

Report on the 1st Workshop on the Perspectives on the Evaluation of Recommender Systems (PERSPECTIVES 2021) at RecSys 2021

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 2021 workshop at the 15th 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: 25 and 30 September 2021.

Website: <https://perspectives-ws.github.io/2021/>.

1 Introduction

The first workshop on the Perspectives on the Evaluation of Recommender Systems (PERSPECTIVES 2021) took place at the 15th ACM Conference on Recommender Systems (RecSys 2021) in Amsterdam, The Netherlands, as well as online. 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. We did not only bring together academia and industry to critically reflect on the evaluation of recommender systems but also ensured that the perspectives from young academic talent to established researchers in the field were represented.

2 Organization and Structure

The workshop was organized with a highly interactive setup. Before the event, we solicited submissions for the workshop. Accepted papers were published in the workshop proceedings [Zangerle et al., 2021]. 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 website¹ and built the basis for the discussion at the workshop.

The workshop was held in two parts: the main online-only event and a second on-site event at the RecSys 2021 venue. The workshop started with a keynote on “Recommender System Evaluation: One Gold Standard, but no Silver Bullets” by Zeno Gantner, Principal Applied Scientist at Zalando. Gantner provided deep insights into the evaluation practices at Zalando. One of the many takeaways is that although A/B testing is considered the golden standard, there is no silver bullet in recommender systems evaluations because every method has its merits and downsides.

A major part of the main (online) workshop was dedicated to discussions. In breakout rooms, participants discussed the challenges they face when evaluating recommender systems (Section 4). The discussion was subsequently wrapped up in the plenum. After a break, Michael D. Ekstrand, Ben Carterette, and Fernando Diaz presented their abstract contribution “Evaluating Recommenders with Distributions”. The discussion topic in the presentation then smoothly transitioned into a general discussion with the plenum addressing various kinds of issues and good practices in the evaluation of recommender systems. The main workshop closed with the announcement of the workshop’s best paper award: The workshop participants anonymously voted to give the best paper award to Scheidt and Beel [2021].

The on-site part of the workshop resembled an open discussion of issues concerning the evaluation of recommender systems. While many of the online participants participated in the on-site workshop as well, there were a large number of new attendees that brought new insights and perspectives into the discussion.

3 Accepted Papers

The workshop received 19 submissions, 12 of these were accepted to be published in the proceedings. While the submissions were heterogeneous in terms of perspectives, all focused on various evaluation topics. All accepted papers were published in the *Proceedings of the Perspectives on the Evaluation of Recommender Systems Workshop* [Zangerle et al., 2021]:

- *Coupled or Decoupled Evaluation for Group Recommendation Methods?*
Peska and Malecek [2021]
- *Evaluating Recommender Systems with and for Children: towards a Multi-Perspective Framework*
Gómez et al. [2021]
- *MOCHI: an Offline Evaluation Framework for Educational Recommendations*
Wang et al. [2021]

¹<https://perspectives-ws.github.io/2021/>

-
- *Modeling Online Behavior in Recommender Systems: The Importance of Temporal Context*
Filipovic et al. [2021]
 - *On Evaluating Session-Based Recommendation with Implicit Feedback*
Diaz [2021]
 - *Prediction Accuracy and Autonomy*
Angwald et al. [2021]
 - *Recommender Systems Meet Species Distribution Modelling*
Zliobaite [2021]
 - *Sequence or Pseudo-Sequence? An Analysis of Sequential Recommendation Datasets*
Woolridge et al. [2021]
 - *Statistical Inference: The Missing Piece of RecSys Experiment Reliability Discourse*
Ihemelandu and Ekstrand [2021]
 - *Time-dependent Evaluation of Recommender Systems* [Best Paper Award]
Scheidt and Beel [2021]
 - *Toward Benchmarking Group Explanations: Evaluating the Effect of Aggregation Strategies versus Explanation*
Barile et al. [2021]
 - *Unboxing the Algorithm: Designing an Understandable Algorithmic Experience in Music Recommender Systems*
Schröder and Ghajargar [2021]

4 Summary and Discussion

During the workshop, we collected ongoing challenges and problems w.r.t. the evaluation of recommender systems and topics to be discussed during the workshop. These discussions revolved around the following main themes:

- Bridging the gap between online and offline experiments
- Valid offline experiments
- Target of evaluation
- Biases that may impact the evaluation
- Domain-dependency of evaluations

Furthermore, the following questions were brought up by participants but were not discussed in the workshop due to time constraints: (i) evaluation of transparency and explanations, (ii) reproducibility of evaluations and code sharing, (iii) effect sizes and their impact on recommender systems in production, (iv) the impact of user intent and its evaluation.

The on-site discussion focused on (i) the call for standardization in evaluation practices and reporting in the RecSys community, (ii) the potential of collaborations of academia and industry (also with cooperative PhD projects), and (iii) differences and similarities to evaluation in related research domains, e.g. information retrieval, machine learning, and human-computer interaction.

Despite all the ongoing efforts in the field of evaluating recommender systems, the workshop demonstrated there are still many open issues that need to be tackled by the recommender systems research and practitioner community.

References

- Anton Angwald, Kalle Areskoug, and Alan Said. Prediction accuracy and autonomy. In *Proceedings of the Perspectives on the Evaluation of Recommender Systems Workshop*, PERSPECTIVES '21. CEUR Workshop Proceedings, September 2021. URL <http://ceur-ws.org/Vol-2955/paper6.pdf>.
- Francesco Barile, Shabnam Najafian, Tim Draws, Oana Inel, Alisa Rieger, Rishav Hada, and Nava Tintarev. Toward benchmarking group explanations: Evaluating the effect of aggregation strategies versus explanation. In *Proceedings of the Perspectives on the Evaluation of Recommender Systems Workshop*, PERSPECTIVES '21. CEUR Workshop Proceedings, September 2021. URL <http://ceur-ws.org/Vol-2955/paper11.pdf>.
- Fernando Diaz. On evaluating session-based recommendation with implicit feedback. In *Proceedings of the Perspectives on the Evaluation of Recommender Systems Workshop*, PERSPECTIVES '21. CEUR Workshop Proceedings, September 2021. URL <http://ceur-ws.org/Vol-2955/paper5.pdf>.
- Milena Filipovic, Blagoj Mitrevski, Diego Antognini, Emma Lejal Glaude, Boi Faltings, and Claudiu Mjusat. Modeling online behavior in recommender systems: The importance of temporal context. In *Proceedings of the Perspectives on the Evaluation of Recommender Systems Workshop*, PERSPECTIVES '21. CEUR Workshop Proceedings, September 2021. URL <http://ceur-ws.org/Vol-2955/paper4.pdf>.
- Emilia Gómez, Vicky Charisi, and Chaudron Stephane. Evaluating recommender systems with and for children: towards a multi-perspective framework. In *Proceedings of the Perspectives on the Evaluation of Recommender Systems Workshop*, PERSPECTIVES '21. CEUR Workshop Proceedings, September 2021. URL <http://ceur-ws.org/Vol-2955/paper2.pdf>.
- Ngozi Ihemelandu and Michael D. Ekstrand. Statistical inference: The missing piece of RecSys experiment reliability discourse. In *Proceedings of the Perspectives on the Evaluation of Recommender Systems Workshop*, PERSPECTIVES '21. CEUR Workshop Proceedings, September 2021. URL <http://ceur-ws.org/Vol-2955/paper9.pdf>.
- Ladislav Peska and Ladislav Malecek. Coupled or decoupled evaluation for group recommendation methods? In *Proceedings of the Perspectives on the Evaluation of Recommender Systems Workshop*, PERSPECTIVES '21. CEUR Workshop Proceedings, September 2021. URL <http://ceur-ws.org/Vol-2955/paper1.pdf>.
- Teresa Scheidt and Joeran Beel. Time-dependent evaluation of recommender systems. In *Proceedings of the Perspectives on the Evaluation of Recommender Systems Workshop*, PERSPECTIVES '21. CEUR Workshop Proceedings, September 2021. URL <http://ceur-ws.org/Vol-2955/paper10.pdf>.
- Anna Marie Schröder and Malileh Ghajargar. Unboxing the algorithm: Designing an understandable algorithmic experience in music recommender systems. In *Proceedings of the Perspectives on the Evaluation of Recommender Systems Workshop*, PERSPECTIVES '21. CEUR Workshop Proceedings, September 2021. URL <http://ceur-ws.org/Vol-2955/paper12.pdf>.

-
- Chunpai Wang, Shaghayegh Sahebi, and Peter Brusilovsky. MOCHI: an offline evaluation framework for educational recommendations. In *Proceedings of the Perspectives on the Evaluation of Recommender Systems Workshop*, PERSPECTIVES '21. CEUR Workshop Proceedings, September 2021. URL <http://ceur-ws.org/Vol-2955/paper3.pdf>.
- Daniel Woolridge, Sean Wilner, and Madeleine Glick. Sequence or pseudo-sequence? An analysis of sequential recommendation datasets. In *Proceedings of the Perspectives on the Evaluation of Recommender Systems Workshop*, PERSPECTIVES '21. CEUR Workshop Proceedings, September 2021. URL <http://ceur-ws.org/Vol-2955/paper8.pdf>.
- Eva Zangerle, Christine Bauer, and Alan Said, editors. *Proceedings of the Perspectives on the Evaluation of Recommender Systems Workshop 2021*, volume 2955 of *CEUR Workshop Proceedings*, 2021. CEUR-WS.org. URL <http://ceur-ws.org/Vol-2955/>.
- Indre Zliobaite. Recommender systems meet species distribution modelling. In *Proceedings of the Perspectives on the Evaluation of Recommender Systems Workshop*, PERSPECTIVES '21. CEUR Workshop Proceedings, September 2021. URL <http://ceur-ws.org/Vol-2955/paper7.pdf>.