Unbiased Learning to Rank: Counterfactual and Online Approaches

Oosterhuis, H.; Jagerman, R.; de Rijke, M.

DOI
10.1145/3366424.3383107

Publication date
2020

Document Version
Final published version

Published in
The Web Conference 2020

License
CC BY

Citation for published version (APA):
https://doi.org/10.1145/3366424.3383107

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)
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:

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

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.
and 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 they were 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://ips.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


