Impressions in Recommender Systems: Present and Future

Discussion Paper

Fernando B. Pérez Maurera[1,2,*], Maurizio Ferrari Dacrema[1,3], Pablo Castells[4] and Paolo Cremonesi[1]

[1]Politecnico di Milano, Milano, Italy
[2]ContentWise, Milano, Italy
[3]ICSC, Bologna, Italy
[4]Universidad Autónoma de Madrid, Madrid, Spain

Abstract

Impressions are a novel data source providing researchers and practitioners with more details about user interactions and their context. In particular, an impression contain the items shown on screen to users, alongside users' interactions toward such items. In recent years, interest in impressions has thrived, and more papers use impressions in recommender systems. Despite this, the literature does not contain a comprehensive review of the current topics and future directions. This work summarizes impressions in recommender systems under three perspectives: recommendation models, datasets with impressions, and evaluation methodologies. Then, we propose several future directions with an emphasis on novel approaches. This work is part of an ongoing review of impressions in recommender systems.

Keywords

Recommender Systems, Impressions, Exposure, Past Recommendations, Slates

1. Introduction

Recommender systems aim to generate user engagement in the short and long term. They achieve this by creating *personalized recommendations*, i.e., a curation of the catalog tailored to users based on their preferences. When such recommendations are relevant, they generate the desired engagement of users toward the recommender system. However, this is not an easy task or goal. A recommender system must be able to predict future user preferences in various conditions, e.g., when new users arrive, when users change their tastes, or when handling anonymous users.

The research community has devised recommenders using distinct data sources to learn users' preferences. Interactions are one of the most relevant and commonly used data sources;

IIR2023: 13th Italian Information Retrieval Workshop, 8th - 9th June 2023, Pisa, Italy

*Corresponding author.
fernandobenjamin.perez@polimi.it (F. B. Pérez Maurera); maurizio.ferrari@polimi.it (M. Ferrari Dacrema); pablo.castells@uam.es (P. Castells); paolo.cremonesi@polimi.it (P. Cremonesi)
0000-0001-6578-7404 (F. B. Pérez Maurera); 0000-0001-7103-2788 (M. Ferrari Dacrema); 0000-0003-0668-6317 (P. Castells); 0000-0002-1253-8081 (P. Cremonesi)

most research works in recommender systems use interactions. Interactions are those actions users perform toward items of a recommender system, e.g., product purchases or movie ratings.

Using additional data sources is an effective approach to improve the recommendation quality. This work focuses on one specific data source: *impressions*, also known as *slates*, *past recommendations*, *exposure*. An impression contain information about items shown on-screen to users, alongside the possible interactions with such items. In some cases, an impression also contains layout information, i.e., the collection of items shown on-screen, the position of items, their placement on-screen, and labels indicating which items were interacted with. Impressions are not exclusive to recommender systems; they can be generated by different entities, e.g., a search engine or editors. This work, however, focuses on impressions in recommender systems.

With impressions, researchers are able to dissect whether a given item has been shown to a user and whether the user preferred it. Impressions contain mixed *signals* of users' preferences, e.g., for a given user and impression, the user may like all, some, or none of the items in the impression. This juxtaposition enables researchers to investigate novel areas or complex attributes of recommender systems using impressions. However, using impressions in recommender systems does come with additional challenges, e.g., the number of impressions is usually higher than the number of interactions, in some cases, by several orders of magnitude.

This work is part of an in-progress systematic literature review on impressions in recommender systems. This work summarizes our review's relevant findings, highlighting the current state of the art, open research questions, and future directions. Throughout this work, we structure the discussion under three perspectives: recommendation models, datasets with impressions, and evaluation methodologies.

2. State of the Art

We start this section by providing the definition of the most common and relevant terms used in this work. Then, we proceed with the discussion of the state of the art. We structure the discussion around three fundamental angles in recommender systems: recommendation models, datasets with impressions, and evaluation methodologies.

The reviewed literature consists of regular conference or journal papers published in high-level venues describing recommenders using impressions. We discovered papers using relevant academic search engines and a query to match papers containing the keywords *impression* (or its synonyms) and *recommender systems* in their content. Most papers were discarded due to the broad definition of impression, e.g., some papers use the word *impression* as how users perceive the recommender system.

2.1. Definitions

We define an *impression* as a selection of N items to be served to the user by a recommender system or another entity (e.g., a search engine). Upon arrangement of an impression on-screen, users decide whether to *interact* with items in the impression or not, i.e., a user's action over an item. Traditionally, the community handles user feedback using user-item pairs, i.e., for a given user and item, the pair indicates whether the user interacted with the item. We refine the definition of user feedback based on users' interactions with impressions: *interacted impression*, *non-interacted impression*, and *non-impressed*. An interacted impression is a user-item pair where the user interacts with the item. A non-interacted impression is a user-item pair where the user does not interact with a served item. Lastly, a non-impressed is a user-item pair where the item has never been served to the user.

Depending on the recommender's system data collection strategy, impressions can be categorized into two classes: *contextual* and *global* impressions. A contextual impression holds the items shown on-screen and the interactions such items receive. On the contrary, a global impression holds only one of the two: the items shown on-screen or the interactions one item receives.

2.2. Recommendation Models

This section describes recommendation models: the module in a recommender system in charge of learning users' preferences and predicting the relevance of items. Recommendation models can be classified according to the design of the recommender.[1] From the literature, we identify five classes of recommenders:

Heuristics Use ad-hoc functions, techniques, or rules to learn users' preferences. Recommenders of this type were published between 2014 and 2017. See [1, 2, 3] for examples.

Statistical Use probabilistic techniques or statistical models to learn users' preferences. One paper was published in 2009, two in 2016, and one in 2017. See [4, 5] for examples.

Machine learning Use machine learning techniques to learn users' preferences. Machine learning and statistical recommenders are the least common recommender type in the literature. See [3, 6] for examples.

Deep learning Use deep learning architectures to learn users' preferences. This is the most popular recommender type in the literature. The most common deep learning architecture is the two-tower framework [7]. See [8, 9, 10] for examples.

Reinforcement learning Model the recommendation task using reinforcement learning, i.e., as a Markov decision process [11]. This is the second most popular recommender type in the literature. See [12, 13, 14] for examples.

2.3. Datasets with Impressions

This section describes datasets with impressions from the reviewed literature. We identify three types of datasets: *public*, *expired*, and *private* datasets. *Public* datasets are those researchers and practitioners can access via the Internet and use in future works as long as the license agreements are respected. *Expired* datasets have been used in competitions (e.g., the ACM RecSys Challenge) and are not accessible anymore. *Private* datasets have never been published nor made available to the community. The downside of expired and private datasets is they cannot be used in future research.

[1] Due to space limitations, we leave out reviewed papers' descriptions and more taxonomies looking at other recommenders' properties.

Inside public datasets, we categorize datasets by the type of impressions they contain: *contextual* or *global* impressions. Three public datasets contain contextual impressions: ContentWise Impressions [15], MIND [16], and FINN.no Slates [17, 18]. Ten public datasets contain global impressions: Yahoo! - R6A [19, 20], Yahoo! - R6B [21, 13], Search Ads[2], PANDOR [22], Ali-CCP [23], Alimama [24], Cross-Shop Combo [25], In-Shop Combo [25], Kwai_FAIR System [26], and Kwai_FAIR Experiment [26].

2.4. Evaluation Methodologies

The last topic addresses the current research goals and the importance of using sound evaluation methodologies to measure progress in the community. Most research papers use offline evaluations. However, many papers by industrial actors are also performing online evaluations via A/B testing. One paper [27] performs user studies.

The most common research goal in the literature is *to improve recommendations quality*, where researchers devise one or several recommendation models to improve a particular evaluation metric, e.g., precision or diversity. Authors may need to alter existing evaluation methodologies or create ad-hoc ones when using impressions. Nonetheless, they must be cautious to avoid data leakages, thus, invalidating their findings. Despite the suggestions made by several previous works [28, 29, 30], we find some papers use improper evaluation methodologies, e.g., creating artificial sessions to evaluate session-based recommenders [10] or not reporting statistical significance in evaluations [31].

3. Future Directions

This section describes existing open research questions or research needs in the same three perspectives of this work: recommendation models, datasets with impressions, and evaluation methodologies. Then, we identify future work directions addressing such questions or needs.

On recommendation models, we identify the lack of recommenders handling *side information* in the literature, i.e., those recommenders designed to leverage interactions and other data sources, e.g., factorization machines [32, 33]. In principle, this type of recommender is suitable for using impressions as side information.

On datasets with impressions, we identify the need for more datasets containing contextual impressions. Contextual impressions are more informative than global impressions: researchers may know all items shown at a particular moment and which are interacted and non-interacted impressions.

On evaluations, an open question *how to use impressions in the evaluation of recommenders?* Felicioni [34] states one future direction is to use impressions to debias evaluation methodologies, e.g., by computing propensity scores. Applying inverse propensity weighting [35] produces an unbiased estimator by adjusting the relevance of items by their propensity score, i.e., the probability of a given user being exposed to a given item. Using impressions to model propensity may be beneficial, as impressions contain the system's exposed and non-exposed items.

[2]https://www.kaggle.com/competitions/kddcup2012-track2

References

[1] N. Buchbinder, M. Feldman, A. Ghosh, J. Naor, Frequency capping in online advertising, J. Sched. 17 (2014) 385–398. URL: https://doi.org/10.1007/s10951-014-0367-z. doi:10.1007/s10951-014-0367-z.

[2] P. Lee, L. V. S. Lakshmanan, M. Tiwari, S. Shah, Modeling impression discounting in large-scale recommender systems, in: S. A. Macskassy, C. Perlich, J. Leskovec, W. Wang, R. Ghani (Eds.), The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA - August 24 - 27, 2014, ACM, 2014, pp. 1837–1846. URL: https://doi.org/10.1145/2623330.2623356. doi:10.1145/2623330.2623356.

[3] D. C. Liu, S. K. Rogers, R. Shiau, D. Kislyuk, K. C. Ma, Z. Zhong, J. Liu, Y. Jing, Related pins at pinterest: The evolution of a real-world recommender system, in: R. Barrett, R. Cummings, E. Agichtein, E. Gabrilovich (Eds.), Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia, April 3-7, 2017, ACM, 2017, pp. 583–592. URL: https://doi.org/10.1145/3041021.3054202. doi:10.1145/3041021.3054202.

[4] C. Wu, C. V. Alvino, A. J. Smola, J. Basilico, Using navigation to improve recommendations in real-time, in: S. Sen, W. Geyer, J. Freyne, P. Castells (Eds.), Proceedings of the 10th ACM Conference on Recommender Systems, Boston, MA, USA, September 15-19, 2016, ACM, 2016, pp. 341–348. URL: https://doi.org/10.1145/2959100.2959174. doi:10.1145/2959100.2959174.

[5] D. Agarwal, B. Chen, P. Elango, Spatio-temporal models for estimating click-through rate, in: Proceedings of the 18th International Conference on World Wide Web, WWW 2009, Madrid, Spain, April 20-24, 2009, ACM, 2009, pp. 21–30. URL: https://doi.org/10.1145/1526709.1526713. doi:10.1145/1526709.1526713.

[6] H. Ma, X. Liu, Z. Shen, User fatigue in online news recommendation, in: J. Bourdeau, J. Hendler, R. Nkambou, I. Horrocks, B. Y. Zhao (Eds.), Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11 - 15, 2016, ACM, 2016, pp. 1363–1372. URL: https://doi.org/10.1145/2872427.2874813. doi:10.1145/2872427.2874813.

[7] X. Yi, J. Yang, L. Hong, D. Z. Cheng, L. Heldt, A. Kumthekar, Z. Zhao, L. Wei, E. H. Chi, Sampling-bias-corrected neural modeling for large corpus item recommendations, in: T. Bogers, A. Said, P. Brusilovsky, D. Tikk (Eds.), Proceedings of the 13th ACM Conference on Recommender Systems, RecSys 2019, Copenhagen, Denmark, September 16-20, 2019, ACM, 2019, pp. 269–277. URL: https://doi.org/10.1145/3298689.3346996. doi:10.1145/3298689.3346996.

[8] R. Xie, S. Zhang, R. Wang, F. Xia, L. Lin, A peep into the future: Adversarial future encoding in recommendation, in: K. S. Candan, H. Liu, L. Akoglu, X. L. Dong, J. Tang (Eds.), WSDM '22: The Fifteenth ACM International Conference on Web Search and Data Mining, Virtual Event / Tempe, AZ, USA, February 21 - 25, 2022, ACM, 2022, pp. 1177–1185. URL: https://doi.org/10.1145/3488560.3498476. doi:10.1145/3488560.3498476.

[9] P. Li, R. Chen, Q. Liu, J. Xu, B. Zheng, Transform cold-start users into warm via fused behaviors in large-scale recommendation, in: E. Amigó, P. Castells, J. Gonzalo, B. Carterette, J. S. Culpepper, G. Kazai (Eds.), SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022,

ACM, 2022, pp. 2013–2017. URL: https://doi.org/10.1145/3477495.3531797. doi:10.1145/3477495.3531797.

[10] S. Gong, K. Q. Zhu, Positive, negative and neutral: Modeling implicit feedback in session-based news recommendation, in: E. Amigó, P. Castells, J. Gonzalo, B. Carterette, J. S. Culpepper, G. Kazai (Eds.), SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022, ACM, 2022, pp. 1185–1195. URL: https://doi.org/10.1145/3477495.3532040. doi:10.1145/3477495.3532040.

[11] M. M. Afsar, T. Crump, B. H. Far, Reinforcement learning based recommender systems: A survey, ACM Comput. Surv. 55 (2023) 145:1–145:38. URL: https://doi.org/10.1145/3543846. doi:10.1145/3543846.

[12] J. McInerney, B. Lacker, S. Hansen, K. Higley, H. Bouchard, A. Gruson, R. Mehrotra, Explore, exploit, and explain: personalizing explainable recommendations with bandits, in: S. Pera, M. D. Ekstrand, X. Amatriain, J. O'Donovan (Eds.), Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018, Vancouver, BC, Canada, October 2-7, 2018, ACM, 2018, pp. 31–39. URL: https://doi.org/10.1145/3240323.3240354. doi:10.1145/3240323.3240354.

[13] S. Li, A. Karatzoglou, C. Gentile, Collaborative filtering bandits, in: R. Perego, F. Sebastiani, J. A. Aslam, I. Ruthven, J. Zobel (Eds.), Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy, July 17-21, 2016, ACM, 2016, pp. 539–548. URL: https://doi.org/10.1145/2911451.2911548. doi:10.1145/2911451.2911548.

[14] Y. Ge, X. Zhao, L. Yu, S. Paul, D. Hu, C. Hsieh, Y. Zhang, Toward pareto efficient fairness-utility trade-off in recommendation through reinforcement learning, in: K. S. Candan, H. Liu, L. Akoglu, X. L. Dong, J. Tang (Eds.), WSDM '22: The Fifteenth ACM International Conference on Web Search and Data Mining, Virtual Event / Tempe, AZ, USA, February 21 - 25, 2022, ACM, 2022, pp. 316–324. URL: https://doi.org/10.1145/3488560.3498487. doi:10.1145/3488560.3498487.

[15] F. B. Pérez Maurera, M. Ferrari Dacrema, L. Saule, M. Scriminaci, P. Cremonesi, Contentwise impressions: An industrial dataset with impressions included, in: M. d'Aquin, S. Dietze, C. Hauff, E. Curry, P. Cudré-Mauroux (Eds.), CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020, ACM, 2020, pp. 3093–3100. URL: https://doi.org/10.1145/3340531.3412774. doi:10.1145/3340531.3412774.

[16] F. Wu, Y. Qiao, J. Chen, C. Wu, T. Qi, J. Lian, D. Liu, X. Xie, J. Gao, W. Wu, M. Zhou, MIND: A large-scale dataset for news recommendation, in: D. Jurafsky, J. Chai, N. Schluter, J. R. Tetreault (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020, Association for Computational Linguistics, 2020, pp. 3597–3606. URL: https://doi.org/10.18653/v1/2020.acl-main.331. doi:10.18653/v1/2020.acl-main.331.

[17] S. Eide, D. S. Leslie, A. Frigessi, J. Rishaug, H. Jenssen, S. Verrewaere, Finn.no slates dataset: A new sequential dataset logging interactions, all viewed items and click responses/no-click for recommender systems research, in: H. J. C. Pampín, M. A. Larson, M. C. Willemsen, J. A. Konstan, J. J. McAuley, J. Garcia-Gathright, B. Huurnink, E. Oldridge (Eds.), RecSys

'21: Fifteenth ACM Conference on Recommender Systems, Amsterdam, The Netherlands, 27 September 2021 - 1 October 2021, ACM, 2021, pp. 556–558. URL: https://doi.org/10.1145/3460231.3474607. doi:10.1145/3460231.3474607.

[18] S. Eide, D. S. Leslie, A. Frigessi, Dynamic slate recommendation with gated recurrent units and thompson sampling, Data Min. Knowl. Discov. 36 (2022) 1756–1786. URL: https://doi.org/10.1007/s10618-022-00849-w. doi:10.1007/s10618-022-00849-w.

[19] L. Li, W. Chu, J. Langford, X. Wang, Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms, in: I. King, W. Nejdl, H. Li (Eds.), Proceedings of the Forth International Conference on Web Search and Web Data Mining, WSDM 2011, Hong Kong, China, February 9-12, 2011, ACM, 2011, pp. 297–306. URL: https://doi.org/10.1145/1935826.1935878. doi:10.1145/1935826.1935878.

[20] W. Chu, S. Park, T. Beaupre, N. Motgi, A. Phadke, S. Chakraborty, J. Zachariah, A case study of behavior-driven conjoint analysis on yahoo!: front page today module, in: J. F. E. IV, F. Fogelman-Soulié, P. A. Flach, M. J. Zaki (Eds.), Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, June 28 - July 1, 2009, ACM, 2009, pp. 1097–1104. URL: https://doi.org/10.1145/1557019.1557138. doi:10.1145/1557019.1557138.

[21] C. Gentile, S. Li, G. Zappella, Online clustering of bandits, in: Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014, volume 32 of *JMLR Workshop and Conference Proceedings*, JMLR.org, 2014, pp. 757–765. URL: http://proceedings.mlr.press/v32/gentile14.html.

[22] S. Sidana, C. Laclau, M. Amini, Learning to recommend diverse items over implicit feedback on PANDOR, in: S. Pera, M. D. Ekstrand, X. Amatriain, J. O'Donovan (Eds.), Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018, Vancouver, BC, Canada, October 2-7, 2018, ACM, 2018, pp. 427–431. URL: https://doi.org/10.1145/3240323.3240400. doi:10.1145/3240323.3240400.

[23] X. Ma, L. Zhao, G. Huang, Z. Wang, Z. Hu, X. Zhu, K. Gai, Entire space multi-task model: An effective approach for estimating post-click conversion rate, in: K. Collins-Thompson, Q. Mei, B. D. Davison, Y. Liu, E. Yilmaz (Eds.), The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018, ACM, 2018, pp. 1137–1140. URL: https://doi.org/10.1145/3209978.3210104. doi:10.1145/3209978.3210104.

[24] Q. Shen, H. Wen, W. Tao, J. Zhang, F. Lv, Z. Chen, Z. Li, Deep interest highlight network for click-through rate prediction in trigger-induced recommendation, in: F. Laforest, R. Troncy, E. Simperl, D. Agarwal, A. Gionis, I. Herman, L. Médini (Eds.), WWW '22: The ACM Web Conference 2022, Virtual Event, Lyon, France, April 25 - 29, 2022, ACM, 2022, pp. 422–430. URL: https://doi.org/10.1145/3485447.3511970. doi:10.1145/3485447.3511970.

[25] C. Zhu, P. Du, W. Zhang, Y. Yu, Y. Cao, Combo-fashion: Fashion clothes matching CTR prediction with item history, in: A. Zhang, H. Rangwala (Eds.), KDD '22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 14 - 18, 2022, ACM, 2022, pp. 4621–4629. URL: https://doi.org/10.1145/3534678.3539101. doi:10.1145/3534678.3539101.

[26] J. Wang, W. Ma, J. Li, H. Lu, M. Zhang, B. Li, Y. Liu, P. Jiang, S. Ma, Make fairness more fair: Fair item utility estimation and exposure re-distribution, in: A. Zhang, H. Rangwala

(Eds.), KDD '22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 14 - 18, 2022, ACM, 2022, pp. 1868–1877. URL: https://doi.org/10.1145/3534678.3539354. doi:10.1145/3534678.3539354.

[27] Q. Zhao, M. C. Willemsen, G. Adomavicius, F. M. Harper, J. A. Konstan, Interpreting user inaction in recommender systems, in: S. Pera, M. D. Ekstrand, X. Amatriain, J. O'Donovan (Eds.), Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018, Vancouver, BC, Canada, October 2-7, 2018, ACM, 2018, pp. 40–48. URL: https://doi.org/10.1145/3240323.3240366. doi:10.1145/3240323.3240366.

[28] M. Ferrari Dacrema, S. Boglio, P. Cremonesi, D. Jannach, A troubling analysis of reproducibility and progress in recommender systems research, ACM Trans. Inf. Syst. 39 (2021) 20:1–20:49. URL: https://doi.org/10.1145/3434185. doi:10.1145/3434185.

[29] P. Castells, A. Moffat, Offline recommender system evaluation: Challenges and new directions, AI Mag. 43 (2022) 225–238. URL: https://doi.org/10.1002/aaai.12051. doi:10.1002/aaai.12051.

[30] J. Beel, C. Breitinger, S. Langer, A. Lommatzsch, B. Gipp, Towards reproducibility in recommender-systems research, User Model. User Adapt. Interact. 26 (2016) 69–101. URL: https://doi.org/10.1007/s11257-016-9174-x. doi:10.1007/s11257-016-9174-x.

[31] Z. Chen, J. Wu, C. Li, J. Chen, R. Xiao, B. Zhao, Co-training disentangled domain adaptation network for leveraging popularity bias in recommenders, in: E. Amigó, P. Castells, J. Gonzalo, B. Carterette, J. S. Culpepper, G. Kazai (Eds.), SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022, ACM, 2022, pp. 60–69. URL: https://doi.org/10.1145/3477495.3531952. doi:10.1145/3477495.3531952.

[32] S. Rendle, Factorization machines, in: G. I. Webb, B. Liu, C. Zhang, D. Gunopulos, X. Wu (Eds.), ICDM 2010, The 10th IEEE International Conference on Data Mining, Sydney, Australia, 14-17 December 2010, IEEE Computer Society, 2010, pp. 995–1000. URL: https://doi.org/10.1109/ICDM.2010.127. doi:10.1109/ICDM.2010.127.

[33] H. Guo, R. Tang, Y. Ye, Z. Li, X. He, Deepfm: A factorization-machine based neural network for CTR prediction, in: C. Sierra (Ed.), Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017, ijcai.org, 2017, pp. 1725–1731. URL: https://doi.org/10.24963/ijcai.2017/239. doi:10.24963/ijcai.2017/239.

[34] N. Felicioni, Enhancing counterfactual evaluation and learning for recommendation systems, in: J. Golbeck, F. M. Harper, V. Murdock, M. D. Ekstrand, B. Shapira, J. Basilico, K. T. Lundgaard, E. Oldridge (Eds.), RecSys '22: Sixteenth ACM Conference on Recommender Systems, Seattle, WA, USA, September 18 - 23, 2022, ACM, 2022, pp. 739–741. URL: https://doi.org/10.1145/3523227.3547429. doi:10.1145/3523227.3547429.

[35] T. Schnabel, A. Swaminathan, A. Singh, N. Chandak, T. Joachims, Recommendations as treatments: Debiasing learning and evaluation, in: M. Balcan, K. Q. Weinberger (Eds.), Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016, volume 48 of *JMLR Workshop and Conference Proceedings*, JMLR.org, 2016, pp. 1670–1679. URL: http://proceedings.mlr.press/v48/schnabel16.html.