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 \( \pi \) the click probability of an on item \( d \) is an expectation over the display rank \( k \):

\[
P(C = 1 | d, \pi) = \sum_{k \in K} \pi(k | d) (\alpha_k P(R = 1 | d) + \beta_k)
\]

\[
= E_k[\alpha_k | \pi] P(R = 1 | d) + E_k[\beta_k | \pi],
\]

where \( \alpha_k \) and \( \beta_k \) are parameters per rank and \( P(R = 1 | d) \) is the probability that a user finds \( d \) relevant.

The Intervention-Oblivious Estimator is based on the inverse:

\[
P(R = 1 | d) = \frac{P(C = 1 | d, \pi) - E_k[\beta_k | \pi]}{E_k[\alpha_k | \pi]}
\]

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 \( \Pi \) be a set that contains the logging policy for each timestep: \( \Pi = \{ \pi_1, \pi_2, \ldots \} \). The click probability can be conditioned on \( \Pi \):

\[
P(C = 1 | d, \Pi) = \frac{1}{|\Pi|} \sum_{\pi \in \Pi} E_k[\alpha_k | \pi] P(R = 1 | d) + E_k[\beta_k | \pi]
\]

\[
= E_k[\alpha_k | \Pi] P(R = 1 | d) + E_k[\beta_k | \Pi].
\]

The Intervention-Aware Estimator is based on the inverse:

\[
P(R = 1 | d) = \frac{P(C = 1 | d, \Pi) - E_k[\beta_k | \Pi]}{E_k[\alpha_k | \Pi]}
\]

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 \( t = 100 \):**

\[
\frac{1}{E_k[\alpha_k | \pi, t]} \quad \frac{1}{E_k[\beta_k | \pi, t]} \quad 100
\]

**Comparison with Online LTR**

- **Full-Information**
- **PDGD (online)**
- **PDGD (counterfactual)**
- **Intervention-Aware (100 int.)**

**Comparison with Counterfactual LTR**

- **Full-Information**
- **Intervention-Oblivious**
- **Intervention-Aware**

**Conclusion**

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


**Acknowledgements**

This research was partially supported by the Netherlands Organization for Scientific Research (NWO) under project 023.001.013 and by the Innovation Centre 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.

**References**