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

Harrie Oosterhuis[1], Maarten de Rijke[23]

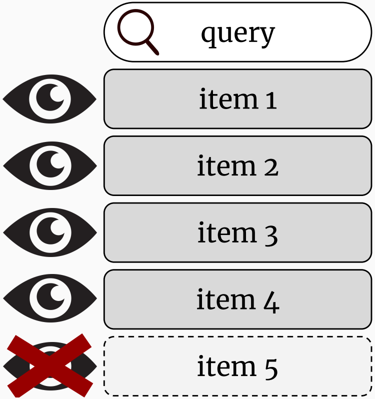
December 3, 2020

Radboud University[1], University of Amsterdam[2], Ahold Delhaize[3]
harrie.oosterhuis@ru.nl, derijke@uva.nl
https://twitter.com/HarrieOos

Introduction: Top-k Ranking

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

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



Background:
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.

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

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

Existing Policy Oblivious Estimator

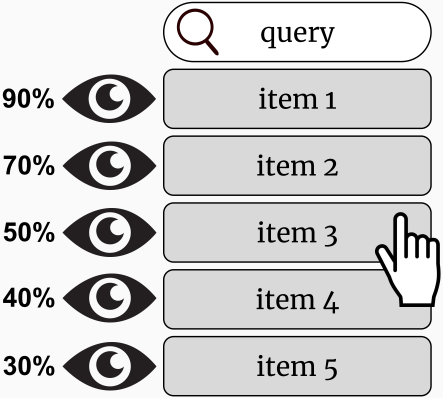
For an item d, displayed ranking $\bar{R}$, and query q,
decompose the click probability according to examination hypothesis:

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

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



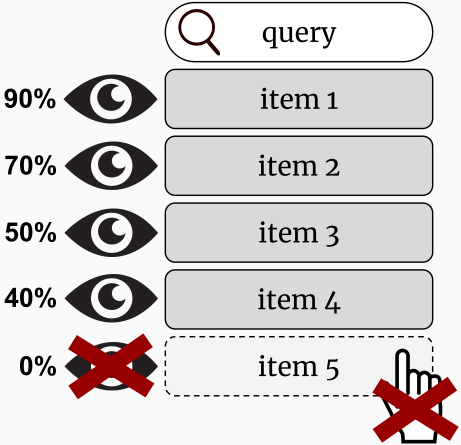
Item-Selection Bias

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**:

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

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

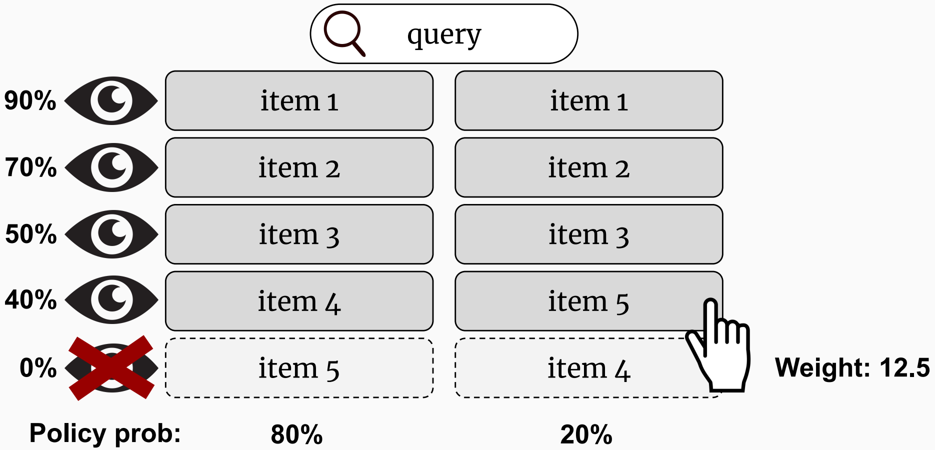
$$P(C=1 \mid \pi, q, d) = \sum_{\bar{R}} \overbrace{\pi(\bar{R} \mid q)}^{\text{policy}} \overbrace{P(E=1 \mid \bar{R}, d)}^{\text{examination}} \overbrace{P(C=1 \mid E=1, q, d)}^{\text{relevance}}. \quad (4)$$

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

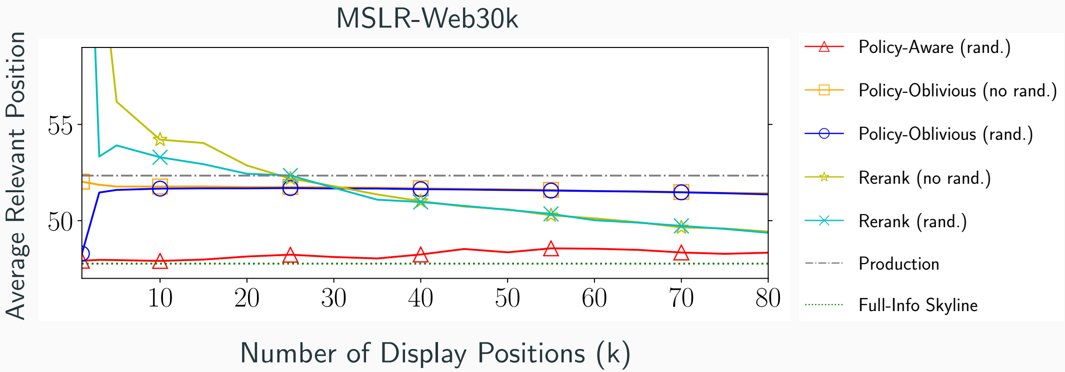
$$\text{relevance}(q, d) \approx \frac{1}{N} \sum_{i=1}^{N} \frac{c_i}{P(E=1 \mid \pi, d)} = \frac{1}{N} \sum_{i=1}^{N} \frac{c_i}{\sum_{\bar{R}} \pi(\bar{R} \mid q) P(E=1 \mid \bar{R}, d)}. \quad (5)$$

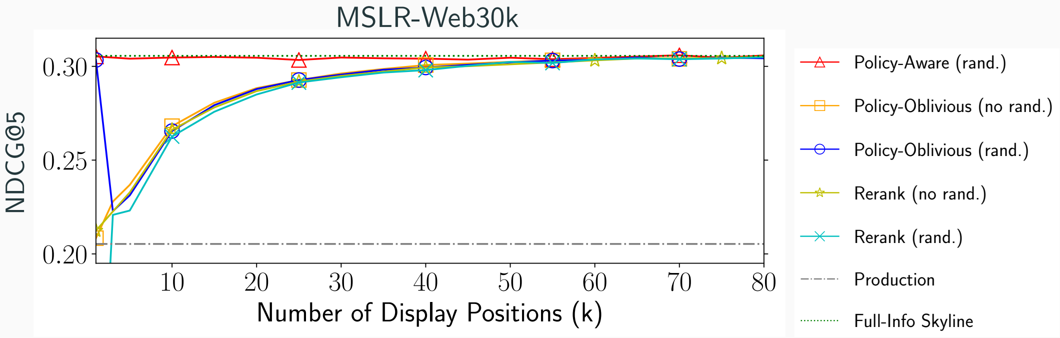
Unbiased if every item has a **non-zero chance** of being displayed in the top-k.

The Novel Policy-Aware Estimator: Visualized



Experimental 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


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

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
  (Not discussed in presentation).

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.

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.

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