Andra-Georgiana Sav, Andrew M. Demetriou, and Cynthia C. S. Liem. Annotation practices in societally impactful machine learning applications: What are popular recommender systems models actually trained on? In Said et al. [2023b]. URL `https://ceur-ws.org/Vol-3476/paper1.pdf`.

Mete Sertkan, Sophia Althammer, Sebastian Hofstätter, Peter Knees, and Julia Neidhardt. Exploring effect-size-based meta-analysis for multi-dataset evaluation. In Said et al. [2023b]. URL `https://ceur-ws.org/Vol-3476/paper2.pdf`.

Lukas Wegmeth, Tobias Vente, Lennart Purucker, and Joeran Beel. The effect of random seeds for data splitting on recommendation accuracy. In Said et al. [2023b]. URL `https://ceur-ws.org/Vol-3476/paper4.pdf`.

Eva Zangerle, Christine Bauer, and Alan Said. Perspectives on the evaluation of recommender systems (PERSPECTIVES). In *15th ACM Conference on Recommender Systems*, RecSys '21, pages 794–795, New York, NY, USA, 2021. ACM. doi: 10.1145/3460231.3470929.

Eva Zangerle, Christine Bauer, and Alan Said. Report on the 1st workshop on the perspectives on the evaluation of recommender systems (PERSPECTIVES 2021) at recsys 2021. *ACM SIGIR Forum*, 55(2), mar 2022a. doi: 10.1145/3527546.3527565. URL `https://doi.org/10.1145/3527546.3527565`.

Eva Zangerle, Christine Bauer, and Alan Said. Second workshop: Perspectives on the evaluation of recommender systems (PERSPECTIVES 2022). In *16th ACM Conference on Recommender Systems*, RecSys '22, pages 652–653, New York, NY, USA, 2022b. ACM. doi: 10.1145/3523227.3547408.

Eva Zangerle, Christine Bauer, and Alan Said. Report on the 2nd workshop on the perspectives on the evaluation of recommender systems (PERSPECTIVES 2022) at recsys 2022. *ACM SIGIR Forum*, 56(2), 2023. doi: 10.1145/3582900.3582919. URL `https://doi.org/10.1145/3582900.3582919`.