Contrasting Neural Click Models and Pointwise IPS Rankers

Philipp Hager[1,2], Maarten de Rijke[1], Onno Zoeter[2]

April 5, 2023

[1]University of Amsterdam [2]Mercury ML Lab - Booking.com
p.k.hager@uva.nl, m.derijke@uva.nl, onno.zoeter@booking.com

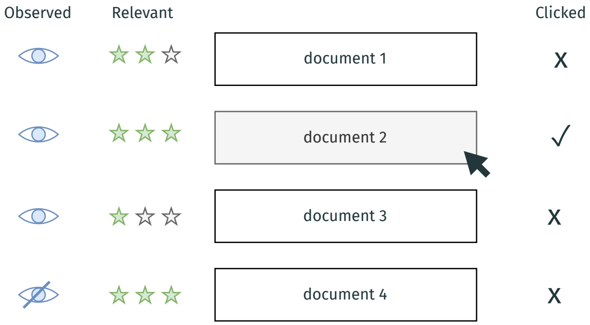
Unbiased LTR

Problem: Clicks are a biased indicator of item relevance

Unbiased LTR: Learn ranking models from biased user interactions

Click biases:

- Selection bias
- **Position bias**
- Trust bias
- ...

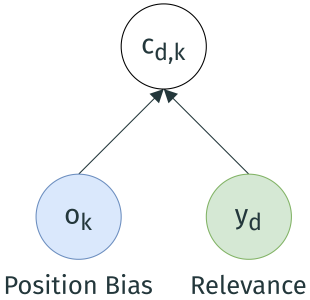


Approaches for Unbiased LTR I

Probabilistic click models

- **Predict biased clicks**
- Jointly infer bias and relevance parameters
- Pointwise rankers

Position-based Model (PBM)

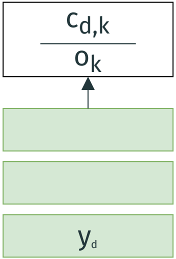


Approaches for Unbiased LTR II

Inverse Propensity Scoring (IPS)

- **Predict unbiased relevance** by weighting clicks inversely to bias
- Decouple bias and relevance estimation
- Generalizes over features to require fewer observations of the same query-document pair
- Introduced as pairwise and listwise rankers

Re-weight clicks in loss using IPS



Relevance

Approaches for Unbiased LTR III

Two sides of the same coin?

- Both methods assume the same user models (PBM, Cascade)
- Neural click models using features are popular in industry applications
- Introduction of pointwise IPS methods

How do both methods compare theoretically and empirically for pointwise ranking under position bias?

Comparing pointwise loss functions

PBM: Binary cross-entropy between predicted $\hat{c}_{d,k} = \hat{y}_d \cdot \hat{o}_k$ and observed clicks:

$$\mathcal{L}_{\text{pbm}}(\hat{y}, \hat{o}) = - \sum_{(d,k) \in D} c_{d,k} \cdot \log(\hat{y}_d \cdot \hat{o}_k) + (1 - c_{d,k}) \cdot \log(1 - \hat{y}_d \cdot \hat{o}_k). \tag{1}$$

IPS: Binary cross-entropy weighting clicks inversely by position bias [Bekker et al., 2019]:

$$\mathcal{L}_{\text{ips}}(\hat{y}, \hat{o}) = - \sum_{(d,k) \in D} \frac{c_{d,k}}{\hat{o}_k} \cdot \log(\hat{y}_d) + \left(1 - \frac{c_{d,k}}{\hat{o}_k}\right) \cdot \log(1 - \hat{y}_d). \tag{2}$$

Comparing unbiasedness

- **IPS** is unbiased when the **true position bias is known** and users behave according the PBM [Saito et al., 2020]
- **Click models jointly inferring parameters** are not always consistent estimators of relevance [Oosterhuis, 2022]

Assuming that position bias is known, we show in our paper that a **click model only inferring relevance** is also unbiased

Experimental setup

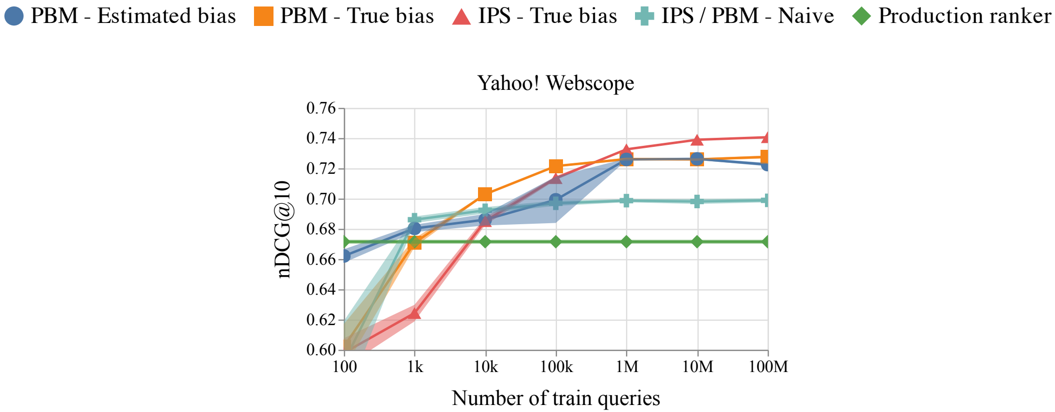
Semi-synthetic clicks on MSLR30K, Yahoo! Webscope, and Istella-S:

- Sample a query uniformly at random
- Pre-rank documents with "production ranker" (LambdaMART 20 training queries)
- Sample clicks using a PBM user model

We simulate:

- **Position bias at rank k:** $o_k = \left(\frac{1}{k}\right)^{\eta}$
- **Graded relevance:** $y_d = \epsilon + (1 - \epsilon) \cdot \frac{2^{s_d}-1}{2^4-1}$, click noise $\epsilon = 0.1$

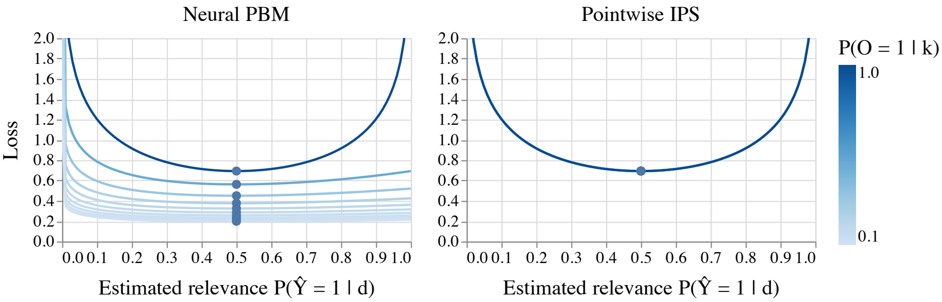
Results



We find significant empirical differences between both methods on 2/3 datasets.

Comparing loss magnitude

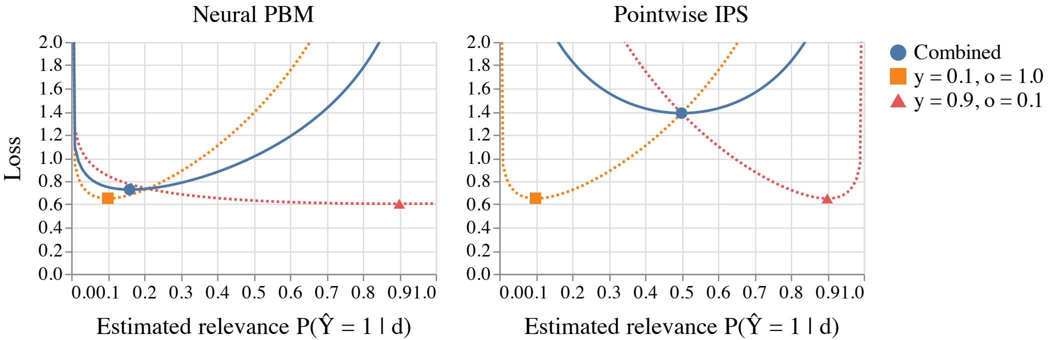
Observation: Items at lower positions contribute less to the click model's loss



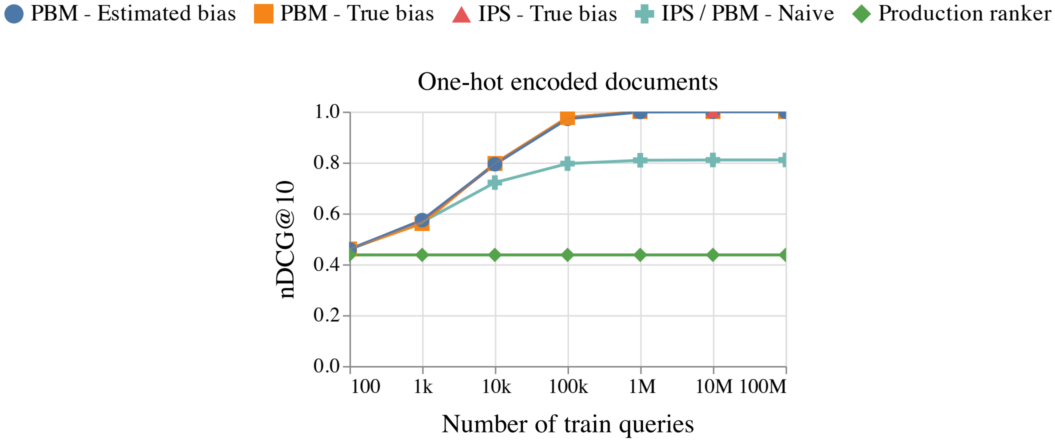
Loss for a single document of relevance $y_d = 0.5$ under increasing position bias

Comparing loss magnitude

Does generalizing over document features introduce bias?



Estimating relevance per query-document pair



Summary

Conclusion

- Click models and IPS perform equivalently **if the position bias is known** and relevance is **estimated per query-document pair**.
- The neural click model used in this work seems to be **affected by position bias** when generalizing over **shared, sometimes conflicting, document features**.

Limitations and future work

- Additive two-tower models, RegressionEM
- Extend analysis to pairwise and listwise methods

p.k.hager@uva.nl

References i

[Bekker et al., 2019] Bekker, J., Robberechts, P., and Davis, J. (2019). **Beyond the selected completely at random assumption for learning from positive and unlabeled data.** In *Machine Learning and Knowledge Discovery in Databases: European Conference (ECML PKDD)*.

[Oosterhuis, 2022] Oosterhuis, H. (2022). **Reaching the end of unbiasedness: Uncovering implicit limitations of click-based learning to rank.** In *International Conference on the Theory of Information Retrieval (ICTIR)*.

[Saito et al., 2020] Saito, Y., Yaginuma, S., Nishino, Y., Sakata, H., and Nakata, K. (2020). **Unbiased recommender learning from missing-not-at-random implicit feedback.** In *International Conference on Web Search and Data Mining (WSDM)*.

Click model only inferring relevance

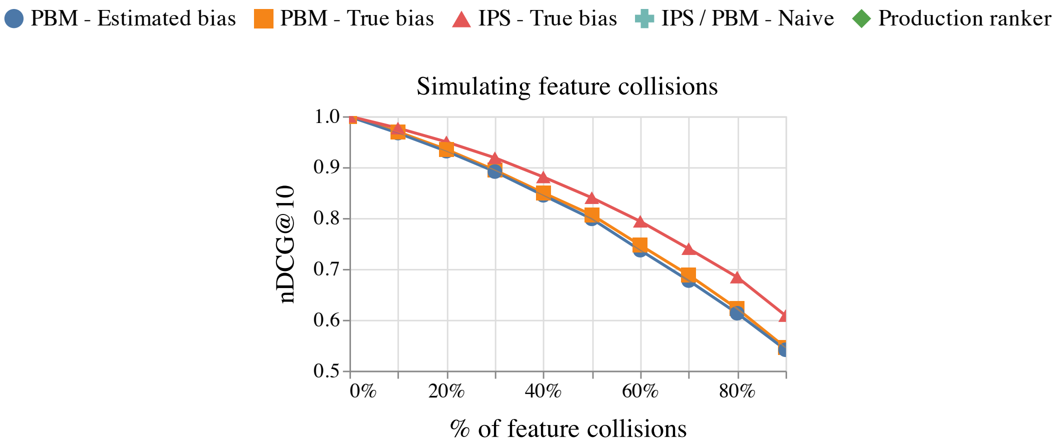
Find the ideal model that minimizes $\mathcal{L}_{\text{pbm}}$:

$$\begin{aligned}\frac{\partial \mathcal{L}_{\text{pbm}}}{\partial \hat{y}} &= 0 \\ -\frac{c - \hat{o}\hat{y}}{\hat{y}(1 - \hat{o}\hat{y})} &= 0 \\ \hat{y} &= \frac{c}{\hat{o}}.\end{aligned} \tag{3}$$

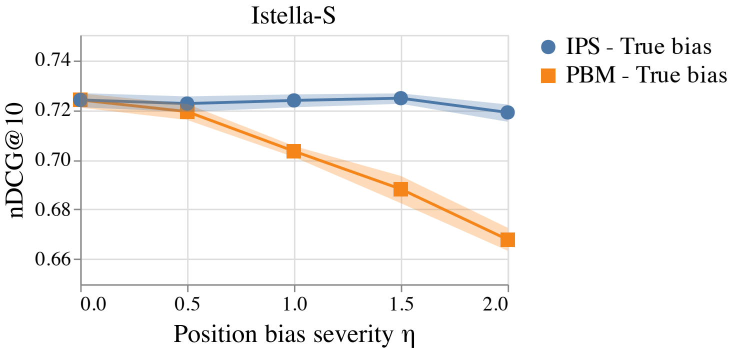
The click model optimizes for unbiased relevance if $\forall k \in K, \hat{o}_k = o_k$

$$\mathbb{E}\left[\hat{y}\right] = \frac{\mathbb{E}\left[c\right]}{\hat{o}} = \frac{oy}{\hat{o}}. \tag{4}$$

Estimating relevance per query-document pair



Mitigating known position bias



Increasing the (known) position bias