Unifying Online and Counterfactual Learning to Rank

A Novel Counterfactual Estimator that Effectively Utilizes Online Interventions

Harrie Oosterhuis[1] and Maarten de Rijke[2]

Radboud University Nijmegen[1], the University of Amsterdam[2] and Ahold Delhaize[2]

Introduction

Unbiased Learning to Rank (LTR) from biased user clicks is traditionally divided into:

- **Online LTR:** Interactive algorithms that correct for bias by randomizing results.
- **Counterfactual LTR:** Algorithms that learn from historical click data, correct using a inferred model of bias.

In this paper, we bridge this traditional division by introducing a method designed for both counterfactual and online LTR:

- A counterfactual method that takes into account the effect of online interventions.

The Intervention-Oblivious Estimator

Based on the methods of Oosterhuis and de Rijke (2020) and Vardasbi et al. (2020), we introduce a single estimator that corrects for position-bias, item-selection bias, and trust-bias. For a logging policy π the click probability of on an item d is an expectation over the display rank k:

$$\begin{aligned} P(C=1 \mid d, \pi) &= \sum_{k=1} \pi(k \mid d)(\alpha_k P(R=1 \mid d) + \beta_k) \\ &= \mathbb{E}_k[\alpha_k \mid \pi] P(R=1 \mid d) + \mathbb{E}_k[\beta_k \mid \pi], \end{aligned}$$

where α_k and β_k are parameters per rank and $P(R=1 \mid d)$ is the probability that a user finds d relevant.

The Intervention-Oblivious Estimator is based on the inverse:

$$P(R=1 \mid d) = \frac{P(C=1 \mid d, \pi) - \mathbb{E}_k[\beta_k \mid \pi]}{\mathbb{E}_k[\alpha_k \mid \pi]}.$$

This is a **counterfactual** approach: it assumes the logging policy is completely static.

The Intervention-Aware Estimator

Insight: An intervention is simply a change of logging policy. Let Π be a set that contains the logging policy for each timestep: $\Pi = \{\pi_1, \pi_2, \ldots\}$. The click probability can be conditioned on Π:

$$\begin{aligned} P(C=1 \mid d, \Pi) &= \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{E}_k[\alpha_k \mid \pi] P(R=1 \mid d) + \mathbb{E}_k[\beta_k \mid \pi] \\ &= \mathbb{E}_k[\alpha_k \mid \Pi] P(R=1 \mid d) + \mathbb{E}_k[\beta_k \mid \Pi]. \end{aligned}$$

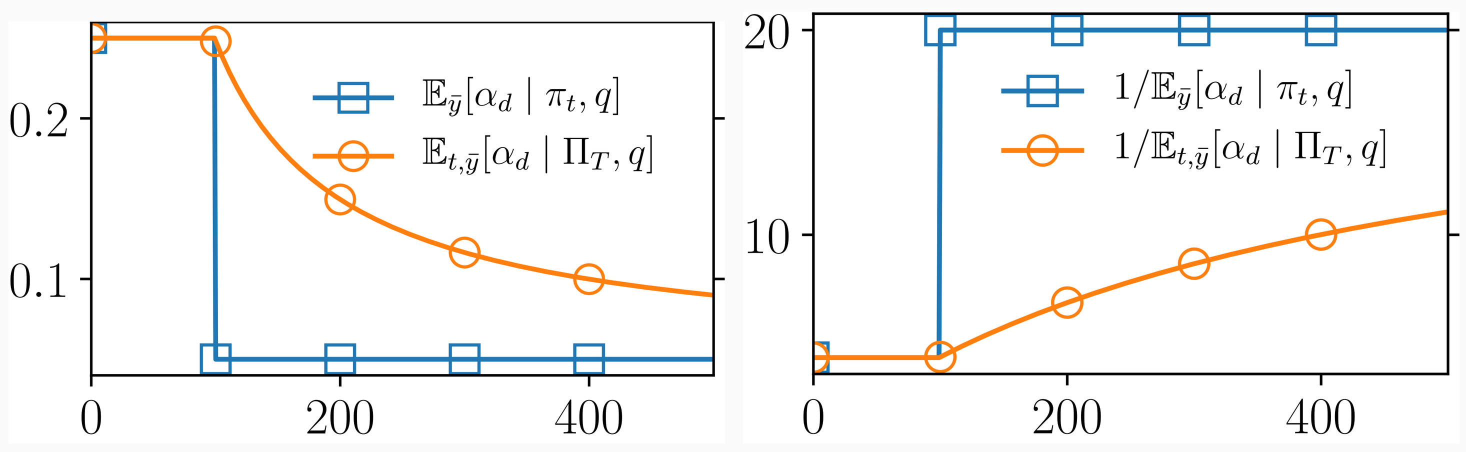
The Intervention-Aware Estimator is based on the inverse:

$$P(R=1 \mid d) = \frac{P(C=1 \mid d, \Pi) - \mathbb{E}_k[\beta_k \mid \Pi]}{\mathbb{E}_k[\alpha_k \mid \Pi]}.$$

This is a **counterfactual** and **online** approach: it takes into account online interventions for all its corrections, but it is also unbiased without any interventions.

Visualization

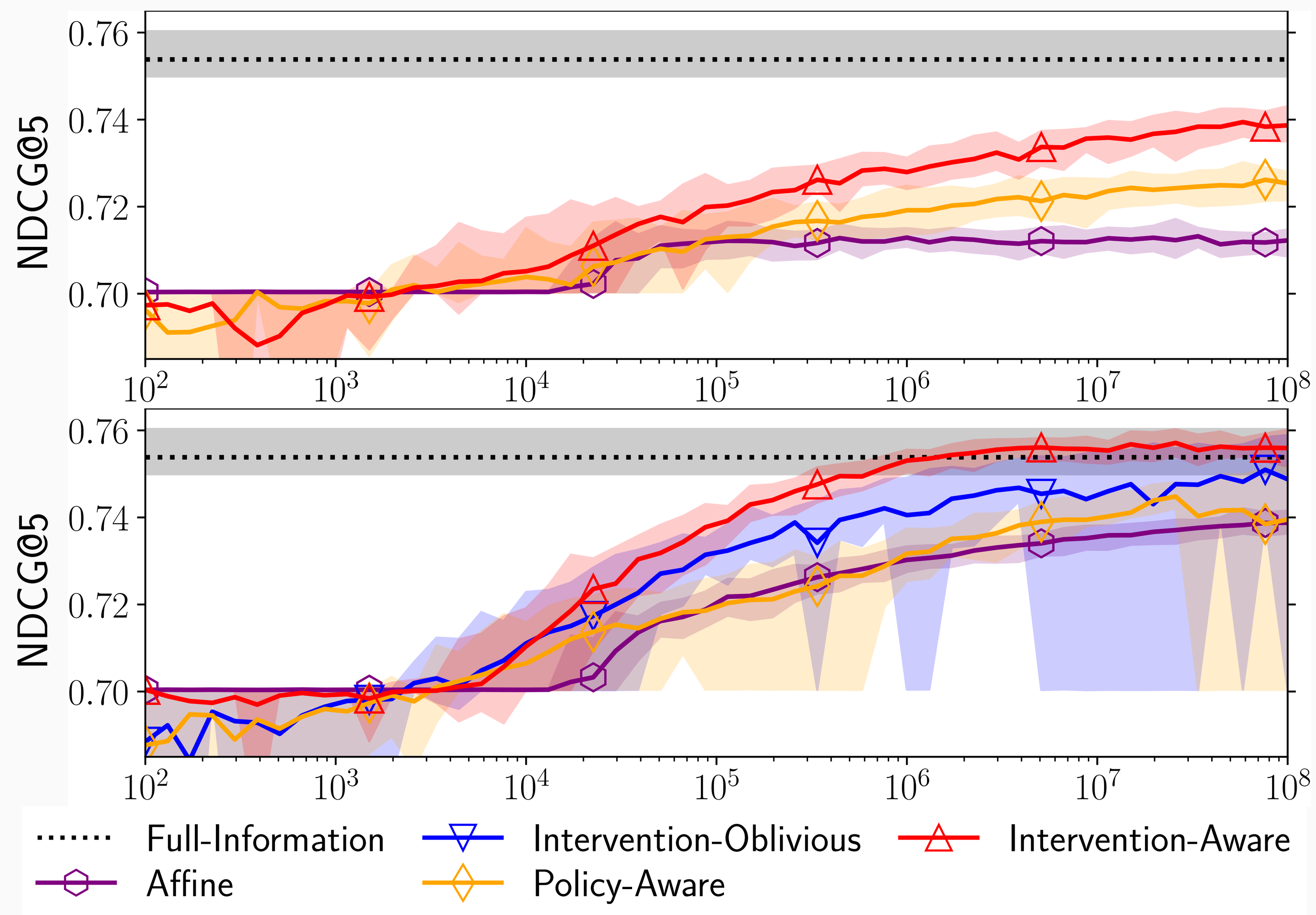
Example of the effect of a single intervention at $t = 100$:



Experimental Setup

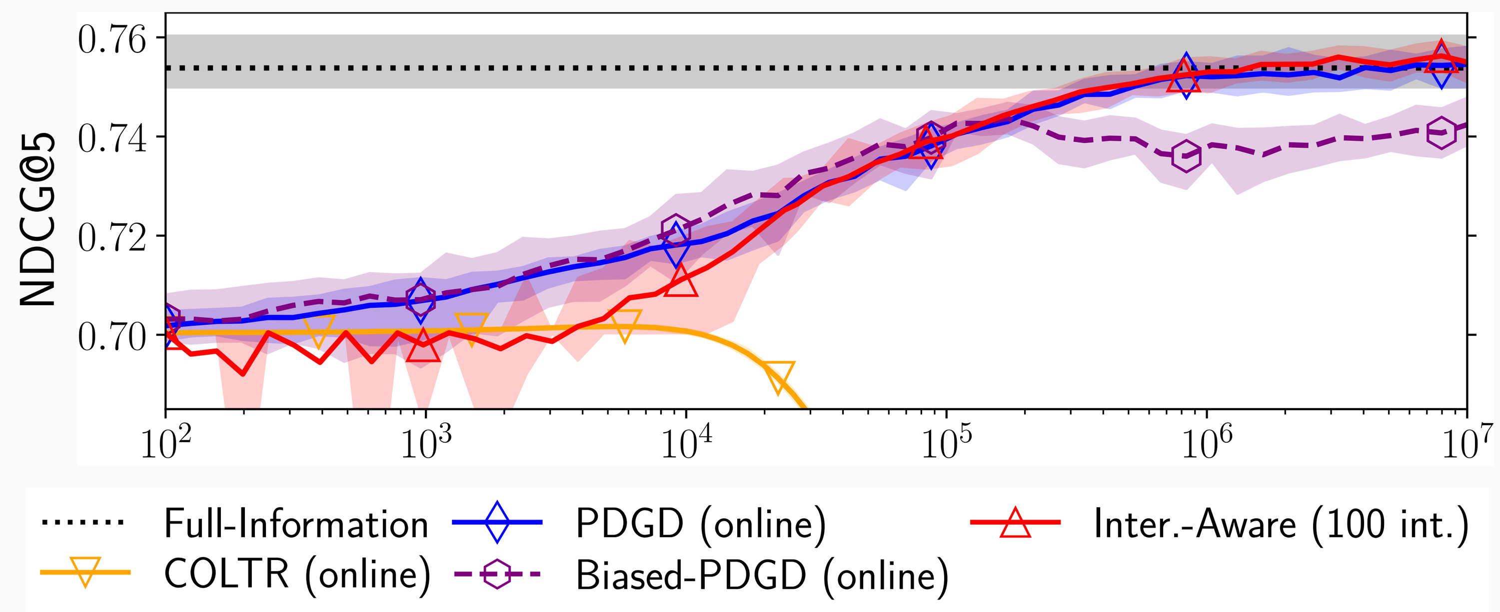
Results based on the Yahoo! Webscope dataset (Chapelle and Chang, 2011) with clicks simulated following a user model inferred by Agarwal et al. (2019) from real-world click data.

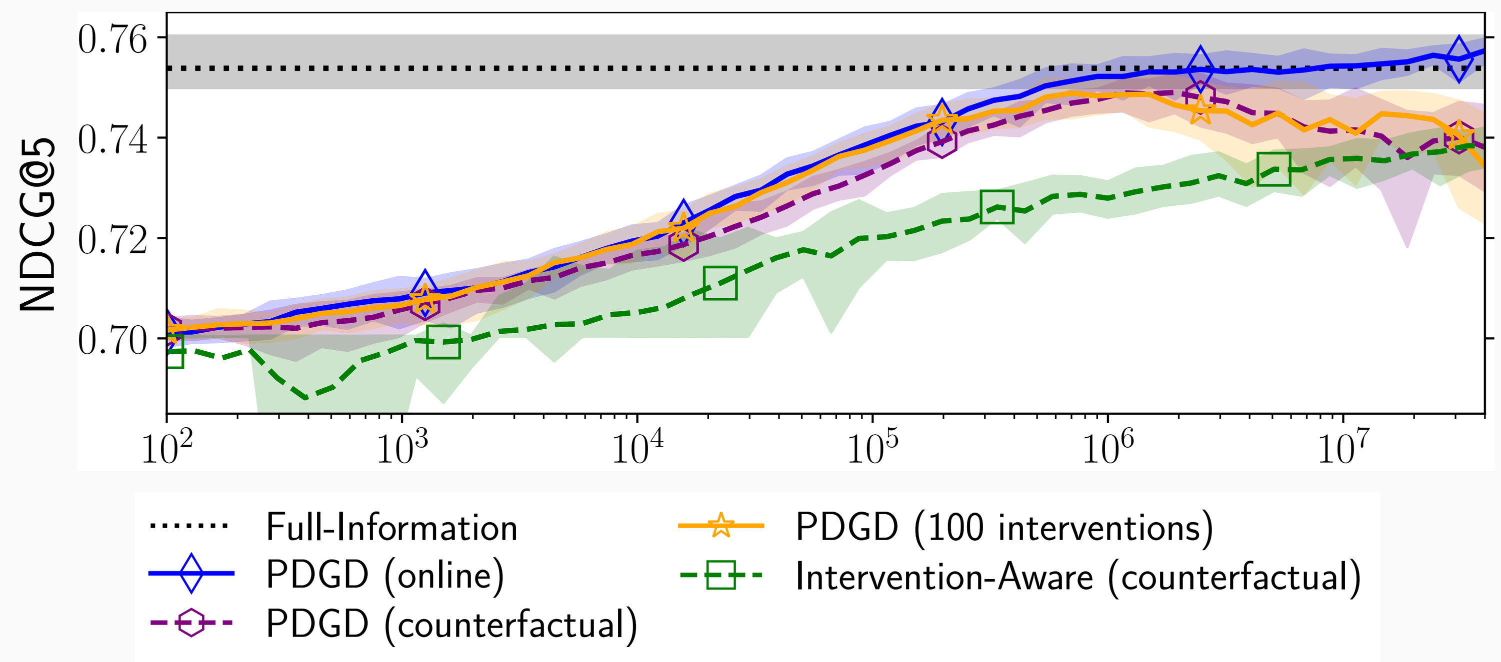
Comparison with Counterfactual LTR



Top: Data gathered with a static policy.
Bottom: Data gathered with 50 online interventions.

Comparison with Online LTR





Conclusion

The intervention-aware estimator is the most reliable choice for counterfactual learning and has online performance comparable to the state-of-the-art.

Public Code: `https://github.com/HarrieO/2021wsdm-unifying-LTR`

References

A. Agarwal, X. Wang, C. Li, M. Bendersky, and M. Najork. Addressing trust bias for unbiased learning-to-rank. In *The World Wide Web Conference*, pages 4–14. ACM, 2019.

O. Chapelle and Y. Chang. Yahoo! Learning to Rank Challenge Overview. *Journal of Machine Learning Research*, 14: 1–24, 2011.

H. Oosterhuis and M. 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*, pages 489–498. ACM, 2020.

A. Vardasbi, H. Oosterhuis, and M. de Rijke. When inverse propensity scoring does not work: Affine corrections for unbiased learning to rank. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2020.

Acknowledgements

This research was partially supported by the Netherlands Organisation for Scientific Research (NWO) under project nr 612.001.551 and by the Innovation Center for AI (ICAI). All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.

Radboud University