Unbiased Learning to Rank: Counterfactual and Online Approaches

Harrie Oosterhuis
University of Amsterdam
Amsterdam, The Netherlands
oosterhuis@uva.nl

Rolf Jagerman
University of Amsterdam
Amsterdam, The Netherlands
rolf.jagerman@uva.nl

Maarten de Rijke
University of Amsterdam
Amsterdam, The Netherlands
derijke@uva.nl

ABSTRACT

This tutorial is about Unbiased Learning to Rank, a recent research field that aims to learn unbiased user preferences from biased user interactions. We will provide an overview of the two main families of methods in Unbiased Learning to Rank: Counterfactual Learning to Rank (CLTR) and Online Learning to Rank (OLTR) and their underlying theory. First, the tutorial will start with a brief introduction to the general Learning to Rank (LTR) field and the difficulties user interactions pose for traditional supervised LTR methods. The second part will cover Counterfactual Learning to Rank (CLTR), a LTR field that sprung out of click models. Using an explicit model of user biases, CLTR methods correct for them in their learning process and can learn from historical data. Besides these methods, we will also cover practical considerations, such as how certain biases can be estimated. In the third part of the tutorial we focus on Online Learning to Rank (OLTR), methods that learn by directly interacting with users and dealing with biases by adding stochasticity to displayed results. We will cover cascading bandits, dueling bandit techniques and the most recent pairwise differentiable approach. Finally, in the concluding part of the tutorial, both approaches are contrasted, highlighting their relative strengths and weaknesses, and presenting future directions of research. For LTR practitioners our comparison gives guidance on how the choice between methods should be made. For the field of Information Retrieval (IR) we aim to provide an essential guide on unbiased LTR to understanding and choosing between methodologies.

ACM Reference Format:
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 (WWW '20 Companion), April 20–24, 2020, Taipei, Taiwan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3366424.3383107

1 INTRODUCTION

Learning to Rank (LTR) has long been a core task in Information Retrieval (IR), as ranking models form the basis of most search and recommendation systems. Traditionally, LTR has been approached as a supervised task where there is a dataset with perfect relevance annotations [12]. However, over time the limitations of this approach have become apparent. Most importantly, datasets are very expensive to create [4] and user preferences do not necessarily align with the annotations [19]. As a result, interest in LTR from user interactions has increased significantly in recent years.

User interactions, often in the form of user clicks, provide *implicit feedback* [9], and while cheap to collect, they are also heavily biased [6, 23]. A prominent form of bias in ranking is *position bias*: users devote more attention to higher ranked documents, and consequently, the order in which documents are displayed affects the interactions that take place [6]. Another common form of bias is *item selection bias*: users can only interact with documents that are displayed; hence, the selection of displayed documents heavily affects which interactions are possible [18]. Naively ignoring these biases during the learning process will result in biased ranking models that are not fully optimized for user preferences [11]. The field of LTR from user interactions is mainly focussed on methods that remove biases from the learning process, resulting in unbiased LTR.

The first approach to unbiased LTR that we discuss in the tutorial is Counterfactual Learning to Rank (CLTR); it has its roots in user modeling [5]. CLTR relies on a user model that models observance probabilities explicitly; this model can be inferred separately [1, 3, 11, 21] or jointly learned [2]. By adjusting for observance probabilities, the effect of position bias can be removed from learning. This type of approach allows for unbiased learning from historical data, i.e., interactions collected in the past, as long as an accurate user model can be inferred.

The second approach is Online Learning to Rank (OLTR), which optimizes by directly interacting with users [22]. An OLTR method repeatedly presents a user with a ranking, observes their interactions, and updates its ranking model accordingly. Initially, these methods were based around interleaving methods [10] that compare rankers unbiasedly from clicks. Dueling Bandit Gradient Descent (DBGD) compares its current ranking model with a slight variation at each step, and updates toward the variation if such a preference is inferred [22]. While this approach has long formed the basis of OLTR [7, 15, 17, 20], recently fundamental problems with this approach were discovered [14]. Currently, there is another OLTR method: Pairwise Differentiable Gradient Descent (PDGD) that does not follow the DBGD procedure and thereby avoids these problems [16]. OLTR promises a responsive learning process where ranking systems adapt to users automatically and continuously.

Overall, we see that a big shift in unbiased LTR has taken place over the last three years: the emergence of CLTR from the field of user modeling and the replacement of the DBGD approach with PDGD in OLTR. It is important that practitioners and academics have a good understanding of each approach, their advantages and

WWW '20 Companion, April 20–24, 2020, Taipei, Taiwan

ACM ISBN 978-1-4503-7024-0/20/04.
https://doi.org/10.1145/3366424.3383107

limitations. Each approach is better suited for a certain situation, and understanding the applicability and effectiveness of each method is essential for LTR practitioners [8]. As the field has recently advanced in these different directions, now is the perfect time for a single tutorial to present all of these approaches together.

2 TUTORIAL OVERVIEW

In this tutorial, we provide an overview of the two main families of approaches to unbiased LTR and their underlying theory. We discuss the situations for which each approach was designed, and the places were they are applicable. Furthermore, we compare the properties of the two approaches and give guidance on how the choice between them should be made. For the field of IR we aim to provide an essential guide on unbiased LTR to understanding and choosing between methodologies.

Brief Schedule

The tutorial is divided in four parts:

- Part 1 **Introduction** (20 min) – Introduction to ranking, traditional LTR and user interactions, so that the audience understands the basic LTR concepts and the need for unbiased LTR.
- Part 2 **Counterfactual Learning to Rank** (70 min) – CLTR methods learn from historical interaction data and deal with biases by using an explicit model of observance probability.
- Part 3 **Online Learning to Rank** (70 min) – OLTR methods learn by directly interacting with users; they deal with biases by adding stochasticity to the displayed results.
- Part 4 **Conclusion** (20 min) – We conclude the tutorial by summarizing the previous sections and fully comparing and contrasting the two different approaches.

We note that a shorter (two-hour) version of this tutorial was part of a full-day tutorial at SIGIR'19 [13]; for WWW'20 the material has been updated and an hour of material has been added.

Publicly Available Material

The slides of this tutorial along with additional information are publicly available at https://ilps.github.io/webconf2020-tutorial-unbiased-ltr/.

ACKNOWLEDGMENTS

The development of the tutorial was partially supported by Ahold Delhaize, the Association of Universities in the Netherlands (VSNU), the Innovation Center for Artificial Intelligence (ICAI), the Netherlands Organisation for Scientific Research (NWO) under project nr 612.001.551. 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, Ivan Zaitsev, Xuanhui Wang, Cheng Li, Marc Najork, and Thorsten Joachims. 2019. Estimating Position Bias without Intrusive Interventions. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. 474–482.

[2] Qingyao Ai, Keping Bi, Cheng Luo, Jiafeng Guo, and W. Bruce Croft. 2018. Unbiased Learning to Rank with Unbiased Propensity Estimation. (2018), 385–394.

[3] Ben Carterette and Praveen Chandar. 2018. Offline Comparative Evaluation with Incremental, Minimally-Invasive Online Feedback. In *Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval* (Ann Arbor, MI, USA) *(SIGIR '18)*. ACM, New York, NY, USA, 705–714.

[4] Olivier Chapelle and Yi Chang. 2011. Yahoo! Learning to Rank Challenge Overview. In *Proceedings of the Learning to Rank Challenge*. 1–24.

[5] Aleksandr Chuklin, Ilya Markov, and Maarten de Rijke. 2015. Click models for web search. *Synthesis Lectures on Information Concepts, Retrieval, and Services* 7, 3 (2015), 1–115.

[6] 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* (Palo Alto, California, USA) *(WSDM '08)*. ACM, New York, NY, USA, 87–94.

[7] Katja Hofmann, Shimon Whiteson, and Maarten de Rijke. 2013. Balancing Exploration and Exploitation in Listwise and Pairwise Online Learning to Rank for Information Retrieval. *Information Retrieval* 16, 1 (Feb 2013), 63–90.

[8] Rolf Jagerman, Harrie Oosterhuis, and Maarten de Rijke. 2019. To Model or to Intervene: A Comparison of Counterfactual and Online Learning to Rank from User Interactions. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Paris, France) *(SIGIR'19)*. ACM, New York, NY, USA, 15–24. https://doi.org/10.1145/3331184.3331269

[9] Thorsten Joachims. 2002. Optimizing Search Engines Using Clickthrough Data. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Edmonton, Alberta, Canada) *(KDD '02)*. ACM, New York, NY, USA, 133–142.

[10] Thorsten Joachims. 2003. Evaluating Retrieval Performance using Clickthrough Data. In *Text Mining*, J. Franke, G. Nakhaeizadeh, and I. Renz (Eds.). Physica/Springer Verlag, 79–96.

[11] 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* (Cambridge, United Kingdom) *(WSDM '17)*. ACM, New York, NY, USA, 781–789.

[12] Tie-Yan Liu. 2009. Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval* 3, 3 (2009), 225–331.

[13] 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: From Gradient Boosting to Neural Networks and Unbiased Learning. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Paris, France) *(SIGIR'19)*. ACM, New York, NY, USA, 1419–1420.

[14] Harrie Oosterhuis. 2018. Learning to rank and evaluation in the online setting. 12th Russian Summer School in Information Retrieval (RuSSIR 2018).

[15] Harrie Oosterhuis and Maarten de Rijke. 2017. Balancing Speed and Quality in Online Learning to Rank for Information Retrieval. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (Singapore, Singapore) *(CIKM '17)*. ACM, New York, NY, USA, 277–286.

[16] 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* (Torino, Italy) *(CIKM '18)*. ACM, New York, NY, USA, 1293–1302.

[17] Harrie Oosterhuis, Anne Schuth, and Maarten de Rijke. 2016. Probabilistic multileave gradient descent. In *European Conference on Information Retrieval*. Springer, 661–668.

[18] Zohreh Ovaisi, Ragib Ahsan, Yifan Zhang, Kathryn Vasilaky, and Elena Zheleva. 2020. Correcting for Selection Bias in Learning-to-rank Systems. *arXiv preprint arXiv:2001.11358* (2020).

[19] Mark Sanderson. 2010. Test Collection Based Evaluation of Information Retrieval Systems. *Foundations and Trends in Information Retrieval* 4, 4 (2010), 247–375.

[20] 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* (San Francisco, California, USA) *(WSDM '16)*. ACM, New York, NY, USA, 457–466.

[21] 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* (Pisa, Italy) *(SIGIR '16)*. ACM, New York, NY, USA, 115–124.

[22] 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* (Montreal, Quebec, Canada) *(ICML '09)*. ACM, New York, NY, USA, 1201–1208.

[23] Yisong Yue, Rajan Patel, and Hein Roehrig. 2010. Beyond Position Bias: Examining Result Attractiveness As a Source of Presentation Bias in Clickthrough Data. In *Proceedings of the 19th International Conference on World Wide Web* (Raleigh, North Carolina, USA) *(WWW '10)*. ACM, New York, NY, USA, 1011–1018.