Report on the Workshop on Learning and Evaluating Recommendations with Impressions (LERI) at RecSys 2023

Maurizio Ferrari Dacrema
Politecnico di Milano
Italy
maurizio.ferrari@polimi.it

Pablo Castells
Universidad Autónoma de Madrid and Amazon
Spain
pablo.castells@uam.es

Justin Basilico
Netflix
USA
jbasilico@netflix.com

Paolo Cremonesi
Politecnico di Milano and ContentWise
Italy
paolo.cremonesi@polimi.it

Abstract

The *Workshop on Learning and Evaluating Recommendations with Impressions* (LERI) was held in conjunction with the 17th ACM Conference on Recommender Systems (RecSys 2023). The program included a keynote, a panel discussion and 7 paper presentations. The proceedings of the workshop are available online.[1] The LERI workshop focused on all aspects related to the use of impressions for recommendation with the aim to bring the community together and share experience and perspectives. Recommender systems typically rely on past user interactions as the primary source of information for making predictions. Impressions are an important source of information that indicate the items displayed on screen when the user interacted (or not) with them, and have the potential to impact the field of recommender systems in several ways. Impressions present new research questions and opportunities, but also bring new challenges.

Date: 19 September 2023.

Website: https://recsyspolimi.github.io/leri2023/.

1 Motivation

In recent years a source of information that was previously almost unavailable to the wider research community has emerged with the potential to impact the field in numerous ways: *impressions*. Impressions [Liang et al., 2016; Shih et al., 2016; Sato et al., 2019; Eide et al., 2021; Xin et al., 2023] refer to the items displayed on the screen when a user interacts (or not) with them and are the output of a whole recommendation engine [Pérez Maurera et al., 2020; Eide et al., 2021; Pérez

[1] https://ceur-ws.org/Vol-3590/

Maurera et al., 2022] or other surfaces within a product. Impressions constitute a nuanced and intricate data source that raises novel research questions, opportunities, and challenges. These may have profound implications for how recommender systems are conceptualized, trained, and evaluated.

Until recently, research on the use of impressions was primarily conducted by companies with the lack of publicly available datasets limiting research on the topic. This is rapidly changing and most of the available datasets including impressions have been published in the last few years: e.g., ContentWise Impressions [Pérez Maurera et al., 2020], MIND [Wu et al., 2020], FINN.no Slates [Eide et al., 2021], Pandor [Sidana et al., 2018], Ali-CCP [Ma et al., 2018], Alimama [Shen et al., 2022], Cross-shop Combo [Zhu et al., 2022], In-Shop Combo [Zhu et al., 2022], Kwai FAIR System [Wang et al., 2022], Kwai FAIR Experiment [Wang et al., 2022]. However, despite the increasing research interest, the efforts devoted to studying the use of impressions are still limited and fragmented. Therefore, one of the core aims of the LERI workshop was to bring together and consolidate the community working on this topic.

Workshop Topics. The *Workshop on Learning and Evaluating Recommendations with Impressions* focused on all aspects related to leveraging impression data to build and evaluate a recommendation engine. The goal was to both help to coalesce researchers exploring the use of impressions from different perspectives, as well as foster increased interest from the community for this new and still largely underexplored topic that has the potential to impact the field in several ways. The workshop aimed to provide a venue for researchers and practitioners to come together in order to: (i) share experience and lessons learned; (ii) identify key challenges in the area; (iii) build a common mental model and conceptual framework for thinking and researching on the use of impressions; (iv) identify emerging topics and new opportunities. The workshop also aimed to lay bridges between practitioners and academics, encourage a wider availability of impression data sources and leverage industry's experience to guide and inform academic research.

2 Workshop Contributions

The workshop was articulated in a keynote, a paper presentation session and a panel discussion.

2.1 Keynote by Jiangwei Pan

The keynote was given by **Jiangwei Pan**, titled: "Recommendation Modeling with Impression Data at Netflix".[2] Jiangwei Pan is a research scientist at Netflix, where he works on algorithms for homepage recommendations. Prior to Netflix, he was a research scientist at Meta. He obtained a Ph.D. from Duke University in 2016.

The keynote focused on how impression data is used to build recommender models at Netflix, in particular for training a recommendation model and compute additional features. One first issue that arises when using impressions to train a model is that impressions usually only contain potentially relevant items. Due to this, a model trained on impressions alone will not be able to distinguish between relevant and non-relevant items, exhibiting poor generalizability. One

[2] https://www.slideshare.net/JiangweiPan/recommendation-modeling-with-impression-data-at-netflixpptx

potential solution is to inject a small number of random recommendations to the user in order to increase the item exploration, but this strategy is limited in real applications by the need to avoid harming the user experience. Another option is to sample negative items during the training process among those that were not impressed, but the right distribution must be chosen.

In applications where the item space is large, in the range of millions, it is common to use a two-pass architecture. The first pass, called *candidate generation*, is done by a fast model tasked to select a few hundreds of loosely-relevant items. The second pass, called *fine-grained ranker*, is tasked to sort the selected items with a more sophisticated model to distinguish between *good* and *excellent* candidates and generate the final recommendations. For this reason, the fine-grained ranker is well suited to be trained on impressions, because it is applied only on relevant item candidates. Furthermore, impressions may have long-term value by making the user more familiar with the content. It was observed that repeated impressions affect the user likelihood of interacting with an item in the future, sometimes increasing it others reducing it. Impressions data can also be used to define features used by the models, such as the number of past impressions or the ratio between engagements and impressions.

Using impressions brings several challenges in real applications. A first challenge is how to account for the user scrolling behavior producing repeated and noisy impressions. Another important challenge is the huge data volume of impressions, much higher than user interactions. Impressions are challenging to log on the heterogeneous client devices as well as to process at scale. Common strategies includes summarizing or sampling them. Lastly, when impressions correlate with the likelihood of user interactions, one might be tempted to continue to recommend that item, but this may not be a good strategy because other factors may be at play. In experiments it was observed that a model learned instead to recommend less impressed items.

Among the most important future challenges are how to do a more efficient item exploration without affecting the user experience, how to sample negative items mixing negatives and impressed but non-interacted items, as well as how to better model the long-term value of impressions.

The keynote was followed by a discussion with the audience in which several aspects related to using impressions in practice were discussed. For example how to select negative samples and how to account for the fact that not all impressions are produced by the recommendation algorithm but may be due to editorial choices or business rules, implying that they may be less relevant for the user but may help in exploring the item space.

2.2 Paper Session

During the paper session, seven contributions were presented:

- *Impression-Informed Multi-Behavior Recommender System: A Hierarchical Graph Attention Approach*; Dong Li; Divya Bhargavi; Vidya Sagar Ravipati [Li et al., 2023]
- *Characterizing Impression-Aware Recommender Systems*; Fernando Benjamín Pérez Maurera; Maurizio Ferrari Dacrema; Pablo Castells; Paolo Cremonesi [Pérez Maurera et al., 2023a]
- *Effects of Human-curated Content on Diversity in PSM: ARD-M Dataset*; Marcel Hauck; Ahtsham Manzoor; Sven Pagel [Hauck et al., 2023]

- *Formulating Video Watch Success Signals for Recommendations on Short Video platforms*; Srijan Saket; Venkata Sai Baba Reddy Velugoti; Rishabh Mehrotra [Saket et al., 2023]
- *Offline Evaluation using Interactions to Decide Cross-selling Recommendations Algorithm for Online Food Delivery*; Manchit Madan [Madan, 2023]
- *Contextual Position Bias Estimation using a Single Stochastic Logging Policy*; Giuseppe Di Benedetto; Ben London; Alexander Buchholz; Yannik Stein; Vito Bellini; Matej Jakimov; Matteo Ruffini; Thorsten Joachims [Di Benedetto et al., 2023]
- *Incorporating Impressions to Graph-Based Recommenders*; Fernando Benjamín Pérez Maurera; Maurizio Ferrari Dacrema; Pablo Castells; Paolo Cremonesi [Pérez Maurera et al., 2023b]

2.3 Panel Discussion

The panel discussion was moderated by Paolo Cremonesi, with panelists: **Jiangwei Pan** (Netflix), **Arnab Bhadury** (YouTube), **Srijan Saket** (Sharechat). The panelists agreed that research in impressions for recommendation systems is primarily driven by companies because this is a very important problem in real applications. One of the factors limiting the participation of academia is the lack of datasets, of which only few were publicly available until recently. One issue however is that publicly available data is preprocessed and cleaned in such a way that might not represent the full spectrum of a real scenario. A further challenge is that impressions is very contextual and depends on the scenario, for example an impression that the user did not interact with may be a weakly positive signal in one context but a weakly negative in another depending on the user interface and the type of content. Due to this it is very difficult to build a generic dataset.

The moderator observed that most presentations in the workshop came from companies working on the video and music recommendation domain, asking whether the difficulty of obtaining good metadata for multimedia content could play a role. The panelists agreed that in multimedia recommendation it is more difficult to obtain good metadata and since items have a limited shelf life impressions are useful to provide more information in a shorter amount of time, especially in large item spaces. This may also be due to the fact that users spend a lot of time on streaming platforms and therefore the industry is more sensitive to the topic.

The panelists then answered questions from the audience. First regarding the issue of assessing causality when including impressions in model training and evaluation as well as accounting for business metrics. The panelists agreed that their impact is significant. Furthermore, the source of impressions too will have a strong impact. For example, if there is a home feed and an autoplay feed the two will generate different types of impressions with different information content. Similarly, on mobile devices the space is much smaller which may require to treat differently impressions shown on mobile compared to those shown on TV or personal computers. The panelists argued that a smaller space with a lower number of impressions would increase their importance by making more likely for the user to actually see them, reducing uncertainty and noise. The discussion then focused on the issue of balancing the data, given that each interaction will be associated to many impressions, as well as the cost/benefit trade-offs of logging and storing that large amount of data.

Acknowledgments

We would like to thank all the workshop attendees, our keynote speaker Jiangwei Pan, our additional panelists Arnab Bhadury and Srijan Saket, as well as all authors of accepted papers and presenting authors. We also thank the RecSys Workshop Chairs Ludovico Boratto, Mi Zhang and Victor Sheng. Finally, we would like to thank our program committee: Antonio Ferrara, Claudio Pomo, Daniele Malitesta, David Massimo, Fernando B. Pérez Maurera, Marco de Gemmis, Marco Polignano, Nicolò Felicioni, Olivier Jeunen, Pengjie Ren, Vito Walter Anelli.

References

Giuseppe Di Benedetto, Ben London, Alexander Buchholz, Yannik Stein, Vito Bellini, Matej Jakimov, Matteo Ruffini, and Thorsten Joachims. Contextual position bias estimation using a single stochastic logging policy. In Maurizio Ferrari Dacrema, Pablo Castells, Justin Basilico, Paolo Cremonesi, and Fernando B. Pérez Maurera, editors, *Proceedings of the Workshop on Learning and Evaluating Recommendations with Impressions co-located with the 17th ACM Conference on Recommender Systems (RecSys 2023), Singapore, September 19th, 2023*, volume 3590 of *CEUR Workshop Proceedings*, pages 55–61. CEUR-WS.org, 2023. URL https://ceur-ws.org/Vol-3590/short4.pdf.

Simen Eide, David S. Leslie, and Arnoldo Frigessi. Dynamic slate recommendation with gated recurrent units and thompson sampling. *CoRR*, abs/2104.15046, 2021. URL https://arxiv.org/abs/2104.15046.

Marcel Hauck, Ahtsham Manzoor, and Sven Pagel. Effects of human-curated content on diversity in PSM: ARD-M dataset. In Maurizio Ferrari Dacrema, Pablo Castells, Justin Basilico, Paolo Cremonesi, and Fernando B. Pérez Maurera, editors, *Proceedings of the Workshop on Learning and Evaluating Recommendations with Impressions co-located with the 17th ACM Conference on Recommender Systems (RecSys 2023), Singapore, September 19th, 2023*, volume 3590 of *CEUR Workshop Proceedings*, pages 34–40. CEUR-WS.org, 2023. URL https://ceur-ws.org/Vol-3590/short1.pdf.

Dong Li, Divya Bhargavi, and Vidya Sagar Ravipati. Impression-informed multi-behavior recommender system: A hierarchical graph attention approach. In Maurizio Ferrari Dacrema, Pablo Castells, Justin Basilico, Paolo Cremonesi, and Fernando B. Pérez Maurera, editors, *Proceedings of the Workshop on Learning and Evaluating Recommendations with Impressions co-located with the 17th ACM Conference on Recommender Systems (RecSys 2023), Singapore, September 19th, 2023*, volume 3590 of *CEUR Workshop Proceedings*, pages 9–21. CEUR-WS.org, 2023. URL https://ceur-ws.org/Vol-3590/paper1.pdf.

Dawen Liang, Laurent Charlin, James McInerney, and David M. Blei. Modeling user exposure in recommendation. In Jacqueline Bourdeau, Jim Hendler, Roger Nkambou, Ian Horrocks, and Ben Y. Zhao, editors, *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11 - 15, 2016*, pages 951–961. ACM, 2016. doi: 10.1145/2872427.2883090. URL https://doi.org/10.1145/2872427.2883090.

Xiao Ma, Liqin Zhao, Guan Huang, Zhi Wang, Zelin Hu, Xiaoqiang Zhu, and Kun Gai. Entire space multi-task model: An effective approach for estimating post-click conversion rate. In Kevyn Collins-Thompson, Qiaozhu Mei, Brian D. Davison, Yiqun Liu, and Emine Yilmaz, editors, *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018*, pages 1137–1140. ACM, 2018. doi: 10.1145/3209978.3210104. URL https://doi.org/10.1145/3209978.3210104.

Manchit Madan. Offline evaluation using interactions to decide cross-selling recommendations algorithm for online food delivery. In Maurizio Ferrari Dacrema, Pablo Castells, Justin Basilico, Paolo Cremonesi, and Fernando B. Pérez Maurera, editors, *Proceedings of the Workshop on Learning and Evaluating Recommendations with Impressions co-located with the 17th ACM Conference on Recommender Systems (RecSys 2023), Singapore, September 19th, 2023*, volume 3590 of *CEUR Workshop Proceedings*, pages 49–54. CEUR-WS.org, 2023. URL https://ceur-ws.org/Vol-3590/short3.pdf.

Fernando B. Pérez Maurera, Maurizio Ferrari Dacrema, Lorenzo Saule, Mario Scriminaci, and Paolo Cremonesi. Contentwise impressions: An industrial dataset with impressions included. In Mathieu d'Aquin, Stefan Dietze, Claudia Hauff, Edward Curry, and Philippe Cudré-Mauroux, editors, *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, pages 3093–3100. ACM, 2020. doi: 10.1145/3340531.3412774. URL https://doi.org/10.1145/3340531.3412774.

Fernando B. Pérez Maurera, Maurizio Ferrari Dacrema, and Paolo Cremonesi. Towards the evaluation of recommender systems with impressions. In Jennifer Golbeck, F. Maxwell Harper, Vanessa Murdock, Michael D. Ekstrand, Bracha Shapira, Justin Basilico, Keld T. Lundgaard, and Even Oldridge, editors, *RecSys '22: Sixteenth ACM Conference on Recommender Systems, Seattle, WA, USA, September 18 - 23, 2022*, pages 610–615. ACM, 2022. doi: 10.1145/3523227.3551483. URL https://doi.org/10.1145/3523227.3551483.

Fernando B. Pérez Maurera, Maurizio Ferrari Dacrema, Pablo Castells, and Paolo Cremonesi. Characterizing impression-aware recommender systems. In Maurizio Ferrari Dacrema, Pablo Castells, Justin Basilico, Paolo Cremonesi, and Fernando B. Pérez Maurera, editors, *Proceedings of the Workshop on Learning and Evaluating Recommendations with Impressions co-located with the 17th ACM Conference on Recommender Systems (RecSys 2023), Singapore, September 19th, 2023*, volume 3590 of *CEUR Workshop Proceedings*, pages 22–33. CEUR-WS.org, 2023a. URL https://ceur-ws.org/Vol-3590/paper2.pdf.

Fernando B. Pérez Maurera, Maurizio Ferrari Dacrema, Pablo Castells, and Paolo Cremonesi. Incorporating impressions to graph-based recommenders. In Maurizio Ferrari Dacrema, Pablo Castells, Justin Basilico, Paolo Cremonesi, and Fernando B. Pérez Maurera, editors, *Proceedings of the Workshop on Learning and Evaluating Recommendations with Impressions co-located with the 17th ACM Conference on Recommender Systems (RecSys 2023), Singapore, September 19th, 2023*, volume 3590 of *CEUR Workshop Proceedings*, pages 62–67. CEUR-WS.org, 2023b. URL https://ceur-ws.org/Vol-3590/short5.pdf.

Srijan Saket, Venkata Sai Baba Reddy Velugoti, and Rishabh Mehrotra. Formulating video watch success signals for recommendations on short video platforms. In Maurizio Ferrari Dacrema, Pablo Castells, Justin Basilico, Paolo Cremonesi, and Fernando B. Pérez Maurera, editors, *Proceedings of the Workshop on Learning and Evaluating Recommendations with Impressions co-located with the 17th ACM Conference on Recommender Systems (RecSys 2023), Singapore, September 19th, 2023*, volume 3590 of *CEUR Workshop Proceedings*, pages 41–48. CEUR-WS.org, 2023. URL https://ceur-ws.org/Vol-3590/short2.pdf.

Masahiro Sato, Janmajay Singh, Sho Takemori, Takashi Sonoda, Qian Zhang, and Tomoko Ohkuma. Uplift-based evaluation and optimization of recommenders. In Toine Bogers, Alan Said, Peter Brusilovsky, and Domonkos Tikk, editors, *Proceedings of the 13th ACM Conference on Recommender Systems, RecSys 2019, Copenhagen, Denmark, September 16-20, 2019*, pages 296–304. ACM, 2019. doi: 10.1145/3298689.3347018. URL https://doi.org/10.1145/3298689.3347018.

Qijie Shen, Hong Wen, Wanjie Tao, Jing Zhang, Fuyu Lv, Zulong Chen, and Zhao Li. Deep interest highlight network for click-through rate prediction in trigger-induced recommendation. In Frédérique Laforest, Raphaël Troncy, Elena Simperl, Deepak Agarwal, Aristides Gionis, Ivan Herman, and Lionel Médini, editors, *WWW '22: The ACM Web Conference 2022, Virtual Event, Lyon, France, April 25 - 29, 2022*, pages 422–430. ACM, 2022. doi: 10.1145/3485447.3511970. URL https://doi.org/10.1145/3485447.3511970.

Ting-Yi Shih, Ting-Chang Hou, Jian-De Jiang, Yen-Chieh Lien, Chia-Rui Lin, and Pu-Jen Cheng. Dynamically integrating item exposure with rating prediction in collaborative filtering. In Raffaele Perego, Fabrizio Sebastiani, Javed A. Aslam, Ian Ruthven, and Justin Zobel, editors, *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy, July 17-21, 2016*, pages 813–816. ACM, 2016. doi: 10.1145/2911451.2914769. URL https://doi.org/10.1145/2911451.2914769.

Sumit Sidana, Charlotte Laclau, and Massih-Reza Amini. Learning to recommend diverse items over implicit feedback on PANDOR. In Sole Pera, Michael D. Ekstrand, Xavier Amatriain, and John O'Donovan, editors, *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018, Vancouver, BC, Canada, October 2-7, 2018*, pages 427–431. ACM, 2018. doi: 10.1145/3240323.3240400. URL https://doi.org/10.1145/3240323.3240400.

Jiayin Wang, Weizhi Ma, Jiayu Li, Hongyu Lu, Min Zhang, Biao Li, Yiqun Liu, Peng Jiang, and Shaoping Ma. Make fairness more fair: Fair item utility estimation and exposure re-distribution. In Aidong Zhang and Huzefa Rangwala, editors, *KDD '22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 14 - 18, 2022*, pages 1868–1877. ACM, 2022. doi: 10.1145/3534678.3539354. URL https://doi.org/10.1145/3534678.3539354.

Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu, and Ming Zhou. MIND: A large-scale dataset for news recommendation. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 3597–3606. Association for Computational Linguistics, 2020. doi: 10.18653/v1/2020.acl-main.331. URL https://doi.org/10.18653/v1/2020.acl-main.331.

Xin Xin, Jiyuan Yang, Hanbing Wang, Jun Ma, Pengjie Ren, Hengliang Luo, Xinlei Shi, Zhumin Chen, and Zhaochun Ren. On the user behavior leakage from recommender system exposure. *ACM Trans. Inf. Syst.*, 41(3):57:1–57:25, 2023. doi: 10.1145/3568954. URL https://doi.org/10.1145/3568954.

Chenxu Zhu, Peng Du, Weinan Zhang, Yong Yu, and Yang Cao. Combo-fashion: Fashion clothes matching CTR prediction with item history. In Aidong Zhang and Huzefa Rangwala, editors, *KDD '22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 14 - 18, 2022*, pages 4621–4629. ACM, 2022. doi: 10.1145/3534678.3539101. URL https://doi.org/10.1145/3534678.3539101.