

# Learning from User Interactions with Rankings: A Unification of the Field

Harrie Oosterhuis  
Radboud University  
*harrie.oosterhuis@ru.nl*

## Abstract

Ranking systems form the basis for online search engines and recommendation services. They process large collections of items, for instance web pages or e-commerce products, and present the user with a small ordered selection. The goal of a ranking system is to help a user find the items they are looking for with the least amount of effort. Thus the rankings they produce should place the most relevant or preferred items at the top of the ranking. Learning to rank is a field within machine learning that covers methods which optimize ranking systems w.r.t. this goal. Traditional supervised learning to rank methods utilize expert-judgements to evaluate and learn, however, in many situations such judgements are impossible or infeasible to obtain. As a solution, methods have been introduced that perform learning to rank based on user clicks instead. The difficulty with clicks is that they are not only affected by user preferences, but also by what rankings were displayed. Therefore, these methods have to prevent being biased by other factors than user preference. This thesis concerns learning to rank methods based on user clicks and specifically aims to unify the different families of these methods.

The first part of the thesis consists of three chapters that look at online learning to rank algorithms which learn by directly interacting with users. Its first chapter considers large scale evaluation and shows existing methods do not guarantee correctness and user experience, we then introduce a novel method that can guarantee both. The second chapter proposes a novel pairwise method for learning from clicks that contrasts with the previous prevalent dueling-bandit methods. Our experiments show that our pairwise method greatly outperforms the dueling-bandit approach. The third chapter further confirms these findings in an extensive experimental comparison, furthermore, we also show that the theory behind the dueling-bandit approach is unsound w.r.t. deterministic ranking systems.

The second part of the thesis consists of four chapters that look at counterfactual learning to rank algorithms which learn from historically logged click data. Its first chapter takes the existing approach and makes it applicable to top- $k$  settings where not all items can be displayed at once. It also shows that state-of-the-art supervised learning to rank methods can be applied in the counterfactual scenario. The second chapter introduces a method that combines the robust generalization of feature-based models with the high-performance specialization of tabular models. The third chapter looks at evaluation and introduces a method for finding the optimal logging policy that collects click data in a way that minimizes the variance of estimated ranking metrics.

---

By applying this method during the gathering of clicks, one can turn counterfactual evaluation into online evaluation. The fourth chapter proposes a novel counterfactual estimator that considers the possibility that the logging policy has been updated during the gathering of click data. As a result, it can learn much more efficiently when deployed in an online scenario where interventions can take place. The resulting approach is thus both online and counterfactual, our experimental results show that its performance matches the state-of-the-art in both the online and the counterfactual scenario.

As a whole, the second part of this thesis proposes a framework that bridges many gaps between areas of online, counterfactual, and supervised learning to rank. It has taken approaches, previously considered independent, and unified them into a single methodology for widely applicable and effective learning to rank from user clicks.

**Awarded by:** University of Amsterdam, Amsterdam, The Netherlands.

**Supervised by:** Maarten de Rijke.

**Available at:** <https://hdl.handle.net/11245.1/8ff3aa38-97fb-4d2a-8127-a29a03af4d5c>.

## Selected Publications

Harrie Oosterhuis and Maarten de Rijke. Differentiable unbiased online learning to rank. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 1293–1302. ACM, 2018.

Harrie Oosterhuis and Maarten de Rijke. Optimizing ranking models in an online setting. In *Advances in Information Retrieval*, pages 382–396, Cham, 2019. Springer International Publishing.

Harrie Oosterhuis and Maarten de Rijke. 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*. ACM, 2020a.

Harrie Oosterhuis and Maarten de Rijke. 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*, pages 137–144, 2020b.

Harrie Oosterhuis and Maarten de Rijke. Unifying online and counterfactual learning to rank. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining (WSDM'21)*. ACM, 2021.