CONSEQUENCES — Causality, Counterfactuals and Sequential Decision-Making for Recommender Systems

Olivier Jeunen
Amazon
Edinburgh, United Kingdom
jeunen@amazon.co.uk

Thorsten Joachims
Cornell University
Ithaca, NY, USA
tj@cs.cornell.edu

Harrie Oosterhuis
Radboud University
Nijmegen, The Netherlands
harrie.oosterhuis@ru.nl

Yuta Saito
Cornell University
Ithaca, NY, USA
ys552@cornell.edu

Flavian Vasile*
Criteo
Paris, France
f.vasile@criteo.com

ABSTRACT

Recommender systems are more and more often modelled as repeated decision making processes – deciding which (ranking of) items to recommend to a given user. Each decision to recommend or rank an item has a significant impact on immediate and future user responses, long-term satisfaction or engagement with the system, and possibly valuable exposure for the item provider. This interactive and *interventionist* view of the recommender uncovers a plethora of unanswered research questions, as it complicates the typically adopted offline evaluation or learning procedures in the field. We need an understanding of *causal inference* to reason about (possibly unintended) consequences of the recommender, and a notion of *counterfactuals* to answer common "what if"-type questions in learning and evaluation. Advances at the intersection of these fields can foster progress in effective, efficient and fair learning and evaluation from logged data. These topics have been emerging in the Recommender Systems community for a while, but we firmly believe in the value of a dedicated forum and place to learn and exchange ideas. We welcome contributions from both academia and industry and bring together a growing community of researchers and practitioners interested in sequential decision making, offline evaluation, batch policy learning, fairness in online platforms, as well as other related tasks, such as A/B testing.

CCS CONCEPTS

• **Computing methodologies → Batch learning**; *Learning from implicit feedback*; **Learning to rank**; **Ranking**.

KEYWORDS

counterfactuals, off-policy evaluation and learning, recommender systems, fairness in rankings

*Authors listed in alphabetical order.

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

ACM ISBN 978-1-4503-9278-5/22/09.
https://doi.org/10.1145/3523227.3547409

ACM Reference Format:
Olivier Jeunen, Thorsten Joachims, Harrie Oosterhuis, Yuta Saito, and Flavian Vasile. 2022. CONSEQUENCES — Causality, Counterfactuals and Sequential Decision-Making for Recommender Systems. In *Sixteenth ACM Conference on Recommender Systems (RecSys '22), September 18–23, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3523227.3547409

1 INTRODUCTION AND MOTIVATION

Current-day, real-world recommender systems can be framed as sequential decision making processes: a policy observes a context, takes an action, and possibly obtains a reward for the chosen action [1, 11, 13, 26, 27, 34–37]. This modern problem setting has evolved from the earlier "rating prediction" days and the Netflix Prize [2], up to a point where the entire framing has undergone a paradigm shift. Indeed, talking about *decisions* instead of *predictions* clearly illustrates that actions taken by the system have *consequences*. If we wish to fully understand the impact of the decisions made by the systems we are designing, it is crucial that we are able to reason about these consequences in all their aspects. This is where the science of "Causal Inference" comes into play – as it allows us to ask questions that are *counterfactual* in nature: "*What would have happened if I had recommended this other item instead?*"

Off-Policy Evaluation (OPE) is arguably the first area in recommender systems research that has fully embraced this paradigm: OPE methods take in logs of previous versions of the system, detailing which recommendations where shown along with the outcomes they led to, and then generate some counterfactual estimate for a metric [36]. For example, "*What would the click-through-rate have been if system A had been deployed instead of system B?*"

Off-Policy Learning (OPL) methods take this process one step further still, and try to directly optimise a system that will maximise the answer to this question. Because of their practical relevance, several estimators and policy learning methods with impressive theoretical results have been proposed in recent years [7, 8, 14, 15, 34, 35, 43]. Nevertheless, the theoretical results often rely on assumptions that do not hold in practice. It is seldom straightforward to apply these advanced methods to real-world applications due to a number of challenges, such as combinatorial/continuous actions [15, 17, 19, 25, 38], safety requirements [9, 40, 41], implementation costs [39], large action sets [16, 18, 19, 28], hyperparameter tuning [26, 29, 42], and distributional shifts [30]. Understanding and addressing the gap between theory and practice has been an issue for the community. Nevertheless, these are important problems, as the decisions of how we rank items that are presented to users affect immediate and long-term outcomes [5].

Additionally, modern recommendation algorithms are used to power multi-sided platforms all over the world wide web, from music streaming platforms to restaurant recommenders and job matching platforms. The decisions made by these systems influence the lives of many people throughout, and it is crucial to be aware of this [31–33]. The potential impact that recommendations have on people and society has been widely established as an essential research topic, and several attempts have been made to define and enforce fairness in online platforms. One way to model this is by enforcing *fairness of exposure*, mandating ranking algorithms to allocate exposure proportional to merit (e.g. probability of relevance) [3, 6, 12, 22, 31, 32]. Nevertheless, currently existing approaches do not always take into account their long-term consequences in large-scale ecosystems, as there is simply no obvious way to do so [20].

It is clear that the decisions made by our recommendation algorithms can have downstream effects on (1) data used to train and evaluate future models, (2) short- as well as long-term user satisfaction, (3) item providers, people, and society at large. Advances in causal and counterfactual inference can help us better understand and reason about these effects. In this context, we believe it is timely to organize a workshop that exposes the community towards the problem of designing and evaluating recommender systems while considering their counterfactual nature and long-term consequences, and to provide an engaging forum for discussion.

We expect the following benefits for attendees to the workshop:

(1) to follow and learn about recent advances and trends in the counterfactual and causal recommendation, and identify important open problems;
(2) to understand the empirical behaviour of online learning, counterfactual reasoning, and fairness in rankings and discuss their importance, applicability, and robustness;
(3) to bridge the gap between academia and industry through developing new methods, building new datasets, and sharing experiences and challenges in applications;
(4) to encourage the future collaboration of students, researchers, and practitioners in their research and development of recommender systems.

2 WORKSHOP FORMAT

The CONSEQUENCES workshop is a full-day event, co-organised with the REVEAL workshop on Reinforcement Learning at Scale. The programme includes:

(1) An in-workshop tutorial by the organisers, covering advanced methods in the area,
(2) Invited talks by leading researchers,
(3) Oral presentations of the best accepted papers over the different tracks topics,
(4) Poster presentations and discussions of all accepted contributions.

2.1 In-Workshop Tutorial

Topics related to causality and counterfactual inference find their roots in statistics, and can often seem rather complex and contrived to new researchers in the field. Advanced topics are typically not covered by conference tutorials, due to time limitations and to attract broad audiences. The goal of this (optional) part of the workshop is to give an advanced introduction and overview of recent advances in the field. The in-workshop tutorial will be given by the organisers at the beginning of the workshop to make sure that every attendee shares important background knowledge, and is aware of recent advances in the research area. The in-workshop tutorial targets the workshop attendees and experts, so it covers recent topics and up-to-date methodologies in more detail than the main conference tutorials. The tutorial will also have plenty of time for involved discussion and Q&A sessions.

2.2 Invited Talks

We are honoured to have two confirmed invited speakers: one academic and one industrial researcher, who are both highly regarded in their fields. We allot 45 minutes for their talks and 15 minutes for Q&A, to allow plenty of room for in-depth discussion.

Guido Imbens is the Applied Econometrics Professor and Professor of Economics at the Stanford Graduate School of Business. After graduating from Brown University Guido taught at Harvard University, UCLA, and UC Berkeley. He joined the GSB in 2012. Imbens specializes in econometrics, and in particular methods for drawing causal inferences. Guido Imbens is a fellow of the Econometric Society and the American Academy of Arts and Sciences. In 2021, he was co-awarded The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel, for "their methodological contributions to the analysis of causal relationships".

Lihong Li is a Senior Principal Scientist at Amazon. He obtained a PhD degree in Computer Science from Rutgers University. After that, he has held research positions in Yahoo!, Microsoft and Google. His main research interests are in reinforcement learning, including contextual bandits, and other related problems in AI. His work has found applications in recommendation, advertising, web search and conversation systems, and has won best paper awards at ICML, AISTATS and WSDM. He regularly serves as area chair or senior program committee member at major AI/ML conferences such as AAAI, AISTATS, ICLR, ICML, IJCAI and NeurIPS.

2.3 Contributed Talks & Poster Session

The contributed talks will be given by selected high-quality contributions of different types of papers over the different tracks, covering methodological research, empirics, software, and open problems. We will allocate about 30 minutes for each contributed talk so that the authors can have an opportunity to give a detailed presentation. We will also have a specific segment of the schedule reserved for discussing open problems and industry experiences to encourage future collaboration amongst the attendees. The morning and afternoon breaks will include poster sessions. The workshop has a dedicated Twitter channel (@CONSEQUENCES_ws) and webpage (sites.google.com/view/consequences2022).

3 ORGANISER BIOGRAPHIES

The workshop organisers and their biographies are listed here in alphabetical order:

Olivier Jeunen (jeunen@amazon.co.uk) is a Postdoctoral Scientist at Amazon with a PhD in Computer Science from the University of Antwerp [10]. The final chapter of his thesis was recognised with the ACM RecSys '21 Best Student Paper award [11]. Olivier's research focuses on applying ideas from causal and counterfactual inference to recommendation and advertising problems. Before joining Amazon, he gained industrial experience through collaborations with Criteo, Facebook and Spotify Research; co-chaired the Dutch-Belgian Information Retrieval workshop in 2020, and co-lectured tutorials at the RecSys Summer School, UMAP and The WebConf.

Thorsten Joachims (tj@cs.cornell.edu) is a Professor in the Department of Computer Science and in the Department of Information Science at Cornell University, and he is an Amazon Scholar. His research interests center on the synthesis of theory and system building in machine learning, with applications in information retrieval and recommendation. His past research focused on support vector machines, learning to rank, learning with preferences, and learning from implicit feedback, text classification, and structured output prediction. Working with his students and collaborators, his papers won 10 Best Paper Awards and 4 Test-of-Time Awards. He is also an ACM Fellow, AAAI Fellow, KDD Innovations Award recipient, and member of the SIGIR Academy.

Harrie Oosterhuis (harrie.oosterhuis@ru.nl) is an Assistant Professor in the Institute for Computing and Information Sciences at the Radboud University Nijmegen, and he is also a Staff Machine Learning Scientist at Twitter. His research lies on the intersection of machine learning and information retrieval and primarily concerns learning from user interactions on rankings for search and recommendation. He received his PhD cum laude from the University of Amsterdam in 2020 on the thesis titled "*Learning from User Interactions with Rankings: A Unification of the Field*" [21]. He is also a recipient of the 2021 Google Research Scholar Award for early career researchers and the WSDM'21 [23], SIGIR'21 [22] and ICTIR'22 [24] best paper awards.

Yuta Saito (ys552@cornell.edu) is a Ph.D. student in the Department of Computer Science at Cornell University, advised by Prof. Thorsten Joachims. He received a B.Eng degree in Industrial Engineering and Economics from Tokyo Institute of Technology in 2021. His current research focuses on OPE of bandit algorithms, learning from human behavior data, and fairness in rankings. Some of his recent work has been published at top conferences, including ICML, NeurIPS, KDD, SIGIR, RecSys, and WSDM. He has also co-lectured a tutorial related to counterfactual inference at RecSys 2021 and won the Best Paper Runner-Up Award at WSDM2022 [16].

Flavian Vasile (f.vasile@criteo.com) is part of the Criteo AI Lab where he works as Principal Scientist, his main focus being on the development of Deep Learning-based Recommendation Systems and on introducing aspects of Causal Inference to Recommendation. Before joining Criteo, he worked as a Senior Researcher in the Twitter Advertising Science team; before that, in the Yahoo! Research Lab where he mostly focused on Content Understanding problems. His current research interests include Deep Sequential Models for Recommendation and understanding Recommendation as a decision-making system with reward uncertainty. Among his recent research publications, the work on Causal Embeddings for Recommendation received the best paper award at RecSys '18 [4] and he co-organised the '18–'20 REVEAL Workshop series in conjunction with ACM RecSys.

ACKNOWLEDGMENTS

This research was supported in part by NSF Awards IIS-1901168 and IIS-2008139. Yuta Saito is supported by Funai Overseas Scholarship. Harrie Oosterhuis is supported by the Google Research Scholar program. All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.

REFERENCES

[1] Aman Agarwal, Soumya Basu, Tobias Schnabel, and Thorsten Joachims. 2017. Effective Evaluation using Logged Bandit Feedback from Multiple Loggers. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, 687–696.

[2] James Bennett, Stan Lanning, et al. 2007. The Netflix prize. In *Proc. of the KDD cup and workshop*, Vol. 2007. 35.

[3] Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. 2018. Equity of attention: Amortizing individual fairness in rankings. In *The 41st international acm sigir conference on research & development in information retrieval*. 405–414.

[4] Stephen Bonner and Flavian Vasile. 2018. Causal embeddings for recommendation. In *Proceedings of the 12th ACM conference on recommender systems*. 104–112.

[5] Minmin Chen, Yuyan Wang, Can Xu, Ya Le, Mohit Sharma, Lee Richardson, Su-Lin Wu, and Ed Chi. 2021. Values of User Exploration in Recommender Systems. In *Fifteenth ACM Conference on Recommender Systems*. 85–95.

[6] Fernando Diaz, Bhaskar Mitra, Michael D Ekstrand, Asia J Biega, and Ben Carterette. 2020. Evaluating stochastic rankings with expected exposure. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 275–284.

[7] Miroslav Dudík, Dumitru Erhan, John Langford, and Lihong Li. 2014. Doubly Robust Policy Evaluation and Optimization. *Statist. Sci.* 29, 4 (2014), 485–511.

[8] Mehrdad Farajtabar, Yinlam Chow, and Mohammad Ghavamzadeh. 2018. More Robust Doubly Robust Off-policy Evaluation. In *Proceedings of the 35th International Conference on Machine Learning*, Vol. 80. PMLR, 1447–1456.

[9] Audrey Huang, Liu Leqi, Zachary Lipton, and Kamyar Azizzadenesheli. 2021. Off-policy risk assessment in contextual bandits. *Advances in Neural Information Processing Systems* 34 (2021).

[10] Olivier Jeunen. 2021. *Offline Approaches to Recommendation with Online Success*. Ph.D. Dissertation. University of Antwerp.

[11] Olivier Jeunen and Bart Goethals. 2021. Pessimistic reward models for off-policy learning in recommendation. In *Fifteenth ACM Conference on Recommender Systems*. 63–74.

[12] Olivier Jeunen and Bart Goethals. 2021. Top-k contextual bandits with equity of exposure. In *Fifteenth ACM Conference on Recommender Systems*. 310–320.

[13] Olivier Jeunen, David Rohde, Flavian Vasile, and Martin Bompaire. 2020. Joint Policy-Value Learning for Recommendation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1223–1233.

[14] Nathan Kallus, Yuta Saito, and Masatoshi Uehara. 2021. Optimal Off-Policy Evaluation from Multiple Logging Policies. In *Proceedings of the 38th International Conference on Machine Learning*, Vol. 139. PMLR, 5247–5256.

[15] Nathan Kallus and Angela Zhou. 2018. Policy Evaluation and Optimization with Continuous Treatments. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*. PMLR, 1243–1251.
[16] Haruka Kiyohara, Yuta Saito, Tatsuya Matsuhiro, Yusuke Narita, Nobuyuki Shimizu, and Yasuo Yamamoto. 2022. Doubly Robust Off-Policy Evaluation for Ranking Policies under the Cascade Behavior Model. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. 487–497.
[17] Shuai Li, Yasin Abbasi-Yadkori, Branislav Kveton, S Muthukrishnan, Vishwa Vinay, and Zheng Wen. 2018. Offline Evaluation of Ranking Policies with Click Models. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Association for Computing Machinery, 1685–1694.
[18] Jiaqi Ma, Zhe Zhao, Xinyang Yi, Ji Yang, Minmin Chen, Jiaxi Tang, Lichan Hong, and Ed H Chi. 2020. Off-policy learning in two-stage recommender systems. In *Proceedings of The Web Conference 2020*. 463–473.
[19] James McInerney, Brian Brost, Praveen Chandar, Rishabh Mehrotra, and Benjamin Carterette. 2020. Counterfactual Evaluation of Slate Recommendations with Sequential Reward Interactions. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Association for Computing Machinery, 1779–1788.
[20] Martin Mladenov, Elliot Creager, Omer Ben-Porat, Kevin Swersky, Richard Zemel, and Craig Boutilier. 2020. Optimizing long-term social welfare in recommender systems: A constrained matching approach. In *International Conference on Machine Learning*. PMLR, 6987–6998.
[21] Harrie Oosterhuis. 2020. *Learning From User Interactions With Rankings: A Unification of the Field*. Ph.D. Dissertation. University of Amsterdam.
[22] Harrie Oosterhuis. 2021. Computationally Efficient Optimization of Plackett-Luce Ranking Models for Relevance and Fairness. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 1023–1032.
[23] Harrie Oosterhuis and Maarten de Rijke. 2021. Unifying Online and Counterfactual Learning to Rank. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining (WSDM'21)*. ACM.
[24] Harrie Oosterhuis and Maarten de Rijke. 2022. Reaching the End of Unbiasedness: Uncovering Implicit Limitations of Click-Based Learning to Rank. In *Proceedings of the 2022 ACM SIGIR International Conference on the Theory of Information Retrieval*. ACM.
[25] Yuta Saito. 2020. Doubly robust estimator for ranking metrics with post-click conversions. In *Fourteenth ACM Conference on Recommender Systems*. Association for Computing Machinery, 92–100.
[26] Yuta Saito, Shunsuke Aihara, Megumi Matsutani, and Yusuke Narita. 2020. Open Bandit Dataset and Pipeline: Towards Realistic and Reproducible Off-Policy Evaluation. *arXiv preprint arXiv:2008.07146* (2020).
[27] Yuta Saito and Thorsten Joachims. 2021. Counterfactual learning and evaluation for recommender systems: Foundations, implementations, and recent advances. In *Fifteenth ACM Conference on Recommender Systems*. 828–830.
[28] Yuta Saito and Thorsten Joachims. 2022. Off-Policy Evaluation for Large Action Spaces via Embeddings. In *International Conference on Machine Learning*. PMLR, 19089–19122.
[29] Yuta Saito, Takuma Udagawa, Haruka Kiyohara, Kazuki Mogi, Yusuke Narita, and Kei Tateno. 2021. Evaluating the Robustness of Off-Policy Evaluation. In *Fifteenth ACM Conference on Recommender Systems*. 114–123.
[30] Nian Si, Fan Zhang, Zhengyuan Zhou, and Jose Blanchet. 2020. Distributionally Robust Policy Evaluation and Learning in Offline Contextual Bandits. In *Proceedings of the 38th International Conference on Machine Learning*, Vol. 119. PMLR, 8884–8894.
[31] Ashudeep Singh and Thorsten Joachims. 2018. Fairness of Exposure in Rankings. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Association for Computing Machinery, 2219–2228.
[32] Ashudeep Singh and Thorsten Joachims. 2019. Policy Learning for Fairness in Ranking. In *Advances in Neural Information Processing Systems*, Vol. 32.
[33] Yi Su, Magd Bayoumi, and Thorsten Joachims. 2021. Optimizing Rankings for Recommendation in Matching Markets. *arXiv preprint arXiv:2106.01941* (2021).
[34] Yi Su, Maria Dimakopoulou, Akshay Krishnamurthy, and Miroslav Dudík. 2020. Doubly Robust Off-Policy Evaluation with Shrinkage. In *Proceedings of the 37th International Conference on Machine Learning*, Vol. 119. PMLR, 9167–9176.
[35] Yi Su, Lequn Wang, Michele Santacatterina, and Thorsten Joachims. 2019. Cab: Continuous Adaptive Blending for Policy Evaluation and Learning. In *Proceedings of the 36th International Conference on Machine Learning*, Vol. 97. PMLR, 6005–6014.
[36] Adith Swaminathan and Thorsten Joachims. 2015. Batch Learning from Logged Bandit Feedback through Counterfactual Risk Minimization. *The Journal of Machine Learning Research* 16, 1 (2015), 1731–1755.
[37] Adith Swaminathan and Thorsten Joachims. 2015. The Self-Normalized Estimator for Counterfactual Learning. In *Advances in Neural Information Processing Systems*, Vol. 28. 3231–3239.
[38] Adith Swaminathan, Akshay Krishnamurthy, Alekh Agarwal, Miro Dudik, John Langford, Damien Jose, and Imed Zitouni. 2017. Off-policy Evaluation for Slate Recommendation. In *Advances in Neural Information Processing Systems*, Vol. 30.
[39] Shengpu Tang and Jenna Wiens. 2021. Model selection for offline reinforcement learning: Practical considerations for healthcare settings. In *Machine Learning for Healthcare Conference*. PMLR, 2–35.
[40] Philip Thomas, Georgios Theocharous, and Mohammad Ghavamzadeh. 2015. High-confidence off-policy evaluation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 29.
[41] Philip Thomas, Georgios Theocharous, and Mohammad Ghavamzadeh. 2015. High confidence policy improvement. In *International Conference on Machine Learning*. PMLR, 2380–2388.
[42] Cameron Voloshin, Hoang M Le, Nan Jiang, and Yisong Yue. 2019. Empirical study of off-policy policy evaluation for reinforcement learning. *arXiv preprint arXiv:1911.06854* (2019).
[43] Yu-Xiang Wang, Alekh Agarwal, and Miroslav Dudık. 2017. Optimal and adaptive off-policy evaluation in contextual bandits. In *Proceedings of the 34th International Conference on Machine Learning*, Vol. 70. PMLR, 3589–3597.