Policy-Aware Unbiased Learning to Rank for Top-k Rankings

Harrie Oosterhuis[1], Maarten de Rijke[12]

July 9, 2020

University of Amsterdam[1], Ahold Delhaize[2]
oosterhuis@uva.nl, derijke@uva.nl
https://staff.fnwi.uva.nl/h.r.oosterhuis
https://staff.fnwi.uva.nl/m.derijke

Main Contributions

Main Contributions

- **Policy-Aware Estimator**
  - **Problem:** **Item-selection bias** in top-k ranking interactions.
  - **Solution:** New estimator can correct under a **stochastic logging policy**.

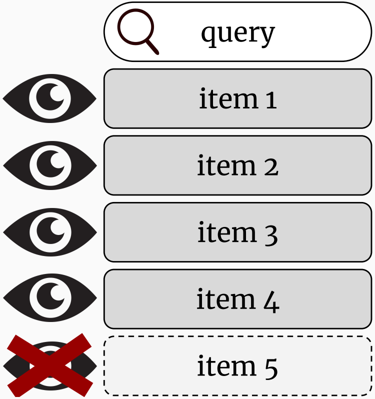
- **Loss functions for top-k Counterfactual Learning to Rank**
  - Adapt **supervised LTR SOTA LambaLoss** to optimize a counterfactual loss.

Introduction: Top-k Ranking

Learning to Top-k Rank

Top-k ranking: very prevalent in **search** and **recommendation**.

Goal: optimize a ranking model for top-k ranking.



Background:
Counterfactual Learning to Rank

Counterfactual Learning to Rank

Learn from **historically logged user clicks** (Joachims et al., 2017; Wang et al., 2016).

Problem:

- Clicks are **biased indicators** of preference (Craswell et al., 2008).

Existing solution:

- **Weight clicks** to correct for position bias.

Position Bias

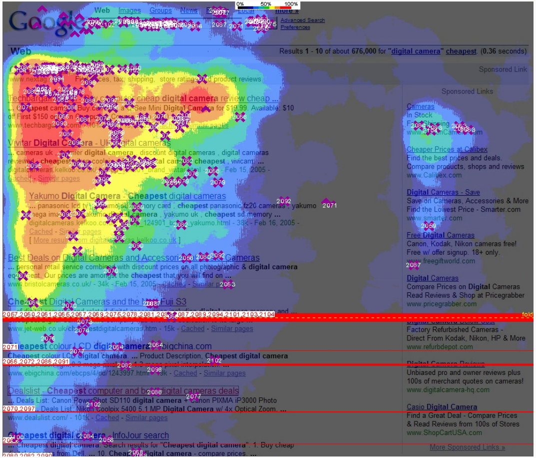


Image source: `http://www.mediative.com/`

Existing Policy Oblivious Estimator

For an item d, displayed ranking $\bar{R}$, and query q,

decompose the click probability according to examination hypothesis:

$$P(C = 1 \mid \bar{R}, q, d) = \overbrace{P(E = 1 \mid \bar{R}, d)}^{\text{examination}} \overbrace{P(C = 1 \mid E = 1, q, d)}^{\text{relevance}}. \tag{1}$$

Existing Policy Oblivious Estimator

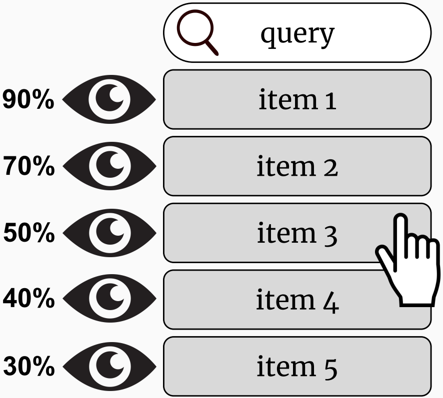
For an item d, displayed ranking $\bar{R}$, and query q,
decompose the click probability according to examination hypothesis:

$$P(C = 1 \mid \bar{R}, q, d) = \overbrace{P(E = 1 \mid \bar{R}, d)}^{\text{examination}} \overbrace{P(C = 1 \mid E = 1, q, d)}^{\text{relevance}}. \tag{1}$$

Existing work corrects for position bias by **Inverse Propensity Scoring** (Joachims et al., 2017; Wang et al., 2016). Given N displayed rankings for query q:

$$\text{relevance}(q, d) \approx \frac{1}{N} \sum_{i=1}^{N} \frac{c_i}{P(E = 1 \mid \bar{R}_i, d)}. \tag{2}$$

Existing Policy Oblivious Estimator: Visualized

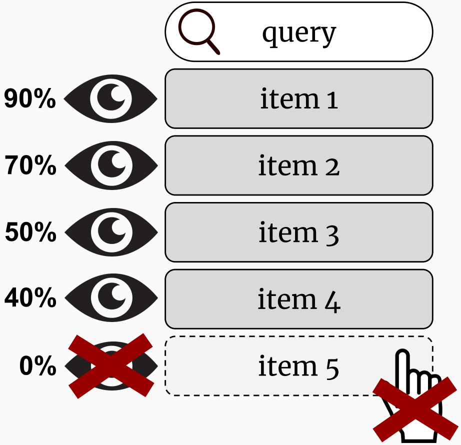


Item-Selection Bias

Item-Selection Bias

Items that are **not displayed cannot be examined**:

$$\text{rank}(d \mid \bar{R}) > k \rightarrow P(E = 1 \mid \bar{R}_i, d) = 0. \qquad (3)$$



Existing approach does **not work in top-k rankings**:

- No clicks to weight!

The Novel Policy-Aware Estimator

The Novel Policy-Aware Estimator

If displayed rankings are sampled from a **stochastic policy** π,
the click probability can be **conditioned** on the **policy**:

$$P(C = 1 \mid \pi, q, d) = \sum_{\bar{R}} \overbrace{\pi(\bar{R} \mid q)}^{\text{policy}} \overbrace{P(E = 1 \mid \bar{R}, d)}^{\text{examination}} \overbrace{P(C = 1 \mid E = 1, q, d)}^{\text{relevance}}. \tag{4}$$

The Novel Policy-Aware Estimator

If displayed rankings are sampled from a **stochastic policy** π,
the click probability can be **conditioned** on the **policy**:

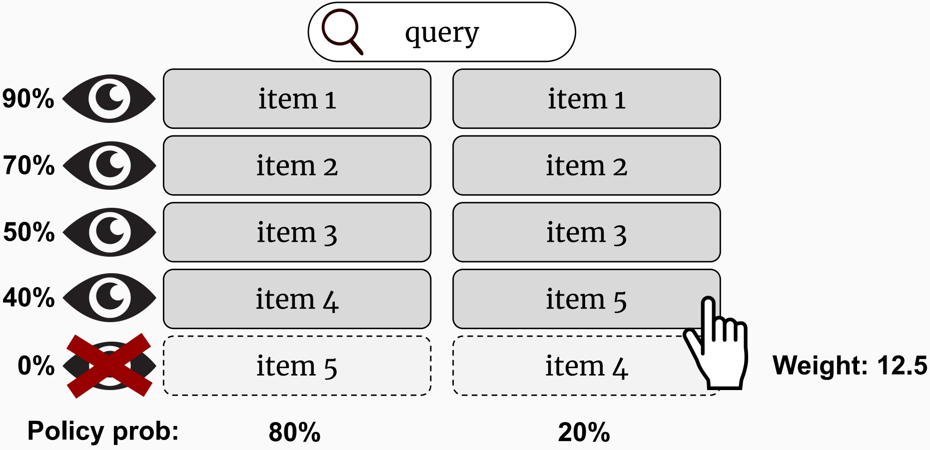
$$P(C = 1 \mid \pi, q, d) = \sum_{\bar{R}} \overbrace{\pi(\bar{R} \mid q)}^{\text{policy}} \overbrace{P(E = 1 \mid \bar{R}, d)}^{\text{examination}} \overbrace{P(C = 1 \mid E = 1, q, d)}^{\text{relevance}}. \quad (4)$$

Our **Policy-Aware Estimator** weights conditioned on the policy:

$$\text{relevance}(q, d) \approx \frac{1}{N} \sum_{i=1}^{N} \frac{c_i}{P(E = 1 \mid \pi, d)} = \frac{1}{N} \sum_{i=1}^{N} \frac{c_i}{\sum_{\bar{R}} \pi(\bar{R} \mid q) P(E = 1 \mid \bar{R}, d)}. \quad (5)$$

Unbiased if every item has a **non-zero chance** of being displayed in the top-k.

The Novel Policy-Aware Estimator: Visualized



Ranking Losses for Top-k Ranking

LambdaLoss for Counterfactual Learning to Rank

LambdaLoss is a state-of-the-art method for **supervised** LTR (Wang et al., 2018).

We adapt it to optimize a **counterfactual** loss (based on a new model f):

$$G_d = \frac{1}{N} \sum_{i=1}^{N} \frac{c_i \cdot \mathbb{1}[d_i = d]}{P(E = 1 \mid \pi, d)}, \qquad D_d = \frac{1}{\log_2(\text{rank}(d \mid f, q) + 1)}. \tag{6}$$

More adaptations to optimize **top-k** losses.

State-of-the-art supervised method applicable to counterfactual LTR!

Experiments

Experimental setup

Semi-synthetic setup based on **commercial LTR datasets:**

- Yahoo Webscope and MSLR-Web30k.

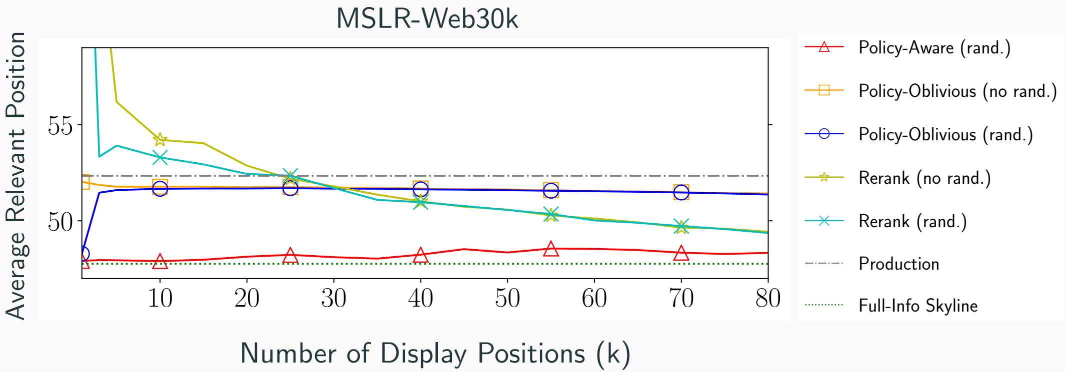
Displayed top-k rankings based on a pretrained '*production*' ranker:

- **without randomization,** and
- **with randomization:** random (remaining) item placed on position k.

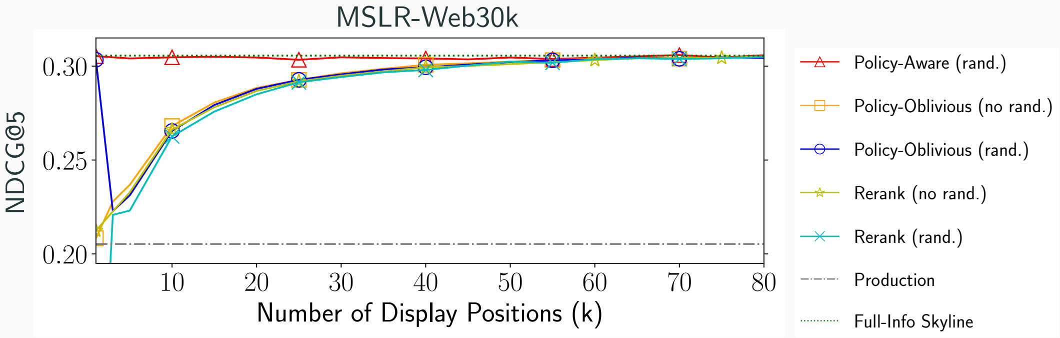
10^8 clicks based on dataset labels, with added **noise** and **position bias**.

Results

Results


MSLR-Web30k
Average Relevant Position
55
50
10
20
30
40
50
60
70
80
Number of Display Positions (k)
Policy-Aware (rand.)
Policy-Oblivious (no rand.)
Policy-Oblivious (rand.)
Rerank (no rand.)
Rerank (rand.)
Production
Full-Info Skyline

Results


MSLR-Web30k
NDCG@5
0.30
0.25
0.20
10
20
30
40
50
60
70
80
Number of Display Positions (k)
Policy-Aware (rand.)
Policy-Oblivious (no rand.)
Policy-Oblivious (rand.)
Rerank (no rand.)
Rerank (rand.)
Production
Full-Info Skyline

Related Work

Related Work

Correcting for Selection Bias in Learning-to-rank Systems by Ovaisi et al. (2020)

- Alternative method for dealing with item-selection bias.

Addressing Trust Bias for Unbiased Learning-to-Rank by Agarwal et al. (2019)

- Mention that LambdaLoss can be used for counterfactual LTR.

Conclusion

Conclusion

Main takeaways:

- Existing Counterfactual LTR cannot correct item-selection bias.
- Novel Policy-Aware estimator can under mild randomization.
- Adapted LambdaLoss works for counterfactual LTR.

References i

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.

N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey. An experimental comparison of click position-bias models. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, pages 87–94. ACM, 2008.

T. Joachims, A. Swaminathan, and T. Schnabel. Unbiased learning-to-rank with biased feedback. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 781–789. ACM, 2017.

Z. Ovaisi, R. Ahsan, Y. Zhang, K. Vasilaky, and E. Zheleva. Correcting for selection bias in learning-to-rank systems. In *Proceedings of The Web Conference 2020*, pages 1863–1873, 2020.

X. Wang, M. Bendersky, D. Metzler, and M. Najork. 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*, pages 115–124. ACM, 2016.

References ii

X. Wang, C. Li, N. Golbandi, M. Bendersky, and M. Najork. The lambdaloss framework for ranking metric optimization. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 1313–1322. ACM, 2018.

Acknowledgments

All content represents the opinion of the author(s), which is not necessarily shared or endorsed by their employers and/or sponsors.

Extra Results