Unbiased Learning to Rank: On Recent Advances and Practical Applications

Shashank Gupta
University of Amsterdam
Amsterdam, The Netherlands
s.gupta2@uva.nl

Philipp Hager
University of Amsterdam
Amsterdam, The Netherlands
p.k.hager@uva.nl

Jin Huang
University of Amsterdam
Amsterdam, The Netherlands
j.huang2@uva.nl

Ali Vardasbi
University of Amsterdam
Amsterdam, The Netherlands
a.vardasbi@uva.nl

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

ABSTRACT

Since its inception, the field of unbiased learning to rank (ULTR) has remained very active and has seen several impactful advancements in recent years. This tutorial provides both an introduction to the core concepts of the field and an overview of recent advancements in its foundations, along with several applications of its methods.

The tutorial is divided into four parts: Firstly, we give an overview of the different forms of bias that can be addressed with ULTR methods. Secondly, we present a comprehensive discussion of the latest estimation techniques in the ULTR field. Thirdly, we survey published results of ULTR in real-world applications. Fourthly, we discuss the connection between ULTR and fairness in ranking. We end by briefly reflecting on the future of ULTR research and its applications.

This tutorial is intended to benefit both researchers and industry practitioners interested in developing new ULTR solutions or utilizing them in real-world applications.

CCS CONCEPTS

• **Information systems → Learning to rank**; *Query log analysis*; Recommender systems.

KEYWORDS

Learning to Rank; Counterfactual Learning to Rank

ACM Reference Format:
Shashank Gupta, Philipp Hager, Jin Huang, Ali Vardasbi, and Harrie Oosterhuis. 2024. Unbiased Learning to Rank: On Recent Advances and Practical Applications. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining (WSDM '24), March 4–8, 2024, Merida, Mexico.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3616855.3636451

1 MOTIVATION

Learning to rank (LTR) algorithms are the cornerstone of modern search and recommender systems. Traditionally, LTR algorithms were based on supervised learning from manually-graded relevance judgments. However, obtaining relevance judgments is costly and often not aligned with actual user preferences [5, 36]. In contrast, click data is cheaper to collect and is generally better aligned with user intents [17]. However, click data is usually a heavily biased signal of user preference [1, 7, 19] which the field of unbiased learning to rank (ULTR) aims to mitigate [24].

Previous tutorials focused on introducing the fundamentals of the field to researchers and practitioners in the information retrieval and recommender system communities [2, 21, 31]. While very relevant at the time, the field of ULTR has matured significantly, and fundamental advancements have been made since then. At the time of the last tutorials, the primarily studied interaction bias was position bias [7, 19, 48]. Since then, the community has addressed new interaction biases, including trust bias [1, 45], item selection bias [28], contextual bias [50, 52, 58, 59], and cascading position bias [20, 44]. For correcting biases, the method most commonly used was inverse propensity scoring (IPS). However, it is now known that IPS is not effective in correcting for all forms of interaction biases [25, 45]. Hence, several new and fundamental estimation techniques have been developed to overcome the limitations of IPS, for instance, affine-corrections [30, 45] and doubly-robust estimation [26], which can both be seen as extensions of IPS for ULTR. Furthermore, estimation methods that are fundamentally different from IPS have also been proposed, such as two-tower models [10, 52, 59] and causal-inference-based methods [32, 43, 57]. While many ULTR methods focus on mitigating bias in historic datasets, the area of online learning to rank mitigates biases while directly interacting with users [27, 39, 54]. A recent line of work addresses both online and offline settings with methods that can be applied to either setting and thereby, aims to unify the ULTR field [29, 30].

Recently, LTR has also seen significant growth from the application side [1, 4, 13, 50, 59], including fair LTR [41, 42, 51]. The focus of the previous tutorials was on the fundamentals of ULTR with a limited emphasis on practical applications. While the focus of LTR was traditionally on relevance ranking, it is now commonly acknowledged that optimizing for relevance alone can result in unfairness issues [3, 40, 53]. In this regard, we believe that the objective of a similar area, such as fair LTR, aligns with ULTR's mission, which is to provide fair and unbiased rankings to the user.

WSDM '24, March 4–8, 2024, Merida, Mexico

ACM ISBN 979-8-4007-0371-3/24/03.
https://doi.org/10.1145/3616855.3636451

To scale up to large-scale applications, fair LTR work relies on unbiased LTR [37, 46, 51], and we hope that our tutorial will encourage further exploration in this area.

Given these significant advancements in the area of ULTR and the increased applications of its methodology, we believe it is the right time to provide an overview of the state-of-the-art of the field. Hereby, we aim to benefit both academic researchers and industry practitioners who are either interested in developing new ULTR solutions or utilizing them in their applications.

2 OBJECTIVES

The tutorial is based on the following two main objectives:

- To motivate and introduce the fundamental concepts of ULTR to academics or practitioners who are new to the topic.
- To provide a comprehensive overview of the important recent developments to the foundations and applications of ULTR, that are useful to both newcomers and experts in the field.

Furthermore, we aim for the following additional objectives:

- Provide the most up-to-date explanation of the mathematical foundations of the ULTR field, covering the different forms of bias that can and cannot be corrected for and the latest estimation techniques. Our tutorial should provide a strong foundation for researchers in the ULTR field for their future work.
- Present an in-depth survey of real-world applications of ULTR so that practitioners can have a realistic expectation of the potential impact of applying ULTR.
- Motivate and stimulate cross-disciplinary research, by enabling researchers from other areas to understand how ULTR could be useful for them. In particular, we will highlight the connection with the topic of *fairness in ranking*.

3 RELEVANCE TO THE INFORMATION RETRIEVAL COMMUNITY

The area of ULTR has grown significantly in the last couple of years, with several fundamental contributions and diverse IR applications.

The earliest tutorial in the area was by Joachims and Swaminathan [18], where they introduced counterfactual learning in the context of search and recommendation. Recent tutorials on counterfactual evaluation and learning have focused mostly on bandit feedback data [34, 35]. In the context of recommender systems, Chen et al. [6] introduce biases and debiasing strategies.

For LTR specifically, there have been tutorials introducing offline ULTR [2, 21, 31] and online LTR [9]. However, to the best of our knowledge, no existing tutorial covers the important advancements in the ULTR field that has been made in the last three years, nor their recent applications, including fair LTR [41, 51].

4 FORMAT AND DETAILED SCHEDULE

The tutorial will consist of three hours, excluding breaks, with the following schedule:

Preliminaries (20 minutes) The first session focuses on the preliminaries; we discuss the basics of supervised LTR and some of the earliest works in position-bias and counterfactual LTR.

- **Learning to rank basics (5 minutes)**: Discuss basics of supervised LTR by introducing pointwise, pairwise, and listwise LTR methods and the concept of learning from user interactions.
- **Position bias (5 minutes)**: Discuss position bias that arises when applying traditional LTR methods on user clicks [7].
- **Counterfactual LTR (10 minutes)**: Introduce the basics of counterfactual LTR for learning from user feedback data with position bias [19].

Biases (40 minutes) In this session, we cover the recent advances in the types of interaction bias that can be tackled with ULTR methods beyond position bias.

- **Trust bias (10 minutes)**: We discuss trust bias, where users are likely to click on items in top positions regardless of item relevance because they trust the search engine [1, 45].
- **Item Selection Bias (10 minutes)**: We discuss item selection bias, where users can only interact with a fixed set of k items, and items outside the top-k position have zero chances of exposure [28].
- **Contextual Bias (10 minutes)**: We discuss contextual bias, where the item's click probability is affected by its surrounding items on the display page [50, 52, 58, 59].
- **Cascading Position Bias (10 minutes)**: Under the cascade model [7], the position bias of an item depends not only on the rank an item is displayed at (as many works in ULTR assume [7, 19, 48]), but also on the relevance of the items the user has inspected before [20, 44]. Thus, cascading position bias is often a more realistic assumption of user behavior, e.g., when users tend to stop searching after finding the first relevant result [7].

Novel Estimation Methods (70 minutes) The most prevalent estimation technique for correcting bias from user interaction data is IPS, introduced in the seminal works by Wang et al. [47] and Joachims et al. [19]. However, recently there have been fundamental contributions in the ULTR field with respect to novel estimation techniques. In this session, we will discuss the recent estimation techniques in detail, as per the following schedule:

- **Affine Correction Method (14 minutes)**: We discuss the affine correction method introduced by Vardasbi et al. [45] as an extension to IPS, as IPS is ineffective for trust-bias correction.
- **Doubly Robust Estimation (14 minutes)**: Despite its popularity in the offline bandit learning literature [8, 14, 16, 33, 49], doubly robust estimation methods for position bias correction in LTR were only proposed recently [26]. We discuss this fundamental contribution to the area which overcomes some of the theoretical and practical disadvantages of IPS [26].
- **Online & Counterfactual methods (14 minutes)**: Online learning methods are an alternative class of methods to counter biases in LTR [27, 39, 54], where the user preferences are learned in an online/interactive fashion, as opposed to purely from offline data. Recently, Oosterhuis and de Rijke [30] argued that the field of ULTR can benefit from using online learning to rank via a novel online intervention-aware counterfactual estimator. Online learning has also been used to collect additional data from the logging policy to minimize the variance of the counterfactual estimate of a new ranking policy [29].
- **Safe Counterfactual Policy Learning (14 minutes)**: ULTR relies on exposure-based IPS, which can provide unbiased and

consistent estimates but often suffers from high variance. Especially when little click data is available, this variance can cause ULTR to learn sub-optimal ranking models, which can subsequently bring significant risks of a negative user experience. Recently, Gupta et al. [12] introduced a risk-aware ULTR method with a novel exposure-based concept of risk regularization with strong theoretical guarantees for safe deployment. The method averts the risk of learning a new policy that is worse than the current production system.
- **Two-tower Models (14 minutes):** The focus of ULTR was primarily on identifying and developing novel bias correction methods, with limited focus on model design. Recently, with the introduction of two-tower networks, this trend has been slowly changing [10, 52, 55, 59]. We discuss the two-tower family of correction methods, their capacity to utilize bias-related signals beyond position (e.g., device type), and their ability to correct click data containing a mixture of different user behaviors [52].

Survey Applications (20 minutes) A significant number of works apply ULTR methods in real-world settings. We discuss the different settings explored by existing work, the practical changes they require, and the reported performance impact. Topics include:

- **Grid layouts**: ULTR methods for tackling bias in 2D grid layouts that are common in the industry [50, 59].
- **Multi-feedback**: Integrating multiple types of user feedback beyond clicks, e.g., views, add-to-cart, or purchase actions [11].

From Unbiased to Fair LTR (20 minutes) Traditionally, fair LTR has relied on manual relevance judgments for learning fair ranking policies [3, 40]. Similar to the arguments in favor of click-based learning for relevance rankings, fair LTR needs to adopt click data for its widespread application. In this part of the tutorial, we discuss applications of ULTR for fair policy learning [23, 51].

Conclusion and Future Work (10 minutes) We close by summarizing the main points discussed in the tutorial, in addition, we also discuss some important limitations of the existing overarching counterfactual approach in the ULTR field [25] and some promising avenues for future research.

4.1 Extensions to Previous Versions

Compared to our previous offering, we have extended our proposal to include the latest research on ULTR from this year's conferences including SIGIR, KDD, and RecSys [12, 15, 22, 38, 56]. We also extended individual sections to have more time for Q&A sessions in-between to encourage engagement with the audience. Finally, we have removed a section related to off-policy bandit learning to maintain a coherent theme of ULTR and to facilitate a longer discussion of fairness in LTR.

Lastly, we are aware of the "Practical Bandits: An Industry Perspective" tutorial proposal and are actively collaborating with the authors to ensure our two presentations complement each other. Many techniques in unbiased learning to rank (ULTR) originated in the bandit literature and were adjusted to the intricacies of the ranking setting. Thus, the two tutorials should provide interested attendees with a comprehensive and complementary curriculum.

5 SUPPLIED MATERIAL

Our slides and all supplementary material is publicly available at: https://sites.google.com/view/wsdm-2024-tutorial-ultr/.

ACKNOWLEDGEMENTS

This research was partially supported by the Netherlands Organisation for Scientific Research (NWO) under project nr. VI.Veni.222.269, the Dreams Lab, and the Mercury Machine Learning Lab. The Dreams Lab is a collaboration between Huawei, the University of Amsterdam, and the Vrije Universiteit Amsterdam. The Mercury Machine Learning Lab is a collaboration between TU Delft, the University of Amsterdam, and Booking.com. All content represents the opinion of the authors, which is not necessarily shared nor endorsed by their respective employers and/or sponsors.

REFERENCES

[1] Aman Agarwal, Xuanhui Wang, Cheng Li, Michael Bendersky, and Marc Najork. 2019. Addressing Trust Bias for Unbiased Learning-to-rank. In *The World Wide Web Conference*. 4–14.
[2] Qingyao Ai, Jiaxin Mao, Yiqun Liu, and W Bruce Croft. 2018. Unbiased learning to rank: Theory and practice. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 2305–2306.
[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] Adam Block, Rahul Kidambi, Daniel N Hill, Thorsten Joachims, and Inderjit S Dhillon. 2022. Counterfactual Learning To Rank for Utility-Maximizing Query Autocompletion. *arXiv preprint arXiv:2204.10936* (2022).
[5] Olivier Chapelle and Yi Chang. 2011. Yahoo! Learning to Rank Challenge Overview. In *Proceedings of the learning to rank challenge*. PMLR, 1–24.
[6] Jiawei Chen, Xiang Wang, Fuli Feng, and Xiangnan He. 2021. Bias Issues and Solutions in Recommender System: Tutorial on the RecSys 2021. In *Proceedings of the 15th ACM Conference on Recommender Systems*. 825–827.
[7] Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. 2008. An Experimental Comparison of Click Position-bias Models. In *Proceedings of the 2008 international conference on web search and data mining*. 87–94.
[8] Miroslav Dudík, John Langford, and Lihong Li. 2011. Doubly robust policy evaluation and learning. *arXiv preprint arXiv:1103.4601* (2011).
[9] Artem Grotov and Maarten De Rijke. 2016. Online learning to rank for information retrieval: Sigir 2016 tutorial. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 1215–1218.
[10] Huifeng Guo, Jinkai Yu, Qing Liu, Ruiming Tang, and Yuzhou Zhang. 2019. PAL: a position-bias aware learning framework for CTR prediction in live recommender systems. In *Proceedings of the 13th ACM Conference on Recommender Systems*. 452–456.
[11] Ruocheng Guo, Xiaoting Zhao, Adam Henderson, Liangjie Hong, and Huan Liu. 2020. Debiasing Grid-Based Product Search in E-Commerce. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2852–2860.
[12] Shashank Gupta, Harrie Oosterhuis, and Maarten de Rijke. 2023. Safe Deployment for Counterfactual Learning to Rank with Exposure-Based Risk Minimization. In *SIGIR 2023: 46th international ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM.
[13] Ziniu Hu, Yang Wang, Qu Peng, and Hang Li. 2019. Unbiased lambdamart: an unbiased pairwise learning-to-rank algorithm. In *The World Wide Web Conference*. 2830–2836.
[14] Jiawei Huang and Nan Jiang. 2020. From importance sampling to doubly robust policy gradient. In *International Conference on Machine Learning*. PMLR, 4434–4443.
[15] Olivier Jeunen. 2023. A Probabilistic Position Bias Model for Short-Video Recommendation Feeds. In *Proceedings of the 17th ACM Conference on Recommender Systems* (Singapore, Singapore) *(RecSys '23)*. 675–681.
[16] Nan Jiang and Lihong Li. 2016. Doubly robust off-policy value evaluation for reinforcement learning. In *International Conference on Machine Learning*. PMLR, 652–661.
[17] Thorsten Joachims. 2002. Optimizing Search Engines Using Clickthrough Data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. 133–142.
[18] Thorsten Joachims and Adith Swaminathan. 2016. Counterfactual Evaluation and Learning for Search, Recommendation and Ad Placement. In *Proceedings*

of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval. 1199–1201.

[19] Thorsten Joachims, Adith Swaminathan, and Tobias Schnabel. 2017. Unbiased Learning-to-rank with Biased Feedback. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. 781–789.

[20] 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.

[21] Claudio Lucchese, Franco Maria Nardini, Rama Kumar Pasumarthi, Sebastian Bruch, Michael Bendersky, Xuanhui Wang, Harrie Oosterhuis, Rolf Jagerman, and Maarten de Rijke. 2019. Learning to Rank in Theory and Practice. (2019).

[22] Dan Luo, Lixin Zou, Qingyao Ai, Zhiyu Chen, Dawei Yin, and Brian D. Davison. 2023. Model-Based Unbiased Learning to Rank. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining* (Singapore, Singapore) *(WSDM '23)*. 895–903.

[23] Marco Morik, Ashudeep Singh, Jessica Hong, and Thorsten Joachims. 2020. Controlling fairness and bias in dynamic learning-to-rank. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*. 429–438.

[24] Harrie Oosterhuis. 2020. *Learning from User Interactions with Rankings: A Unification of the Field*. Ph. D. Dissertation. Informatics Institute, University of Amsterdam.

[25] Harrie Oosterhuis. 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*.

[26] Harrie Oosterhuis. 2023. Doubly Robust Estimation for Correcting Position Bias in Click Feedback for Unbiased Learning to Rank. *ACM Transactions on Information Systems* 41, 3 (2023), 1–33.

[27] Harrie Oosterhuis and Maarten de Rijke. 2018. Differentiable unbiased online learning to rank. In *Proceedings of the 27th ACM international conference on information and knowledge management*. 1293–1302.

[28] Harrie Oosterhuis and Maarten de Rijke. 2020. Policy-aware Unbiased Learning to Rank for Top-k Rankings. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 489–498.

[29] Harrie Oosterhuis and Maarten de Rijke. 2020. Taking the Counterfactual Online: Efficient and Unbiased Online Evaluation for Ranking. In *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*. 137–144.

[30] Harrie Oosterhuis and Maarten de Rijke. 2021. Unifying Online and Counterfactual Learning to Rank: A Novel Counterfactual Estimator that Effectively Utilizes Online Interventions. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. 463–471.

[31] Harrie Oosterhuis, Rolf Jagerman, and Maarten de Rijke. 2020. Unbiased Learning to Rank: Counterfactual and Online Approaches. In *Companion Proceedings of the Web Conference 2020*. 299–300.

[32] Zohreh Ovaisi, Kathryn Vasilaky, and Elena Zheleva. 2021. Propensity-Independent Bias Recovery in Offline Learning-to-Rank Systems. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1763–1767.

[33] Yuta Saito. 2020. Doubly robust estimator for ranking metrics with post-click conversions. In *Proceedings of the 14th ACM Conference on Recommender Systems*. 92–100.

[34] 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.

[35] Yuta Saito and Thorsten Joachims. 2022. Counterfactual Evaluation and Learning for Interactive Systems: Foundations, Implementations, and Recent Advances. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4824–4825.

[36] Mark Sanderson, Monica Lestari Paramita, Paul Clough, and Evangelos Kanoulas. 2010. Do User Preferences and Evaluation Measures Line Up?. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. 555–562.

[37] Fatemeh Sarvi, Maria Heuss, Mohammad Aliannejadi, Sebastian Schelter, and Maarten de Rijke. 2021. Understanding and Mitigating the Effect of Outliers in Fair Ranking. *arXiv preprint arXiv:2112.11251* (2021).

[38] Fatemeh Sarvi, Ali Vardasbi, Mohammad Aliannejadi, Sebastian Schelter, and Maarten de Rijke. 2023. On the Impact of Outlier Bias on User Clicks. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Taipei, Taiwan) *(SIGIR '23)*. 18–27.

[39] Anne Schuth, Harrie Oosterhuis, Shimon Whiteson, and Maarten de Rijke. 2016. Multileave gradient descent for fast online learning to rank. In *proceedings of the ninth ACM international conference on web search and data mining*. 457–466.

[40] 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*. 2219–2228.

[41] Ashudeep Singh and Thorsten Joachims. 2019. Policy Learning for Fairness in Ranking. *arXiv preprint arXiv:1902.04056* (2019).

[42] Ashudeep Singh, David Kempe, and Thorsten Joachims. 2021. Fairness in ranking under uncertainty. *Advances in Neural Information Processing Systems* 34 (2021), 11896–11908.

[43] Mucun Tian, Chu Guo, Vito Claudio Ostuni, and Zhen Zhu. 2020. Counterfactual Learning to Rank using Heterogeneous Treatment Effect Estimation. *ArXiv* abs/2007.09798 (2020).

[44] Ali Vardasbi, Maarten de Rijke, and Ilya Markov. 2020. Cascade model-based propensity estimation for counterfactual learning to rank. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2089–2092.

[45] Ali Vardasbi, Harrie Oosterhuis, and Maarten de Rijke. 2020. When Inverse Propensity Scoring does not Work: Affine Corrections for Unbiased Learning to Rank. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 1475–1484.

[46] Ali Vardasbi, Fatemeh Sarvi, and Maarten de Rijke. 2022. Probabilistic Permutation Graph Search: Black-Box Optimization for Fairness in Ranking. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Madrid, Spain) *(SIGIR '22)*. Association for Computing Machinery, New York, NY, USA, 715–725. https://doi.org/10.1145/3477495.3532045

[47] Xuanhui Wang, Michael Bendersky, Donald Metzler, and Marc Najork. 2016. Learning to Rank with Selection Bias in Personal Search. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 115–124.

[48] Xuanhui Wang, Nadav Golbandi, Michael Bendersky, Donald Metzler, and Marc Najork. 2018. Position Bias Estimation for Unbiased Learning to Rank in Personal Search. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. 610–618.

[49] Xiaojie Wang, Rui Zhang, Yu Sun, and Jianzhong Qi. 2019. Doubly robust joint learning for recommendation on data missing not at random. In *International Conference on Machine Learning*. PMLR, 6638–6647.

[50] Xinwei Wu, Hechang Chen, Jiashu Zhao, Li He, Dawei Yin, and Yi Chang. 2021. Unbiased learning to rank in feeds recommendation. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. 490–498.

[51] Himank Yadav, Zhengxiao Du, and Thorsten Joachims. 2021. Policy-Gradient Training of Fair and Unbiased Ranking Functions. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1044–1053.

[52] Le Yan, Zhen Qin, Honglei Zhuang, Xuanhui Wang, Mike Bendersky, and Marc Najork. 2022. Revisiting two tower models for unbiased learning to rank. (2022).

[53] Tao Yang and Qingyao Ai. 2021. Maximizing Marginal Fairness for Dynamic Learning to Rank. In *Proceedings of the Web Conference 2021* (Ljubljana, Slovenia) *(WWW '21)*. Association for Computing Machinery, New York, NY, USA, 137–145. https://doi.org/10.1145/3442381.3449901

[54] Yisong Yue and Thorsten Joachims. 2009. Interactively optimizing information retrieval systems as a dueling bandits problem. In *Proceedings of the 26th Annual International Conference on Machine Learning*. 1201–1208.

[55] Yunan Zhang, Le Yan, Zhen Qin, Honglei Zhuang, Jiaming Shen, Xuanhui Wang, Michael Bendersky, and Marc Najork. 2023. Towards Disentangling Relevance and Bias in Unbiased Learning to Rank. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.

[56] Yunan Zhang, Le Yan, Zhen Qin, Honglei Zhuang, Jiaming Shen, Xuanhui Wang, Michael Bendersky, and Marc Najork. 2023. Towards Disentangling Relevance and Bias in Unbiased Learning to Rank. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Long Beach, CA, USA) *(KDD '23)*. 5618–5627.

[57] Haiyuan Zhao, Jun Xu, Xiao Zhang, Guohao Cai, Zhenhua Dong, and Ji-Rong Wen. 2022. Unbiased Top-k Learning to Rank with Causal Likelihood Decomposition. *arXiv preprint arXiv:2204.00815* (2022).

[58] Zhi Zheng, Zhaopeng Qiu, Tong Xu, Xian Wu, Xiangyu Zhao, Enhong Chen, and Hui Xiong. 2022. CBR: Context Bias aware Recommendation for Debiasing User Modeling and Click Prediction. In *Proceedings of the ACM Web Conference 2022*. 2268–2276.

[59] Honglei Zhuang, Zhen Qin, Xuanhui Wang, Michael Bendersky, Xinyu Qian, Po Hu, and Dan Chary Chen. 2021. Cross-positional attention for debiasing clicks. In *Proceedings of the Web Conference 2021*. 788–797.