Enhancing Counterfactual Evaluation and Learning for Recommendation Systems

Nicolò Felicioni
nicolo.felicioni@polimi.it
Politecnico di Milano
Milan, Italy

ABSTRACT

Evaluating recommendation systems is a task of utmost importance and a very active research field. While online evaluation is the most reliable evaluation procedure, it may also be too expensive to perform, if not unfeasible. Therefore, researchers and practitioners resort to offline evaluation. Offline evaluation is much more efficient and scalable, but traditional approaches suffer from high bias. This issue led to the increased popularity of counterfactual techniques. These techniques are used for evaluation and learning in recommender systems and reduce the bias in offline evaluation. While counterfactual approaches have a solid statistical basis, their application to recommendation systems is still in a preliminary research phase. In this paper, we identify some limitations of counterfactual techniques applied to recommender systems, and we propose possible ways to overcome them.

ACM Reference Format:
Nicolò Felicioni. 2022. Enhancing Counterfactual Evaluation and Learning for Recommendation Systems. In *Sixteenth ACM Conference on Recommender Systems (RecSys '22), September 18–23, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3523227.3547429

1 INTRODUCTION

Evaluating recommender systems is a task of the utmost importance in the research field [11]. It is well-known that the golden standard of recommender systems' evaluation is an online A/B test [8]. However, online evaluation is highly expensive and not scalable [7, 22]. For these reasons, finding offline evaluation procedures that simulate the outcomes of an online A/B test reliably is a very active research field among both researchers and practitioners. Logged implicit feedback is very cheap to get. The evaluation with this kind of data, though, carries a critical issue: the data collected has a bias due to the fact that the deployed recommender influenced users [7, 10]. This issue contributed to the increase in popularity of a new methodology for offline evaluation called *Counterfactual Evaluation* [3]. The fundamental idea is to view a recommendation system as a *contextual bandit* [12]: the user is a *context* observed by the decision-maker (the recommender). The *actions* that the decision-maker may take are the items in the catalog. The recommended item is sampled from a probability distribution conditioned on the observed user (which we call *policy*), and finally, a *reward* is observed. With this interpretation, we can account for the bias of the recommender that collected the data in order to get an unbiased offline evaluation of another recommendation algorithm. The objective of the counterfactual evaluation is to find an estimator of the expected reward of the evaluation policy that carries statistical guarantees on the reliability of the estimation (e.g., unbiasedness, consistency). Once having obtained a well-grounded estimator, we can also use it for *Counterfactual Learning*, namely, to learn a policy based on the data logged by the logging policy by maximizing the estimated expected reward. Despite the promising qualities, both counterfactual evaluation and counterfactual learning have drawbacks. In the following, we investigate some major research questions in the field of counterfactual evaluation and learning for recommendation systems, and we propose some possible future research directions.

RecSys '22, September 18–23, 2022, Seattle, WA, USA

ACM ISBN 978-1-4503-9278-5/22/09.
https://doi.org/10.1145/3523227.3547429

2 RESEARCH QUESTIONS

Exploiting Side Information. Many state-of-the-art counterfactual estimators exploit an assumption called *full support*, i.e., the logging recommender explored sufficiently the catalog during the logging phase. Whenever this assumption is not valid, we say that we have *deficient support* [18, 23], and usual estimators are biased in the general case. One possible mitigation could be imputing rewards for the missing context-action pairs with a regression model [15]. The drawback of this approach is that model misspecification can lead to a high bias, and it is difficult to have an accurate reward model for each user-item pair [4, 15, 20]. Hence, a possible future research direction is to focus on estimators without a regression model for deficient support. For instance, we could take inspiration from the various *content-based* methods in the recommender systems literature [1, 2, 14, 25] to create an estimator that considers the *similarity* among the actions computed exploiting side information. This could lead to a precise estimation even in the presence of deficient support.

Tractable Counterfactual Slate Recommendation. Real-world recommender systems often aim to suggest a ranked list of items to a user. This problem is called *slate* (or *top-n*) recommendation [6, 9]. Framing this problem as a contextual bandit problem is challenging. Naively, we could interpret every possible permutation of the items as an action and then apply standard counterfactual estimators, but the action space becomes combinatorially large and intractable. One way to circumvent this issue is to make reasonable assumptions about the user interaction behavior with the slate (e.g., we can assume that the reward for each item is influenced only by the preceding items in the recommendation list). All the state-of-the-art proposals make assumptions about user behavior. A direction for future research could be to relax those assumptions with a milder assumption. For example, we could assume that the reward obtained from a slate depends only on a more compact embedding of the slate. A very recent work that goes in a similar direction is [19]. The authors assume that the reward of a given action depends only on a more compact embedding of the action. While very promising, their estimator requires complete knowledge of the logging policy, which is usually impossible to obtain. Thus, they resort to estimating the logging policy, incurring bias. A possible future direction could be to build an estimator upon their assumptions that do not require complete knowledge of the logging policy.

Counterfactual Interpretation of Impressions. While a counterfactual approach is statistically sound and valuable, we need a particular type of dataset to use this approach. Specifically, for an unbiased counterfactual evaluation, we would need a dataset with logged propensities, or a dataset where the ground truth is completely known. Unfortunately, many widely used datasets in the recommendation systems literature do not have the propensity available and have missing data for which we have no ground truth. A popular approach for making counterfactual evaluation and learning with such recommender datasets is a so-called *semi-synthetic* approach [21]: a recommender algorithm is trained on the given dataset and produces a set of scores for each user-item pair. These scores are assumed to be the *ground truth* relevance, and in this way, we have no more missing values. This mechanism is simple but has some drawbacks, such as the fact that the selection of the oracle recommender is somehow arbitrary. Recently, many *impression datasets* were released [5, 16, 24], and their usage has seen a rapid increase in popularity [13, 17]. An impression dataset is halfway between an ideal counterfactual dataset and a classic recommender dataset. A possible research direction is to take advantage of this new type of information for estimating the propensity of the logging recommender.

ACKNOWLEDGMENTS

I would like to thank my advisor Prof. Paolo Cremonesi and Dr. Maurizio Ferrari Dacrema for their support.

REFERENCES

[1] Chumki Basu, Haym Hirsh, and William W. Cohen. 1998. Recommendation as Classification: Using Social and Content-Based Information in Recommendation. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence and Tenth Innovative Applications of Artificial Intelligence Conference, AAAI 98, IAAI 98, July 26-30, 1998, Madison, Wisconsin, USA*, Jack Mostow and Chuck Rich (Eds.). AAAI Press / The MIT Press, 714–720. http://www.aaai.org/Library/AAAI/1998/aaai98-101.php

[2] Cesare Bernardis and Paolo Cremonesi. 2021. NFC: a deep and hybrid item-based model for item cold-start recommendation. *User Modeling and User-Adapted Interaction* (2021), 1–34. https://doi.org/10.1007/s11257-021-09303-w

[3] Léon Bottou, Jonas Peters, Joaquin Quiñonero Candela, Denis Xavier Charles, Max Chickering, Elon Portugaly, Dipankar Ray, Patrice Y. Simard, and Ed Snelson. 2013. Counterfactual reasoning and learning systems: the example of computational advertising. *J. Mach. Learn. Res.* 14, 1 (2013), 3207–3260. http://dl.acm.org/citation.cfm?id=2567766

[4] Miroslav Dudík, John Langford, and Lihong Li. 2011. Doubly Robust Policy Evaluation and Learning. In *Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011*, Lise Getoor and Tobias Scheffer (Eds.). Omnipress, 1097–1104. https://icml.cc/2011/papers/554_icmlpaper.pdf

[5] Simen Eide, David S. Leslie, Arnoldo Frigessi, Joakim Rishaug, Helge Jenssen, and Sofie Verrewaere. 2021. FINN.no Slates Dataset: A new Sequential Dataset Logging Interactions, all Viewed Items and Click Responses/No-Click for Recommender Systems Research. In *RecSys '21: Fifteenth ACM Conference on Recommender Systems, Amsterdam, The Netherlands, 27 September 2021 - 1 October 2021*, Humberto Jesús Corona Pampín, Martha A. Larson, Martijn C. Willemsen, Joseph A. Konstan, Julian J. McAuley, Jean Garcia-Gathright, Bouke Huurnink, and Even Oldridge (Eds.). ACM, 556–558. https://doi.org/10.1145/3460231.3474607

[6] Nicolò Felicioni, Maurizio Ferrari Dacrema, and Paolo Cremonesi. 2021. A Methodology for the Offline Evaluation of Recommender Systems in a User Interface with Multiple Carousels. In *Adjunct Publication of the 29th ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2021, Utrecht, The Netherlands, June 21-25, 2021*, Judith Masthoff, Eelco Herder, Nava Tintarev, and Marko Tkalcic (Eds.). ACM, 10–15. https://doi.org/10.1145/3450614.3461680

[7] Alexandre Gilotte, Clément Calauzènes, Thomas Nedelec, Alexandre Abraham, and Simon Dollé. 2018. Offline A/B Testing for Recommender Systems. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM 2018, Marina Del Rey, CA, USA, February 5-9, 2018*, Yi Chang, Chengxiang Zhai, Yan Liu, and Yoelle Maarek (Eds.). ACM, 198–206. https://doi.org/10.1145/3159652.3159687

[8] Alois Gruson, Praveen Chandar, Christophe Charbuillet, James McInerney, Samantha Hansen, Damien Tardieu, and Ben Carterette. 2019. Offline Evaluation to Make Decisions About PlaylistRecommendation Algorithms. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM 2019, Melbourne, VIC, Australia, February 11-15, 2019*, J. Shane Culpepper, Alistair Moffat, Paul N. Bennett, and Kristina Lerman (Eds.). ACM, 420–428. https://doi.org/10.1145/3289600.3291027

[9] Eugene Ie, Vihan Jain, Jing Wang, Sanmit Narvekar, Ritesh Agarwal, Rui Wu, Heng-Tze Cheng, Tushar Chandra, and Craig Boutilier. 2019. SlateQ: A Tractable Decomposition for Reinforcement Learning with Recommendation Sets. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, Sarit Kraus (Ed.). ijcai.org, 2592–2599. https://doi.org/10.24963/ijcai.2019/360

[10] Amir Hossein Jadidinejad, Craig Macdonald, and Iadh Ounis. 2022. The Simpson's Paradox in the Offline Evaluation of Recommendation Systems. *ACM Trans. Inf. Syst.* 40, 1 (2022), 4:1–4:22. https://doi.org/10.1145/3458509

[11] Olivier Jeunen. 2019. Revisiting offline evaluation for implicit-feedback recommender systems. In *Proceedings of the 13th ACM Conference on Recommender Systems, RecSys 2019, Copenhagen, Denmark, September 16-20, 2019*, Toine Bogers, Alan Said, Peter Brusilovsky, and Domonkos Tikk (Eds.). ACM, 596–600. https://doi.org/10.1145/3298689.3347069

[12] John Langford and Tong Zhang. 2007. The Epoch-Greedy Algorithm for Multi-armed Bandits with Side Information. In *Advances in Neural Information Processing Systems 20, Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 3-6, 2007*, John C. Platt, Daphne Koller, Yoram Singer, and Sam T. Roweis (Eds.). Curran Associates, Inc., 817–824. https://proceedings.neurips.cc/paper/2007/hash/4b04a686b0ad13dce35fa99fa4161c65-Abstract.html

[13] Pei Lee, Laks V. S. Lakshmanan, Mitul Tiwari, and Sam Shah. 2014. Modeling impression discounting in large-scale recommender systems. In *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA - August 24 - 27, 2014*, Sofus A. Macskassy, Claudia Perlich, Jure Leskovec, Wei Wang, and Rayid Ghani (Eds.). ACM, 1837–1846. https://doi.org/10.1145/2623330.2623356

[14] Tianqiao Liu, Zhiwei Wang, Jiliang Tang, Songfan Yang, Gale Yan Huang, and Zitao Liu. 2019. Recommender Systems with Heterogeneous Side Information. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, Ling Liu, Ryen W. White, Amin Mantrach, Fabrizio Silvestri, Julian J. McAuley, Ricardo Baeza-Yates, and Leila Zia (Eds.). ACM, 3027–3033. https://doi.org/10.1145/3308558.3313580

[15] Ben London and Thorsten Joachims. 2020. Offline policy evaluation with new arms. In *Offline Reinforcement Learning Workshop at Neural Information Processing Systems.*

[16] Fernando Benjamín Pérez Maurera, Maurizio Ferrari Dacrema, Lorenzo Saule, Mario Scriminaci, and Paolo Cremonesi. 2020. ContentWise Impressions: An Industrial Dataset with Impressions Included. In *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, Mathieu d'Aquin, Stefan Dietze, Claudia Hauff, Edward Curry, and Philippe Cudré-Mauroux (Eds.). ACM, 3093–3100. https://doi.org/10.1145/3340531.3412774

[17] James McInerney, Ehtsham Elahi, Justin Basilico, Yves Raimond, and Tony Jebara. 2021. Accordion: A Trainable Simulator forLong-Term Interactive Systems. In *RecSys '21: Fifteenth ACM Conference on Recommender Systems, Amsterdam, The Netherlands, 27 September 2021 - 1 October 2021*, Humberto Jesús Corona Pampín, Martha A. Larson, Martijn C. Willemsen, Joseph A. Konstan, Julian J. McAuley, Jean Garcia-Gathright, Bouke Huurnink, and Even Oldridge (Eds.). ACM, 102–113. https://doi.org/10.1145/3460231.3474259

[18] Noveen Sachdeva, Yi Su, and Thorsten Joachims. 2020. Off-policy Bandits with Deficient Support. In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge*

Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020, Rajesh Gupta, Yan Liu, Jiliang Tang, and B. Aditya Prakash (Eds.). ACM, 965–975. https://doi.org/10.1145/3394486.3403139

[19] Yuta Saito and Thorsten Joachims. 2022. Off-Policy Evaluation for Large Action Spaces via Embeddings. In *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA (Proceedings of Machine Learning Research, Vol. 162)*, Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato (Eds.). PMLR, 19089–19122. https://proceedings.mlr.press/v162/saito22a.html

[20] Yuta Saito, Takuma Udagawa, Haruka Kiyohara, Kazuki Mogi, Yusuke Narita, and Kei Tateno. 2021. Evaluating the Robustness of Off-Policy Evaluation. In *RecSys '21: Fifteenth ACM Conference on Recommender Systems, Amsterdam, The Netherlands, 27 September 2021 - 1 October 2021*, Humberto Jesús Corona Pampín, Martha A. Larson, Martijn C. Willemsen, Joseph A. Konstan, Julian J. McAuley, Jean Garcia-Gathright, Bouke Huurnink, and Even Oldridge (Eds.). ACM, 114–123. https://doi.org/10.1145/3460231.3474245

[21] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. 2016. Recommendations as Treatments: Debiasing Learning and Evaluation. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016 (JMLR Workshop and Conference Proceedings, Vol. 48)*, Maria-Florina Balcan and Kilian Q. Weinberger (Eds.). JMLR.org, 1670–1679. http://proceedings.mlr.press/v48/schnabel16.html

[22] Guy Shani and Asela Gunawardana. 2011. Evaluating Recommendation Systems. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor (Eds.). Springer, 257–297. https://doi.org/10.1007/978-0-387-85820-3_8

[23] Hung Tran-The, Sunil Gupta, Thanh Nguyen-Tang, Santu Rana, and Svetha Venkatesh. 2021. Combining Online Learning and Offline Learning for Contextual Bandits with Deficient Support. *CoRR* abs/2107.11533 (2021). arXiv:2107.11533 https://arxiv.org/abs/2107.11533

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

[25] Feipeng Zhao, Min Xiao, and Yuhong Guo. 2016. Predictive Collaborative Filtering with Side Information. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*, Subbarao Kambhampati (Ed.). IJCAI/AAAI Press, 2385–2391. http://www.ijcai.org/Abstract/16/340