Report on the 2nd Workshop on the Perspectives on the Evaluation of Recommender Systems (PERSPECTIVES 2022) at RecSys 2022

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

Christine Bauer
Utrecht University
The Netherlands
c.bauer@uu.nl

Alan Said
University of Gothenburg
Sweden
alansaid@acm.org

Abstract

Evaluation is a central step when it comes to developing, optimizing, and deploying recommender systems. The PERSPECTIVES 2022 workshop at the 16th ACM Conference on Recommender Systems brought together academia and industry to critically reflect on the evaluation of recommender systems. The primary goal of the workshop was to capture the current state of evaluation from different, and maybe even diverging or contradictory perspectives.

Date: 22 September, 2022.

Website: https://perspectives-ws.github.io/2022/.

1 Introduction

The second workshop on Perspectives on the Evaluation of Recommender Systems (PERSPEC-TIVES 2022) [Zangerle et al., 2022b] took place at the 16th ACM Conference on Recommender Systems (RecSys 2022) in Seattle, WA, USA, where the first of two parts 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, and maybe even diverging or contradictory perspectives. In this workshop, we continued the discussion from the first edition of the workshop [Zangerle et al., 2021]. In the 2022 edition, we brought together academia and industry to critically reflect on the evaluation of recommender systems, 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 [Zangerle

et al., 2022c]. Furthermore, authors of accepted papers (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 workshop was held in two parts over one day: the first part was a hybrid event and the second was an in-person-only event at the RecSys 2022 venue. The workshop started with a keynote on "What is the end goal and how do we evaluate it?" by Kim Falk, Staff Recommender Engineer at Shopify. Falk provided insights into evaluation practices at Shopify. In his talk, he elaborated on how evaluations in industry differ from evaluations in academia and how the research community could bridge this gap.

After the keynote, the authors of the two accepted abstract submissions ("Should Algorithm Evaluation Extend to Testing? We Think So" [Michiels et al., 2022] and "Multi-domain news recommender systems at Globo: Challenges, Approaches and Evaluation Metrics" [Pinho Lucas et al., 2022]) pitched their ideas, each followed by a brief questions-&-answers session. 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. During the hybrid part of the workshop, workshop participants anonymously voted on the best paper of the workshop, and the hybrid part of the workshop closed with the announcement of the winner of the best paper award: Verachtert et al. [2022].

The on-site part of the workshop resembled an open discussion of issues concerning the evaluation of recommender systems. We discussed specific challenges that have been addressed in the submissions. On the other hand, the discussion embraced general issues such as reproducibility. This interactive setting brought new insights into a diverse set of topics and perspectives into the discussion.

3 Accepted Contributions

The workshop received 12 paper submissions; 6 of these were accepted to be published in the proceedings [Zangerle et al., 2022c]. Further, the workshop received 3 abstract submissions; 2 of these were accepted as discussion starters for the workshop. 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 Perspectives on the Evaluation of Recommender Systems Workshop* [Zangerle et al., 2022c]:

Paper Are We Forgetting Something? Correctly Evaluate a Recommender System With an Optimal Training Window [**Best Paper**]
[Verachtert et al., 2022]

Paper CaMeLS: Cooperative Meta-Learning Service for Recommender Systems
[Wegmeth and Beel, 2022]

Paper Diversifying Sentiments in News Recommendation
[Sertkan et al., 2022]

Paper Learning Choice Models for Simulating Users' Interactions with Recommender Systems
[Hazrati and Ricci, 2022]

---

[1]https://perspectives-ws.github.io/2022/

Paper	Recommender Systems Alone Are Not Everything: Towards a Broader Perspective in the Evaluation of Recommender Systems [Loepp, 2022]
Paper	Towards Comparing Recommendation to Multiple-Query Search Sessions for Talent Search [Kaya and Bogers, 2022]
Abstract	Should Algorithm Evaluation Extend to Testing? We Think So [Michiels et al., 2022]
Abstract	Multi-domain News Recommender Systems at Globo: Challenges, Approaches and Evaluation Metrics [Pinho Lucas et al., 2022]

The corresponding videos are available on the workshop website[2].

4 Summary and Discussion

During the workshop, participants discussed a wide range of topics revolving around the evaluation of recommender systems. These topics included (1) which aspects of recommender systems should be tested and evaluated, (2) how user intent can be incorporated and evaluated, (3) the choice of suitable baselines for a given recommendation scenario.

Inspired by the discussions during the workshop, interested researchers formed a group to create a set of guidelines for the evaluation of recommender systems to help unify the evaluation process and foster the scientific integrity of the process.

This was the second edition of the Perspectives on the Evaluation of Recommender Systems workshop (see Zangerle et al. [2022a] for the 2021 report). The discussions during the workshop (again) 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.

References

Naieme Hazrati and Francesco Ricci. Learning choice models for simulating users' interactions with recommender systems. In *Proceedings of the 2nd Perspectives on the Evaluation of Recommender Systems Workshop*, PERSPECTIVES '22. CEUR Workshop Proceedings, September 2022. URL https://ceur-ws.org/Vol-3228/paper4.pdf.

Mesut Kaya and Toine Bogers. Towards comparing recommendation to multiple-query search sessions for talent search. In *Proceedings of the 2nd Perspectives on the Evaluation of Recommender Systems Workshop*, PERSPECTIVES '22. CEUR Workshop Proceedings, September 2022. URL https://ceur-ws.org/Vol-3228/paper6.pdf.

Benedikt Loepp. Recommender systems alone are not everything: Towards a broader perspective in the evaluation of recommender systems. In *Proceedings of the 2nd Perspectives on the Evalua-*

---

[2]https://perspectives-ws.github.io/2022/

tion of Recommender Systems Workshop, PERSPECTIVES '22. CEUR Workshop Proceedings, September 2022. URL https://ceur-ws.org/Vol-3228/paper5.pdf.

Lien Michiels, Robin Verachtert, Kim Falk, and Bart Groethals. Abstract: Should algorithm evaluation extend to testing? we think so. In *Proceedings of the 2nd Perspectives on the Evaluation of Recommender Systems Workshop*, PERSPECTIVES '22. CEUR Workshop Proceedings, September 2022. URL https://ceur-ws.org/Vol-3228/abstract1.pdf.

Joel Pinho Lucas, Letícia Freire de Figueiredo, and Felipe Alves Ferreira. Abstract: Multi-domain news recommender systems at globo: Challenges, approaches and evaluation metrics. In *Proceedings of the 2nd Perspectives on the Evaluation of Recommender Systems Workshop*, PERSPECTIVES '22. CEUR Workshop Proceedings, September 2022. URL https://ceur-ws.org/Vol-3228/abstract2.pdf.

Mete Sertkan, Sofia Althammer, Sebastian Hoftstätter, and Julia Neidhardt. Diversifying sentiments in news recommendation. In *Proceedings of the 2nd Perspectives on the Evaluation of Recommender Systems Workshop*, PERSPECTIVES '22. CEUR Workshop Proceedings, September 2022. URL https://ceur-ws.org/Vol-3228/paper3.pdf.

Robin Verachtert, Lien Michiels, and Bart Goethals. Are we forgetting something? correctly evaluate a recommender system with an optimal training window. In *Proceedings of the 2nd Perspectives on the Evaluation of Recommender Systems Workshop*, PERSPECTIVES '22. CEUR Workshop Proceedings, September 2022. URL https://ceur-ws.org/Vol-3228/paper1.pdf.

Lukas Wegmeth and Joeran Beel. Camels: Cooperative meta-learning service for recommender systems. In *Proceedings of the 2nd Perspectives on the Evaluation of Recommender Systems Workshop*, PERSPECTIVES '22. CEUR Workshop Proceedings, September 2022. URL https://ceur-ws.org/Vol-3228/paper2.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. *SIGIR Forum*, 55(2), mar 2022a. ISSN 0163-5840. 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. *Proceedings of the Perspectives on the Evaluation of Recommender Systems Workshop 2022*, volume 3228 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2022c. ISBN 1613-0073. URL http://ceur-ws.org/Vol-3228/.